

LESSON 1: INTRODUCTION

The most reliable and efficient neuroimaging techniques are PET and MR. Instead, CT is largely used in clinical analysis but not in research fields and this is reason why we will just look on these two instrumentations. Another important technology is the fMRI. It is one of the most important tools in neuroscience to understand which area of the brain is activated during a specific task.

Information classification

Generally, in the fields of neuroimaging, we have to distinguish between anatomical and functional images.

- *Anatomical Images:* Refers to the detail of the anatomical parcellation of the brain. This organ is the most complex in our body and it is made up of different structures. During the course we will investigate deeply all the subdivision, but one of the first classification that is performed for a certain area is due to the localization inside the brain of the area itself. Mainly, we can recognize two different positioning, the inner localizations known also as subcortical regions and the outer localizations or cortical regions. In research studies and in clinical investigation we could have available some 3D or 2D anatomical structure. Usually, we must consider a high spatial resolution when the voxel that made up the 3D image have dimension of $1 \times 1 \times 1 \text{ mm}$.
- *Functional Images:* These types of images do not refer to the anatomical organization of the organ or the system but give us some information about the regions activated. Nowadays, in the market, there are available different kind of methodology to access to this information. In most of the study the functional image is representing a 4D object, where the first three dimension are the one already explained also in the anatomical images (physical encumbrance) while the fourth dimension represent the time. Thanks to this information we can see how the brain activity evolves through time. Due to the necessity of accessing to a fourth dimension the spatial resolution decrease (1.5 to 4 mm). Generally, the spatial and the time resolution are inversely proportional. This is due to the necessity of a certain period to collect enough data to correctly reconstruct the image.

Anatomical images acquisition techniques:

X-Ray

CT

NMR

Functional images acquisition techniques:

EEG

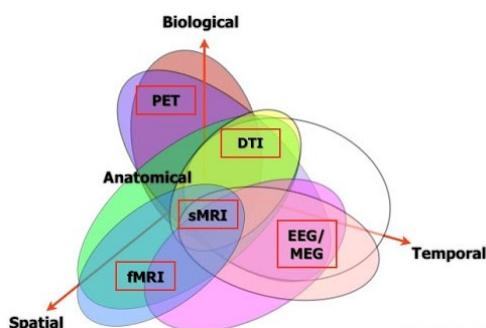
PET

MRI

Optical Imaging

Scales of analysis

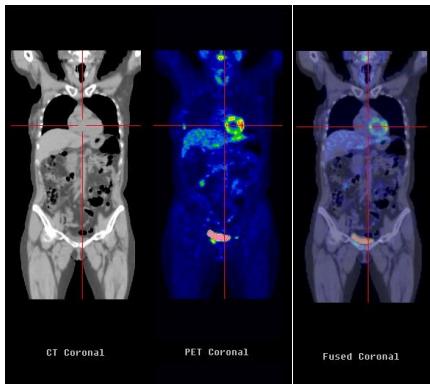
Now when we talk about neuroimaging, we have different techniques largely used in this field and all these differs one from each other by the scales of biological resolution, temporal resolution, and spatial resolution. In the graph down below is represented the differences in terms of scale of all the technics analysed.



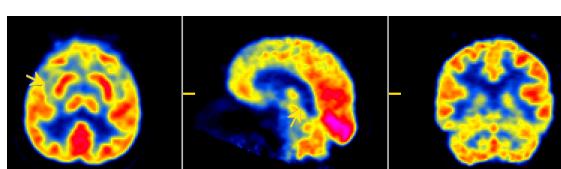
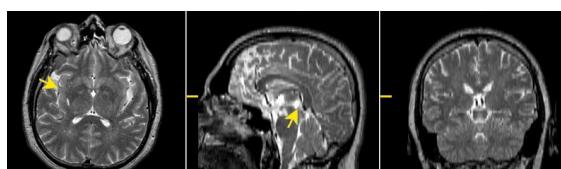
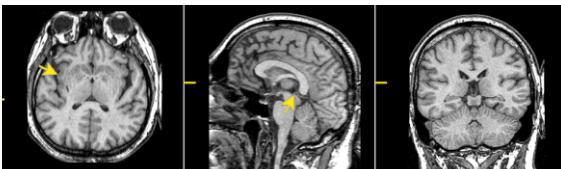
The key methodology that is used to correctly read the content of this graph is the next one. By as instance, focusing on EEG or MEG techniques we could say that these methodologies have a very high temporal resolution but a very low spatial and biological resolution. In fact, these two-measurement systems could not access to a very low biological level (cellular level) and could only describe respectively the electrical and magnetic activity of population of neurons. Instead, the PET is characterised by a very high biological resolution because it can assess to the activity of certain group of neurons localized in a specific region of the brain (this carry also a physiological meaning) but by looking on the temporal and spatial resolution they are very low due to the spatial imprecision and high amount of time necessary to develop completely a PET image.

Merging of information

Down below are reported some images obtain, by the usage of different techniques of the whole body of a subject and also some images representative of just the head.



In the left section of the page are reported three different images. The first one is referred to a CT scan executed on the coronal section. The second one is the result obtained from a coronal PET scan and finally the last image report the merge of the two previous cited images. Nowadays is available the techniques called co-registration of multiple information at the same time to get a better clinical picture of the subject.



The information that could be merged are not just only functional and anatomical images like in the ones reported before, but we can also merge different MR scan and PET data. As instance like the example reported in the left section of this page, we can combine the information that come from a T1 MRI scan with a T2 MRI scan to investigate the anatomical structure and the position of the cerebral lesion or alteration. Usually, in clinical practice the T1 MRI scan is used to access to the anatomical information while the T2 MRI scan is used to highlight anatomical abnormalities inside the results obtained with previous technique. Moreover, to get the functional approval of all the results we can also use some PET data.

Tools in neuroscience

We can recognize two main tools in neuroscience. In the first class are contained the *surface – based acquisition* while in the second, we could find the *tomographic – based acquisition*. The EEG (Electroencephalography), the fNIRS (functional Near – Infrared Spectroscopy) and the MEG

(Magnetoencephalography) are part of the first class because of the usage of electrodes, probes and other surface techniques to record all the possible information. Some example of the second class are PET and MRI because they use some reconstruction algorithm to get as results all the information.

Brain study's aim

The aim of study the brain is to improve three major categories: neuroscience field, medicine filed and finally computing field. In the first case we are interested to achieve a unified, multi-level understanding of the human brain that integrates data about the healthy and diseased brain. In terms of medical field is trying to find out some specific biomarkers to identify biological signatures of specific brain diseases and finally, by talking about the computing field we are interested to understand how the brain work to develop some new neurorobotic technologies based on the brain's circuitry. Moreover, thanks to the efficiency that is shown from the brain by understanding it we could be able to produce more efficient and more reliable computer and machine. In general, the methodology to find out new and relevant results are not obtained by just recording some images or time series, but we need to perform also some mathematical, geometrical, and analytical evaluation of our data and find out some physiological meaning to associate to it. This is how a study should be performed.

Brain connectivity

Some methodology that are used to compute information that regard the brain is as instance the graph theory. This is a very powerful tool which was developed thought the last years and look to the brain as a graph. The key element of this analysis are the nodes and the edges where the first one refers to a certain brain region (population of neurons, specific cell) while the second element represent the connection between different nodes. Usually, by applying this technique we can perform two types of studies, the first one is also known as anatomical analysis that highlight the real connection between brain regions, while the second one could be performed with a functional analysis and shows the functional connection between different brain regions. There are several companies that are investing on this field, such as Neuralink (Elon Musk's colossus), alleninstitue, and also IEEE.

BCI

BCI are a non – new research field which is catching on widely nowadays. They were first introduced in earlier 2000 from Wolpaw and his colleagues. This technology is meanly based on a close loop that implement different and algorithm. Exists different types of BCI and they depend on the technology used to assess to the brain activity information. Usually, we could subdivide the BCI classes in three: invasive, semi-invasive and finally non-invasive. At this moment, the only company that is trying to develop a BCI and put it in the market is Neuralink.

LESSON 2: BASIC MRI PRINCIPLE

Nueromarketing and neurocriminology

Nowadays there are some other field of application of neuroimaging as instance neuromarketing and neuroeconomics. Neuromarketing is a new field of marketing which uses medical technologies such as functional Magnetic Resonance Imaging to study the brain's responses to marketing stimuli. The first application of neuromarketing is dated 2002. In this field are used results coming from different imaging techniques merge then together to study the brain response when a subject facing with specific product. This knowledge raised a lot of issues about the ethical purposes. Even if all the imaging techniques used are non – invasive the information carried out could be considered one of the most ethical involved data. In 1991 was first used some neuroimaging techniques to study the relation between the law and the field of neuroscience. Neurolaw explores the effects of discoveries in neuroscience on legal rules. Neurocriminology is documenting structural and functional brain impairments not just in antisocial, violent, and psychopathic individuals, but also in spouse abusers and white – collar criminals. Nuerocriminologists are proposing a neurodevelopmental contribution to crime causation. Neurolaw encompasses ethical questions regarding nootropics, more commonly known as mind – enhancing drugs. Nootropics referred to as smart drugs which are memory, neuro, cognitive and intelligence enhancers.

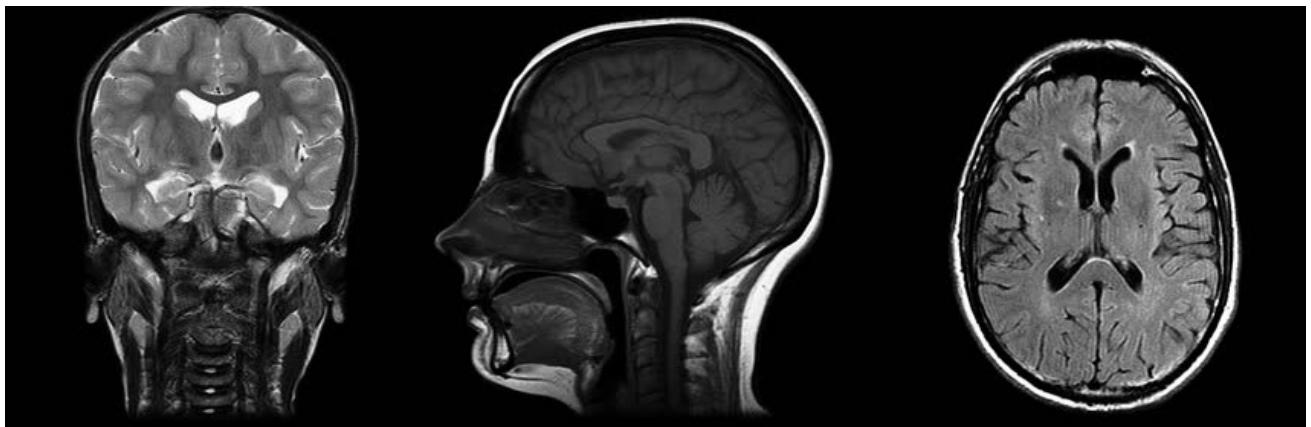
Terminology

Before introducing and summarize the main brain imaging technique we need to introduce some frequent terminology used in this course.

- *Volumes*: Cerebral volumes in a specific analysis.
- *Number of volumes*: Number of times that the same brain volume has been acquired over time.
- *Slice*: Section of the brain with a preferred direction. The slice has an important characteristic defined from the thickness of the slice. This let us see the slice as a 3D object and not just a 2D object.
- *Pixel or voxel*: Smallest unit inside the image, the first one is founded in 2D image while the second in 3D or 4D images.
- *Dicom format*: The usual format of medical images. Matlab could read Dicom format through the usage of the appropriate function. This is a very heavy and old file format.
- *Nifti format*: It is format lighter than Dicom more suit to do elaboration and image processing. It is a file format readable from Matlab too.

When we take some old dataset to perform new type of analysis, we may face with some other file format as ECAT. These ones are heavier, and they are not readable directly from Matlab. The data organization inside these types of files is quite different from the newest one.

One other important terminology that has to be explored is the different plane that we can get from an imaging technique. The first is the *axial* one. The acquisition in this case starts from the end of the neck and point towards the end of the head. The second type is the *sagittal* one. It is a lateral section of the brain, the neuroimage is made up from an ear and end in the other. The last typology is the coronal. The acquisition starts from the ineon with end in the nasion.



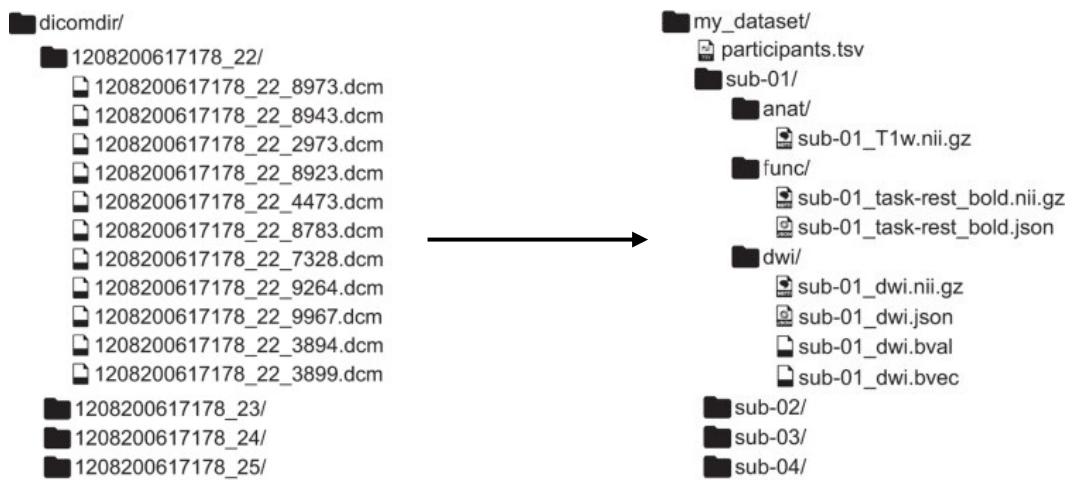
Coronal plane

Sagittal plane

Axial plane

Brain Imaging Data Structure

One important aspect that we need to consider when we deal with brain imaging data is how to organise files. MRI has been used to study the human brain for over 20 years. Despite similarities in experimental designs and data types each researcher tend to organize and describe their data in their own way. Because of this until this last years researchers have produced a lot of heterogeneity in data description practices and this leads to problem in sharing data (even within the same lab), unnecessary manual metadata input when running processing pipelines and no way to automatically validate completeness of a given dataset. To solve this problem has been introduced the Brain Imaging Data Structure (*BIDS*). It is a new standard for organizing results of a human neuroimaging experiment. The adoption of *BIDS* is crucial, it is important to do not reinvent the wheel in the field of file organization, some metadata are better than no metadata. Moreover, it does not rely on external software as instance databases or complicated file formats (*RDF*). The aim of the application of this method is to give an international standard to organize the data to share more easily research's results between different researchers teams.



In the image reported above we can appreciate what the *BIDS* stand for. It is a data reorganization strategy that combine the possibility to easy organize the data structure that cover the most common experiments.

Positron Emission Tomography (PET)

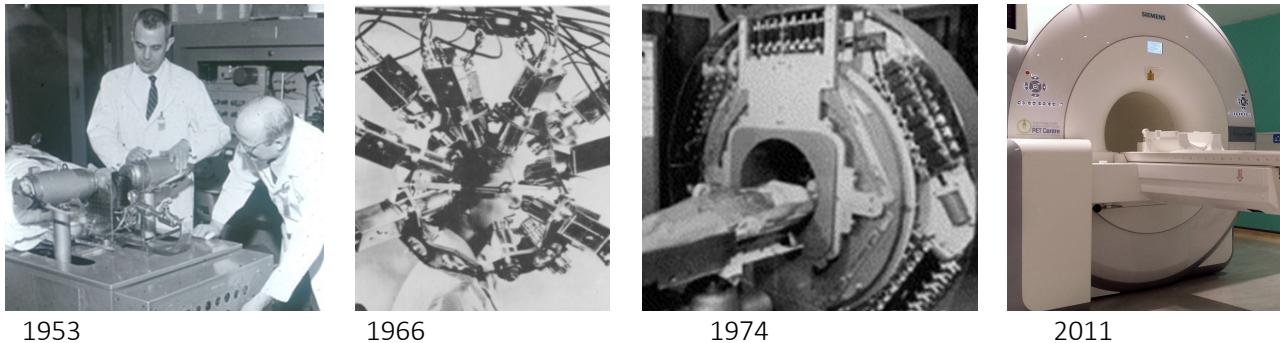
The Positron Emission Tomography, also known as PET, is a nuclear medical imaging technique which produces a three – dimensional image of functional processes in the body. It requires a scanner for the image acquisition and a cyclotron for radiolabelled molecules production. Usually, we can find this instrumentation inside the nuclear medicine department. It gives just a functional image and not an anatomical one. Sometimes not all the hospital has a cyclotron because it is a very expensive structure. Just the most advanced research centres have both the instrumentations as instance the San Thomas in London, the John Redcliff in Oxford and Addenbrooke's Hospital in Cambridge. The hospital of Padova instead does not have the cyclotron and the nearest is located in Verona. All the hospital not equipped with cyclotrons need to know the time needed to provide the radiotracers from the nearest clinical centre available. This could be restrictive respect to the tracer usable in a certain situation. The radiotracers are based on the next radionuclides.

RADIOMUCLIDE	T _{1/2}	
¹¹ C	20.4 min	"naturale"
¹³ N	10.0 min	"naturale"
¹⁵ O	2.0 min	"naturale"
¹⁸ F	109.8 min	"pseudo naturale"

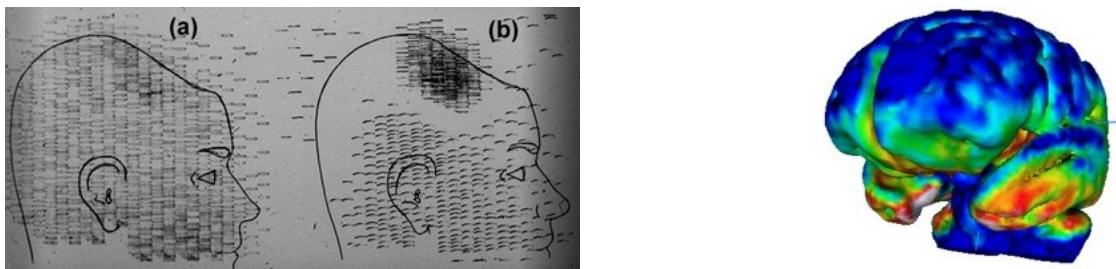
The decay time or the half – life could be very high or very low. All the minutes order half – life radionuclides are a very big problem because to use them the cyclotron has to be in loco. The radiotracers that could be produced by the usage of the previous mentioned radionuclides are a lot and the choose of one instead of another is to study deeply one typology of metabolism than another. Down below is reported a tab that present the main radiotracers:

1	Radiotracer	Targets	Reference	25	[¹¹ C]NNC112	Dopamine (D1) receptor	Halldin et al - 1998
2	[¹⁸ F]FDG	Glucose utilization	Kuhl et al - 1976	26	[¹⁸ F]A-85380	Nicotinic acetylcholine receptors	Horti et al - 1998
3	[¹¹ C]methionine	Amino-acid transport	Comar et al - 1976	27	[¹⁸ F]fallypride	Dopamine (D2) receptor	Mukherjee et al - 1999
4	[¹⁵ O]oxygen	Oxygen utilization	Frackowiak et al - 1980	28	[¹¹ C] α -methyl-l-tryptophan	Tryptophan activity	Shoaf et al - 2000
5	[¹⁵ O]water	Blood flow	Frackowiak et al - 1980	29	[¹¹ C]DASB	Serotonin transporter (SERT/5-HTT)	Ginovart et al - 2001
6	[¹¹ C]leucine	Protein synthesis	Barrio et al - 1983	30	[¹¹ C]Ro15-4513	GABA-benzodiazepine receptors	Lingford-Hughes et al - 2002
7	[¹⁸ F]F-DOPA	Dopamine synthesis	Garnett et al - 1983	31	[¹¹ C]temazolamide	Temazolamide pharmacokinetics	Saleem et al - 2003
7	[¹¹ C]methyl-spiperone	Dopamine and serotonin receptors	Wagner et al - 1983	32	[¹⁸ F]SPA-RQ	Neurokinin-1 receptor	Solin et al - 2004
8	[¹¹ C]PK-11195	Peripheral benzodiazepine receptors	Camsome et al - 1984	33	[¹¹ C]PIB	β -Amyloid	Klunk et al - 2004
9	[¹¹ C]diprenorphine	Nonselective opiate receptors	Jones et al - 1985	34	[¹⁸ F]fluoroethyl-l-tyrosine	Brain tumor protein synthesis	Pauleit et al - 2005
10	[¹¹ C]carfentanil	μ -Opioid receptor	Frost et al - 1985	35	[¹⁸ F]fluorothymidine	Brain tumor proliferation	Chen et al - 2005
11	[¹¹ C]flumazenil (FMZ)	Central benzodiazepine receptors	Samson et al - 1985	36	[¹¹ C]harmine	Monoamine oxidase type-A (MAO-A)	Ginovart et al - 2006
12	[¹¹ C]raclopride	Dopamine type 2 (D2) receptor	Ehrin et al - 1985	37	[¹⁸ F]MK-9470	Cannabinoid receptor type 1 (CBR-1)	Burns et al - 2007
13	[¹¹ C]Schering-23390	Dopamine type 1 (D1) receptor	Halldin et al - 1986	38	[¹¹ C]methylreboxetine (MRB)	Norepinephrine transporter (NET)	Logan et al - 2007
14	[¹¹ C]nomifensine	Dopamine transporter (DAT)	Aquilonius et al - 1987	39	[¹¹ C]ABP688	Glutamate receptor 5 (mGluR5)	Ametamey et al - 2007
15	[¹¹ C]deprendyl	Monoamine oxidase type-B (MAO-B)	Fowler et al - 1987	40	[¹⁸ F]fluoromisonidazole	Brain tumor hypoxia	Spence et al - 2008
16	[¹¹ C]McNeil 5652	Serotonin transporter (SERT/5-HTT)	Suchiro et al - 1993	41	[¹¹ C]AZ10419369a	Serotonin 5-HT1B receptor	Pierson et al - 2008
17	[¹¹ C]WAY 100635	Serotonin 5-HT1A receptor	Pike et al - 1994	42	[¹⁸ F]SP-203a	Glutamate receptor 5 (mGluR5)	Brown et al - 2008
18	[¹¹ C]JFBL 457	Dopamine (D2/3) receptors	Halldin et al - 1995	43	[¹⁸ F]galacto-RGD	Brain tumor angiogenesis	Schnell et al - 2009
19	L-1-[¹¹ C]tyrosine	Brain tumor protein synthesis	Willemse et al - 1995	44	[¹¹ C]SB-207145	Serotonin 5-HT4 receptor	Marner et al - 2009
20	[¹¹ C]MDL 100907	Serotonin 5-HT2A receptor	Lundkvist et al - 1996	45	[¹¹ C]GSK189254a	Histamine-3 receptor	Ashworth et al - 2010
21	[¹¹ C] β -CIT-FE	Dopamine transporter (DAT)	Halldin et al - 1996	46	[¹¹ C]P943a	Serotonin 5-HT1B receptor	Gallezot et al - 2010
22	[¹¹ C]JPMP	Acetylcholinesterase (ACE)	Kilbourn et al - 1996	47	[¹¹ C]GSK931145a	Glycine transporter 1 (GlyT1)	Passchier et al - 2010
23	[¹¹ C]verapamil	P-glycoprotein (P-gp) substrate	Elsinga et al - 1996	48	[¹¹ C]GSK215083a	Serotonin 5-HT6 receptor	Parker et al - 2012
24	[¹¹ C]MP4A	Acetylcholinesterase (ACE)	Iyo et al - 1997				

The radiotracers are made up thanks to the synthetization between the radionuclides and molecules that could be freely found inside our body. Different molecules lead to study different pathways. Down below is reported the evolution of the PET scanner through the years.



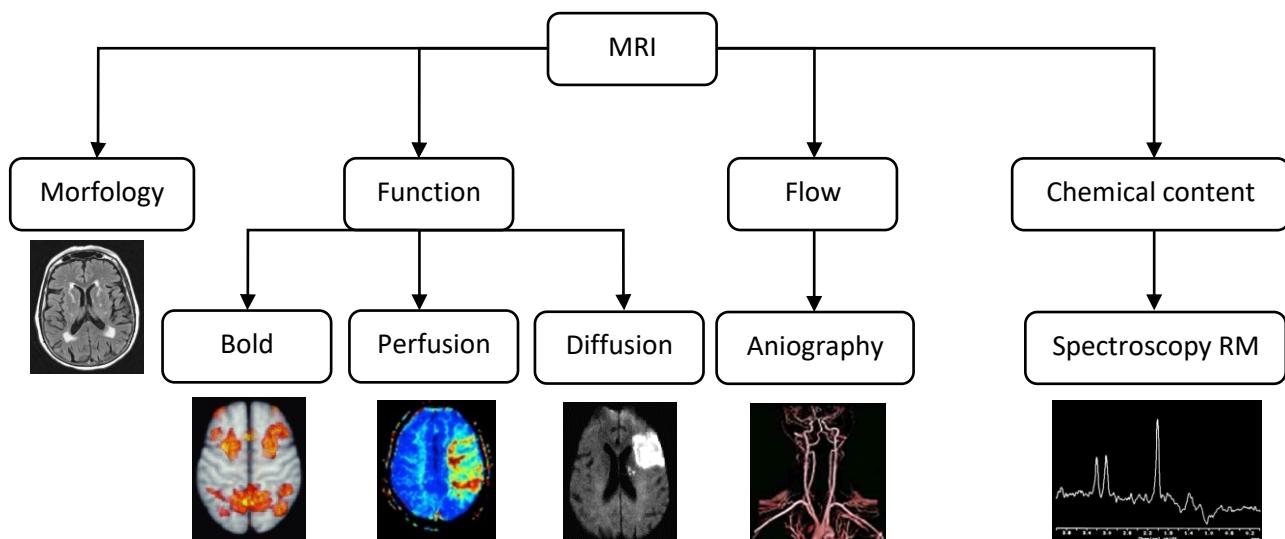
In the images below is reported the comparison between the results obtain from the modern technique and the results obtain with the application of the first prototype of PET scanner. In this last image the shape of the head of the subject was drawn by the researcher's ones they have obtained the results. Moreover, the results show the presence of brain tumour versus a healthy subject.



Nowadays the resolution achieved with the PET technique is of $2 \times 2 \times 2$ or $3 \times 3 \times 3$ mm. Currently are available also hybrid systems like the *PET/CT Scanner* where the functional and the anatomical information are recorded at the same time.

Magnetic resonance

The magnetic resonance imaging is a more complex technique to investigate the brain. It does not investigate the brain from the anatomical point of view but also from the functional, from the chemical content and from the flow point of view. All these are different analysis with different objective.



One of the most particular analysis that could be done with the MRI technique is the brain blood flow studies. With this it can be studied the blood flow in the brain. This methodology has a specific name of angiography. Other important one, rather than the anatomical and functional investigation are the chemical content studies. These ones, thanks to the knowledge of spectroscopy, it can be estimated the chemical content of a certain region of the brain. After having seen all the possible application and utilization of the MR technique we move on how it is made up an MR scan and how it is achieving an MR image. First, by speaking about the MR scanner this is made up of several coil:

- *Main coil* it is responsible for the generation of the static magnetic field.
- *Secondary coils* are used to generate the magnetic field able to move the magnetic dipole from the initial position (referent condition) to a position characterised of a certain flip angle (stimulated condition).
- *Measurement coils*, usually these windings could be the same that stimulate the tissue. In fact, after has delivered the radiofrequency stimulation they move to reading state and they analyse the decay signal generate from all the dipoles inside the body.

The referent system adopt in this technology is the one reported below. In these figures, the one on the right side describe the referent system and the terminology adopt while the image on the left describe how it is applied the static magnetic field from the main coil that produce the polarization of all the magnetic dipoles relative to each molecule and atoms. This magnetic field lead to same orientation of all the magnetic dipoles inside our body. This is a crucial step, which we will describe it later deeply, to have the same referent position and orientation for all the dipoles of our body.

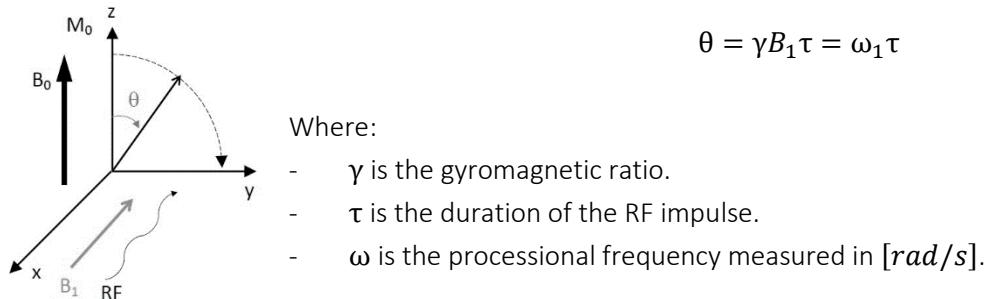


To get some information about the tissue on a specific body region it is needed a radiofrequency stimulation maintain for a certain interval of time. This leads to the changing of orientation of all the dipoles inside the stimulated tissue. This phenomenon exploits one important physics result. All the matter inside the universe from the atomic or molecular point of view present a magnetic moment. When we look on a whole body due to the characteristic of the matter the overall magnetization vector could be considered as null due to the stochastic direction of the magnetic dipoles of all the atoms and molecules. Obviously, the magnetization vector could be considered null if and only if the body analysed is not a magnet. So, this preliminary assumption is verified when we deal with our body. To obtain a magnetization of a volume, which consist in the same orientation of all the small magnetic dipoles, we need to apply a static magnetic field. This is the effect that is exploited first with the application of the static magnetic field in the MRI technique. From the atomic point of view we have that, by consider a little volume, when we apply a magnetic field with a certain direction, also all the dipoles magnetic momentum of the volume will be oriented parallel to it. This concept could be extended for all the volume of our body. This preliminary operation is very important because the

properties that we want to exploit is defined from the magnetization of a tissue after a certain interval of radiofrequency stimulation. Usually, the static magnetic field present an intensity which range from 0.9 T to 11 T. Generally, in clinical application, the intensity of the magnetic field is about 1.5 T to 3 T. Instead, for the research field we could use magnetic field with a magnitude of 3 to 7 T. MRI scanners can be classified by considering the value of the B_0 in:

- Ultrahigh filed (7 T – 11.2 T)
- High filed (1.5 T – 3 T)
- Mid field (0.5 T – 1.4 T)
- Low field (0.2 T – 0.4 T)
- Ultralow field (less than 0.2 T)

The Radio Frequency (*RF*) perturbation is needed to investigate the properties of the specific tissue. This perturbation is needed to change the orientation of all the magnetic dipoles inside the analysed tissue. The orientation is changed usually to get an angle from the initial direction of 90°, this angle is the so called flip angle. For this reason, the radiofrequency perturbation is applied until this condition is achieved from the major quantity of the magnetic momentum. Generally, the application of this magnetic field is maintained for 3 ms. The magnetic field applied to get the perturbation is usually named as B_1 . In the most of cases $B_1 \perp B_0$. B_1 is an oscillating field at Larmor frequency, the magnitude of this field is of 50 mT. The analytical and physical relation that could be applied in this case are the one reported below.



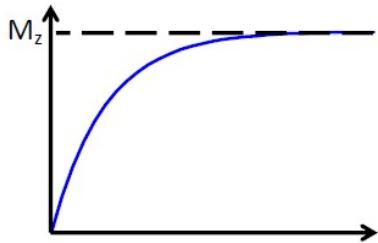
From the analytical relation the flip angle is depending on the duration of the RD pulse and the intensity of the magnetic field B_1 applied. So, to obtain a flip angle of 90° we need to maintain the radiofrequency stimulation for a time defined as $\tau_{\pi/2}$.

$$\theta = \pi/2 = \gamma B_1 \tau_{\pi/2} \Rightarrow \tau_{\pi/2} = \frac{\pi/2}{\gamma B_1}$$

Assuming that the flip angle is equal to $\pi/2$ when the flip is completed, the initial value of M_z is completely assumed in the $x - y$ plane (M_{xy}). The next phase is to read the magnetization in the z direction thanks to appropriate winding. To achieve better results sometimes the RF pulse to apply is not maintain just to obtain a 90° flip angle but to obtain a 180° flip angle.

The main thing that is studied is the time needed from the dipoles to reach again the initial value and direction of the magnetic momentum. In fact, once the radiofrequency stimulation interrupt, immediately all the dipoles turn back to an initial state with a certain delay. The time needed from the dipoles to reach again the original position depends on the tissue analysed. In the MRI studies there are some key time intervals. These ones are described down below.

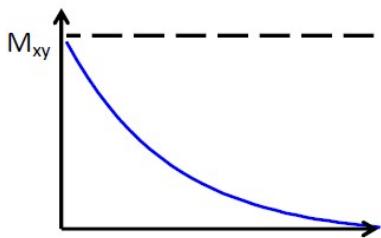
Longitudinal relaxation time (T_1)



T_1 is the longitudinal relaxation time (thermal relaxation time or spin lattice relaxation time). The T_1 relaxation time characterizes the rate at which the M_z component recovers its initial magnetization M_0 .

$$M_z(t) = M_0(1 - e^{-t/T_1})$$

Transverse relaxation time (T_2)



T_2 instead, is also known as the transverse relaxation time (spin – spin relaxation time). The T_2 relaxation time characterizes the rate at which the M_{xy} component decays.

$$M_{xy}(t) = M_0 e^{-t/T_2}$$

As already said before, T_1 and T_2 are tissue specific time.

Free induction decay

The free induction decay is the signal that could be captured after the radiofrequency stimulation. It is due to the spin dephasing (namely, spin – spin interactions and external magnetic field inhomogeneities) by the time the flipped spins diphase slightly and the signal coming from the spins will be slightly less than it was originally. The signal could be described by the next relation:

$$M_{xy}(t) = M_0 e^{-t/T_2^*} \cos(\omega_0 t)$$

In this case it depends on T_2^* and not T_2 . T_2^* depends on the external magnetic field (inhomogeneity of B_0) and spin – spin interaction (T_2). Mathematically:

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \gamma \Delta B$$

Where the $\frac{1}{T}$ is defined as the relaxation rate and it is measured as a frequency, while the value ΔB is defined as the inhomogeneity term.

Repetition time (TR) and Echo Time (TE)

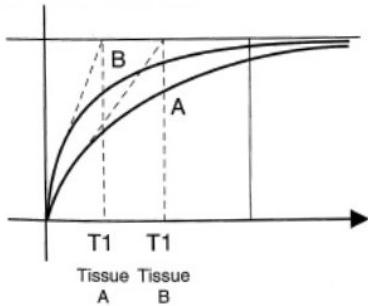
T_R and T_E are defined respectively repetition time and echo time. These two can be controlled and adjusted by the operator. By appropriate setting of T_R and T_E , we can put more weight on T_1 or T_2 depending on the type of clinical application.

Tissue contrast

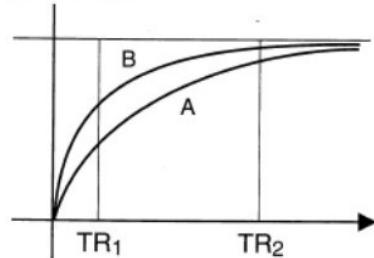
From the general point of view the signal that could be captured from the winding is defined as proportional to:

$$SI \propto M_0(1 - e^{-T_R/T_1})e^{-T_E/T_2^*} \propto N(H)(1 - e^{-T_R/T_1})e^{-T_E/T_2^*}$$

Where $N(H)$ represent the quantity of water inside the tissue. We are looking on the quantity of water because we are studying the free induction decay produced from the atoms of hydrogen and knowing that the main component of water is hydrogen, and the main component of the organism is water. That is the reason why we get a signal proportional to the molecule of water inside a tissue. Moreover, it is important to distinguish different situation by analysing differences in T_E , T_R , T_1 and T_2 .



In the image reported above, represent the free induction decay signal measured for two different tissue denoted with A and B. The tissue A present a lower T_1 than B because we must remember that T_1 or the longitudinal relaxation time is at the denominator of the exponent of the exponential function.

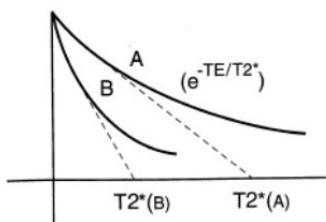


In the image above, instead, are analysed two different tissue A and B. The best time T_R to get the best contrast between these two tissues is the time T_{R_1} . In fact, we can see that the free induction decay signal is used to create the final image and if we read the content of this pattern with a certain time, we could get a good contrast between the different tissue.

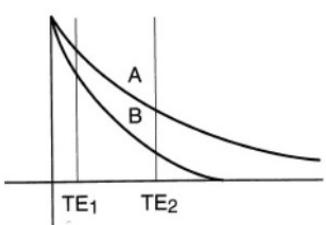
By looking on the formula that describe the signal intensity measured we could see that if we have a very long T_R (repetition time) and so the interval between two different radiofrequency stimulation is very high we could get that mathematically:

$$e^{-T_R/T_1} \rightarrow 0 \Rightarrow 1 - e^{-T_R/T_1} \rightarrow 1 \Rightarrow SI \propto M_0 e^{-T_E/T_2}$$

The signal detected will be more dependent on T_2 instead of T_1 . On the other hand, if the value of T_R is very low and so two different radiofrequency stimulations are very close, we will have a signal which is quite more dependent on T_1 than before.



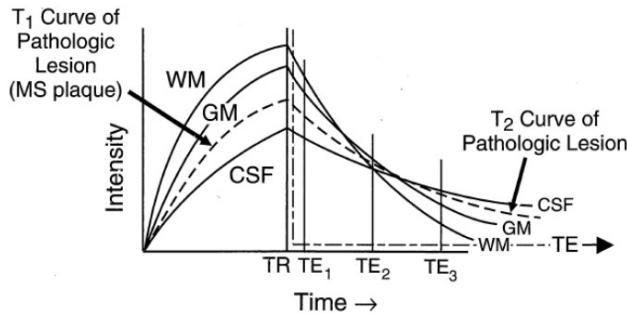
In the image reported above are compared two tissue by looking on the relaxation of the magnetic vector M_{xy} . In this case the T_2^* greater is the one of the tissues with a lower slope. This means that the transversal relaxation of the tissue A instead of B is much slower.



In the image reported above, if we want to inspect and find the image that represent the best contrast between all the tissue by just looking on the curve obtain from the analysis of the M_{xy} proportional signal decay we need to select the echo time T_{E_1} .

In this case if the repetition time T_R is very long so the effect of T_1 could be neglected and if we select T_E long enough the effect of T_2^* on the free induction decay signal revealed is higher. Instead, if we choose T_R very short to enhance the effect of T_1 relaxation instead of T_2^* we need to select a value of T_E which is short for our application. In this way we can enhance the contribution of T_1 instead of T_2 to obtain T_1 and T_2 weighted MRI images.

In the figure below reported we have included a T_1 recovery curve and a T_2 decay curve for a pathological lesion. If we are looking for Multiple Sclerosis (MS) plaques, let's first look at a T_2 weighted image with different value of echo time. Remember from previous paragraph that to get a T_2 weighted image we need to select a T_E which is higher as possible. Moreover, we need also to select T_R very long to decrease the contribute of T_1 decay time.



The most appropriate T_E in this case is not T_{E_3} or T_{E_2} but T_{E_1} . This because when we look to the detected signal choosing $T_E = T_{E_1}$ all the signal coming from the different tissue have the highest contrast. While for the other echo time the different tissue could not been recognized from the image. In fact, if we selected a $T_E = T_{E_3}$ the free induction decay signal of the Cerebral Spinal Fluid (CSF) and MS plaques would not be distinguished. The free induction decay signal presents a maximum value related to the protonic density ρ . The FID oscillating signal decays as a function of 2 time constant T_1 and T_2 . Obviously, we need to pay attention that the previous description of SI represent just the envelop. By using different way to perturb the magnetic dipole, we can weight more the T_1 , T_2 or ρ information of the FID. Down below is reported a table that summarize the different combination of the echo and repetition time usable.

	Short T_E	Long T_E
Short T_R	T_1 weighted	
Long T_R	Proton density	T_2 weighted

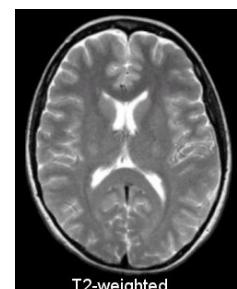
T_1 weighted images

Below are reported an example of T_1 weighted images. By varying T_R , we can introduce varying amounts of T_1 weighting into images.



T_2 weighted images

By varying T_E we can introduce varying amount of T_2 weighting into image.



At 1.5 T the value of $N(H)$ is defined in the different tissue as:

	$N(H)$
White matter	0.61
Gray matter	0.69
Edema	0.86
CSF	1

Down below is reported the different time T_1 and T_2 for the three types of tissues and the image that they made up when it is correctly selected the value of T_R and T_E depending on the type of study that we want to perform.

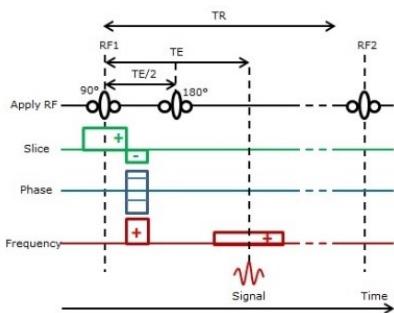
	T_1 [ms]	Image T_1 weighted	T_2 [ms]	Image T_2 weighted
White matter	640		70	
Gray matter	880		80	
CSF	1500		225	

As we can see from the colours associated to all the different tissue the images that will be obtained will show a very high contrast. This is due only to the correct choice of the repetition time and the echo time. One more important concept is that all the relaxations time depend also on the intensity of the static magnetic field B_0 applied to the tissue.

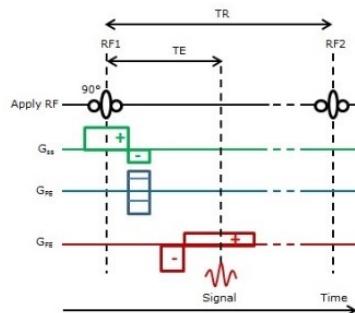
MRI protocols

There are some very important protocols to follow by referring to different image to acquired. The protocols establish how all the stimulation must be applied on our body to achieve a certain result. We must know that the only thing that we have not talk of is how we can select the slice that we want to record. This is achieved thank to the composition of the MRI instrumentation. In the lines before we have talked of secondary coils. These are also called gradient windings. These are three mainly and they introduce different magnetic field gradient to change all the voxel magnetic field associated and consequently also the Lamour resonance frequency. By the usage of a RF stimulation at the Lamour frequency we can excite a specific voxel and study deeper its content. The methods and the timing in the application of all the magnetic gradients create different protocols that are listed below:

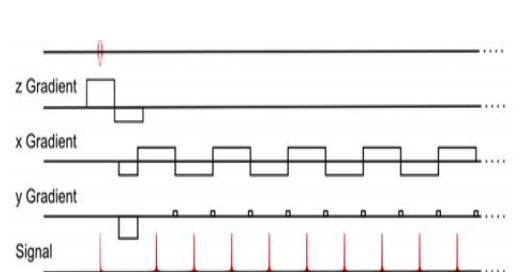
- Spin echo



- Gradient echo



- Echo planar imaging



```
1 %
2 % Title: Laboratory 1
3 %
4 % Note: Loading and handling of DICOM and NIFTY images
5 %
6 % Author: Matteo Martin
7
8 % FILES:
9 % G1_06_dicom: Contain a lot of DICOM files. It is an original dataset
10 % that comes from Pisa. It is an old dataset. This is quite
11 % unusual dataset because are data that come directly from
12 % the scanner. This organization does not follow Bits.
13 % All the files are labeled with numbers. Every file is
14 % representing an image obtain with MRI scanner (information
15 % about CBF). Each of the file represent a slice. Dynamic
16 % susceptibility contrast image.
17
18 % FILES:
19 % T1.nii.gz : File which is very complex and modified version of the Nifty
20 % file. T1 weighted image.
21 % FLAIR.nii : Nifty file of T2 weighted image.
22
23 % FILES:
24 % datiRaw : 4D object of size (128 128 12 48), the last two
25 % dimension are related to the number of slices and the
26 % different collection instant. To check which one is
27 % related with the instant of time we have visualize the
28 % data. In this case 48 is related to the time.
29 % frameBolo : Bolus given.
30
31 % CONVENTION:
32 % In the field of neuroscience there are two main convention of a brain
33 % slice:
34 %
35 % Radiological view: See the brain with direction starting from the legs
36 % and towards the head. In this case the right is the left.
37 %
38 % Neurological view: See the brain from the top of the head towards the
39 % feet of the subject. In this situation the right is on the right.
40
41 %% IMAGE - LOADING (SCANNER RAW DATA)
42
43 % DICOM - LOAD: Read of the image of a single file
44 data1 = double(dicomread('1323994'));
45
46 % All the data upload are represented in double format.
47 % DICOM format is a mix of a lot information and so it may not be just in
48 % double. To access all the information we need to use other function.
49
50 % DICOM - INFO: Read all the information contain into the DICOM file
51 info1 = dicominfo('1323994');
52
53 % What we obtain is a structure of information. Usually the most important
54 % information are the implementation version name (version of the scanner),
```

```
55 % study date stand for the date when an image is analysed.  
56 % We can retrieved different information. All the information inside the  
57 % information section of the DICOM file about the patient and also the  
58 % clinician are hidden in order to respect the privacy law.  
59 % Slice thickness is defined as the thinkness of the slice taken. The  
60 % discretization of the image, one dimension of the voxel that made up the  
61 % image is defined by this element. With the image loaded we have slice  
62 % thickness value equal to 4.8 mm and this means that the voxel could not  
63 % refer to just one tissue. Other important information are the repetition  
64 % time and the echo time (TR and TE). This information could be used in the  
65 % model that we have seen or not. NumberOfAverages is the number of  
66 % sequence of acquisition applied in order to increse the SNR of the image.  
67 % The transmitcoilname is the name of the coil used to deliver the  
68 % radiofrequency pulse. FlipAngle is also contain in this structure.  
69 % SAR represent the amount of electromagnetic waves caught from the  
70 % subject. Some other fields are private and are contain not always  
71 % viewable information.  
72  
73 % DICOM - INFO, Access information: Acess information from the info file  
74 info1.PatientAge;  
75  
76 % REMARK!  
77 % MR data does not have specific unit of measurement.  
78 % PET data has instead specific unit of measruement.  
79  
80 %% IMAGE - LOADING (DICOM)  
81 % CLEAR ALL - Starting a new elaboration  
82 clear all  
83  
84 % DICOM - LOAD: Load of an image T2 weighted  
85 data2 = dicomread('IM_0019');  
86  
87 % The dimension is different from the previous one. It is a T2 weighted  
88 % image. In this case the single file contain the whole volume. It is  
89 % easier to copy. The information that could be used in this case are the  
90 % usual. The main problem come with the Nifty file and not with the DICOM  
91 % one. Inside the Nifty file some information of the DICOM are aviable and  
92 % some not.  
93  
94 %% IMAGE - LOADING (NIFTY) [T2 - w, 1]  
95 % CLEAR ALL - Starting a new elaboration  
96 clear all  
97  
98 % FUNCTIONS:  
99 % load_nii : Load the Nirfty file  
100 % load_untouch_nii : Usual function that is used. It is used this  
101 % beacuse we do not load with this operation  
102 % the image not scaled by scaling factor.  
103 % load_untouch_header_only : Load the header only  
104 % save_nii : Save the file in nifty format  
105  
106 % NIFTY - LOAD: Load of the information from the Nifty file  
107 img = load_untouch_nii('FLAIR.nii');  
108
```

```
109 % The information loaded are contained inside a structure. The numerical
110 % part of the information loaded could be retrieved with the following
111 % statement. In this way we have convert into a readable and representable.
112 mat = double(img.img);
113
114 %% IMAGE - LOADING (NIFTY) [T2 - w, 2]
115 % CLEAR ALL - Starting a new elaboration
116 clear img
117
118 img = load_nii('FLAIR.nii');
119 mat1 = double(img.img);
120
121 %% VISUALIZATION - IMAGE COMPARISON [1] [2]
122 % VISUALIZATION - Visualization of the content of the two matrix
123
124 figure,
125 subplot(1,2,1),imagesc(mat(:,:,25)) , title('UNTOUCH') , colormap 'hot'
126 subplot(1,2,2),imagesc(mat1(:,:,25)), title('CLASSICAL'), colormap 'hot'
127
128 % These two standrd obtain images that are flipped beacuse one is the
129 % radiological version while the other is the nuerologist version.
130 % By the usage of load untouch we dont flip the image. To see if there is
131 % also a scaling factor we should apply the same scale for the two images.
132
133 %% VISUALIZATION - IMAGE COPARISON [1] [2] (AXIAL)
134 % VISUALIZATION AXIAL - Visualization of the content of a lot of image
135 figure,
136
137 for i = 1:50
138     subplot(1,2,1),imagesc(mat(:,:,i)) , title('UNTOUCH') , colormap 'hot'
139     subplot(1,2,2),imagesc(mat1(:,:,i)), title('CLASSICAL'), colormap 'hot'
140     pause
141 end
142
143 %% VISUALIZATION - IMAGE COPARISON [1] [2] (CORONAL)
144 % VISUALIZATION CORONAL - Visualization of the content of a lot of image
145 figure,
146
147 for i = 1:288
148     subplot(1,2,1),imagesc(squeeze(mat(:,i,:))) , title('UNTOUCH') ,
149     colormap 'hot'
150     subplot(1,2,2),imagesc(squeeze(mat1(:,i,:))), title('CLASSICAL'),
151     colormap 'hot'
152     pause
153 end
154
155 % First images are stricly correlated with noise. We are moving from the
156 % ineon to the naision.
157
158 %% VISUALIZATION - IMAGE COPARISON [1] [2] (SAGGITAL)
159 % VISUALIZATION SAGGITAL - Visualization of the content of a lot of image
160 figure,
161
162 for i = 1:50
```

```
163 subplot(1,2,1),imagesc(mat(i,:,:)), title('UNTOUCH') , colormap 'hot'
164 subplot(1,2,2),imagesc(mat1(i,:,:)), title('CLASSICAL'), colormap 'hot'
165 pause
166 end
167
168 %% IMAGE - LOADING (NIFTY) [T1 - w]
169 % CLEAR ALL - Starting a new elaboration
170 clear all
171
172 % NIFTY - LOAD: Load the nifty file with the untouched method the T1.nii.gz
173 img = load_untouch_nii('T1.nii.gz');
174
175 %% IMAGE - LOADING (RAW DATA) [4D OBJ]
176 % CLEAR ALL - Starting a new elaboration
177 clear all
178
179 % RAW - LOAD: Raw file load
180 load('GpreRAW')
181
182 % The objective is to plot the evolution in time. One important thing that
183 % help us to understand when a certain slice is taken we need to have some
184 % information about the repetition time TR (usually get from the DICOM
185 % format).
186
187 TR = 1.56;
188 t = linspace(TR,TR*48,48)';
189
190 % Find out the information about the evolution of the content of a VOXEL
191 % in time. It could be used to find out if something changes during the
192 % experiment.
193
194 timeCourse = squeeze(datiRaw(50,40,6,:));
195
196 % VISUALIZATION - TIME COURSE VISUALIZATION
197 % VISUALIZATION - Visualization of the timeCourse
198
199 figure, plot(t, timeCourse), grid on, grid minor, box on
200 title('Time course'), xlabel('Time [s]'), ylabel('[u]')
201 xlim([0, t(end)])
202
203 % The pattern is very important and is MRI characteristic. The problem in
204 % an analysis like this one is to find out the correct position and instant
205 % of time of analysis.
206
207 % IMAGE - PROCESSING: AVERAGE VOLUME
208 % We can sum all the data to arrive at an average volume. To do so we need to
209 % sum all the elements respect to the fourth dimension.
210
211 volumeSum = sum(datiRaw, 4);
212
213 % VISUALIZATION - Visualization of the 6 th slice
214 figure, imagesc(volumeSum(:,:,6))
215 bw = roipoly;
216
```

217 % *roipoly* : Function that let us to represent a certain brain area. Open
218 % inside the image last opened a command that let us select a
219 % certain area to create a ROI. The output is a matrix of
220 % dimension 128 x 128 of logical value where 1 it is placed over
221 % the location of the area selected.

LESSON 3: MRI, MRI ANATOMICAL INFORMATION, CO REGISTRATION

MRI artifacts

There are many types of artifact that we could record and detect on an image acquired by an MRI system.

- *Image processing artifact*: Aliasing, chemical shift, truncation, partial volume effect.
- *Patient – related artifacts*: Motion artifact, magic angle.
- *Radio frequency (RF) related artifact*: Cross – talk, zipper artifacts, RF feed through, RF noise.
- *External magnetic field artifact*: Magnetic inhomogeneity.
- *Magnetic susceptibility artifacts*: Diamagnetic, paramagnetic, ferromagnetic, and metal.
- *Gradient – related artifacts*: Eddy currents, geometric distortion, and nonlinearity.
- *Flow – related artifacts*.

One of the most important that we have to describe better are the Eddy currents artifact. This is cause from the electrical currents caused by rapidly switching gradients. It appears as a reduced peripheral signal intensity. The correction that has to be apply in this case lead to improve the precision of gradient pulses and use the so – called Eddy current compensation algorithm. When Eddy current compensation or correction is sub – optimal, bright rims can be observed at the edge of the diffusion anisotropy map (as could be seen in the image reported below). With improved Eddy current compensation or correction, this artifact can be effectively reduced.



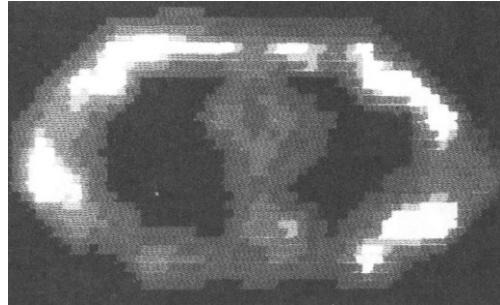
MRI Parameters

The primary parameters are those that are set directly. Down below is reported the list and their main contribute:

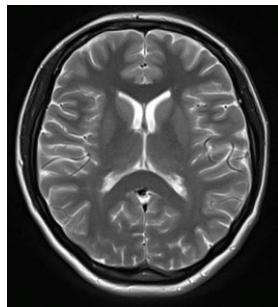
- *Contribute to image contrast*: T_R , T_E , T_I , fa (Flip Angle).
- *Contribute to coverage*: Δz represent the slice thickness or the interslice gap.
- *Contribute to resolution*: FOV_x , FOV_y , N_x , N_y and ***NEX bandwidth***. Where N_x are the number of frequency encoding steps while N_y represent the number of phase encoding steps. Moreover, this category could be also classified under the parameters which mainly contribute to improve the S/N ratio.

To determine the MRI sequence, we need to decide different parameters. The one reported above are not all but just on the bigger set. The first human MR image is the one showed below.

In the image below is reported the first human MR image.



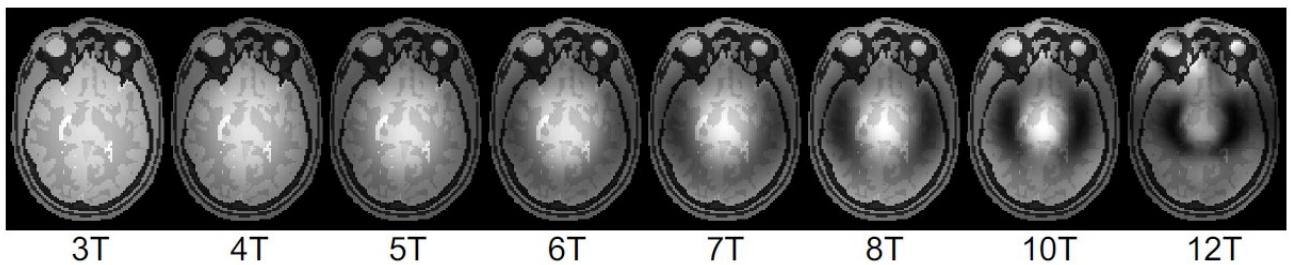
The RF coil was positioned around the chest. The value of B_0 used at that time was of 0.05 T. The technology has been developed a lot from that historical moment, and the quality of the image raises a lot too. The only thing that it is not changed is the phenomenon exploited. The image reported above was obtained in 4 hours. Nowadays, depending on the type of study to be perform, the acquisition time decreases a lot and the results obtainable are the one below reported.



We can see from the previous image, that the resolution increased a lot, in fact now we can even apply segmentation technique on MRI data to, in the case of brain images, recognize the grey and the white matter inside this organ.

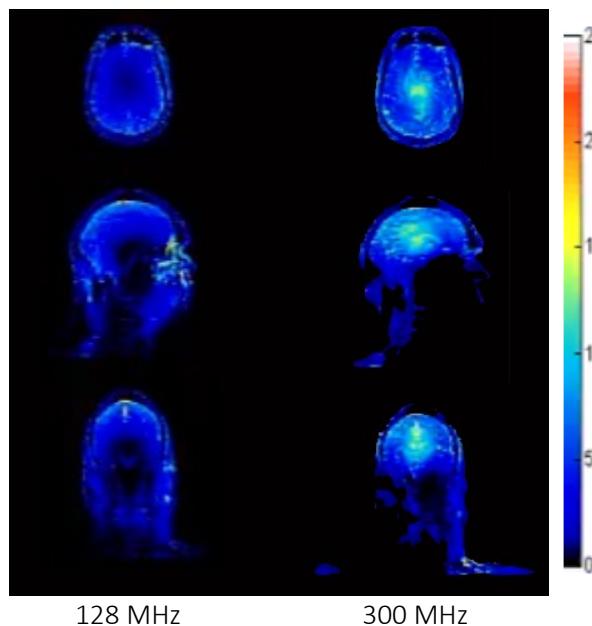
RF inhomogeneity

The price that we pay are inhomogeneities of the RF stimulation when it is activated. Here we have an example of the same slice taken with the same protocol (*gradient echo*) with different magnetic field applied. We can clearly see the growth of some black structure inside the grey matter by increasing the field strength. This change also the interaction between the RF stimulation and the tissue.



SAR

The absorbed power, also known as *SAR* or *Specific Absorption Rate*, is a very important concept to consider when we deal with electromagnetic field generator instrumentation. This represents the quantity of energy absorbed from biological tissue. It is defined in the bioheat equation and it is involved to determine the tissue's temperature after and while a RF stimulation. Generally, this entity is weighted over 10g of tissues. This quantity is regulated from some international law. The organ that set this limitation is the IEEE, that every number of years update its guidelines. During an MRI acquisition, in any case, we could see an increment of the *SAR head distribution*. When the frequency of the RF stimulation increases, the SAR distribution changes too.



In the image reported above the applied static magnetic field is of 7 T, instead, the radiofrequency stimulation changes. In the first case researchers applied a 128 Hz stimulation, while in the second a 300 Hz one. The change of RF used leads to the increase of tissue's temperature in the same position where the SAR is maximum absorbed. Generally, the increment of the temperature is a direct indicator of the possible thermal tissues damage. To limit the maximum temperature because of the head is delicate body's part, IEEE has established that the maximum temperature reachable has to be equal to 38°C. To conclude, when we deal with MR technology, we have to take into account the SAR and the temperature of our head.

Hybrid Imaging system

Hybrid imaging system compose different imaging technique together. One of the most important system combinations is the PET/ MRI one. This is our best in this moment. The combination of the two previous mentioned technology is very important because lead to the possibility of acquiring a functional and an anatomical information at the same time. Obviously, the time needed for the overall process is equal to the highest acquisition time between the two techniques involved. Other important improvement find out by the usage of hybrid system are that it improves the quantitative accuracy of functional/molecular imaging studies, it enables clinical or preclinical studies to be done in less time (two exams done in one session) and

reduce a lot the cost. One of the first hybrid system was made by Philips and its name is *Ingenuity*. They put a PET and an MRI system very close one to each other. It was really challenging to combine the PET system and MRI one. In fact, the magnetic field developed from the MRI instrumentation act on PET system by altering its measures. Caused by this the development of this system takes a lot of time at least 10 years just to solve the compatibility problem. In 2011 Siemens first integrate the MRI and PET in the same structure. Nowadays, the last – born hybrid system is the Signa GE that combine MR and PET together. This instrumentation is very expensive. It produces a magnetic field with an intensity of 3 T to get the MRI information. In Padova we have one of them. By the combination of PET and MR we could perform different types of studies. These are the ones reported below. We could see that we are combining different results obtainable with PET and MR scanners.

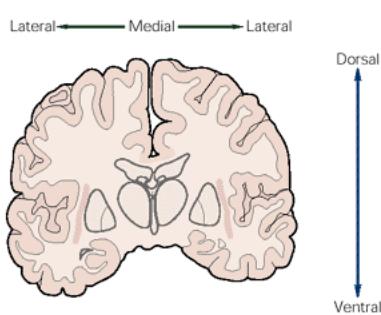
Table 1: Assessment of biological properties by MR and PET.

MR	PET
Morphology	Flow ($H_2^{15}O$)
Water diffusion capacity (DWI)	Metabolism (^{18}F -FDG)
Vascular anatomy (MRA)	Blood volume ($C^{15}O$)
Perfusion (PWI, DCE-MRI)	Oxygen consumption (^{15}O)
Tissue metabolites (MRS)	Hypoxia (^{18}F -MISO)
Functional activation (fMRI)	Vascular permeability (labeled AA)
Cerebral fiber tracts (DTI)	Nuclide acid synthesis (^{18}F -FLT)
Oxygen consumption (^{17}O)	Transmitters (e.g., ^{18}F -DOPA)
Migration of cells (Fe labeling)	Enzymatic activity (e.g., MP4A)
	Angiogenesis (e.g., ^{18}F -RGD)
	Distribution and kinetics of tracers and drugs (labeled compounds)
	Enzymatic activity in transfected cells

Normally when MR is presented in comparison to PET, we have different resolution. In fact, the PET has a lower spatial resolution and a lower temporal resolution than the MRI results and so when the images are post processed the clinician need to consider this important difference. In the table reported below are shown the different MRI and PET exams executable by highlighting the spatial resolution, the acquisition time and the final dimension of the image.

MRI				PET			
T1 3D	~ 4 minutes	512 x 512 x 140	0.7 mm x 0.7 mm x 0.7 mm				
DSC-MRI	~2 minutes	128 x 128 x 25	2 mm x 2mm x 4 mm				
ASL	~ 4 minutes	64 x 64 x 6	4mm x 4mm x 6mm				
fMRI	~20-45 minutes	64 x 64 x 40	4mm x 4mm x 6mm				
				15O	~ 6 minutes	128 x 128 x 40	2 x 2 x 2 mm
				18F	~90 minutes	128 x 128 x 40	2 x 2 x 2 mm
				11C	~ 60 minutes	128 x 128 x 40	1.21x1.21x1.23 mm

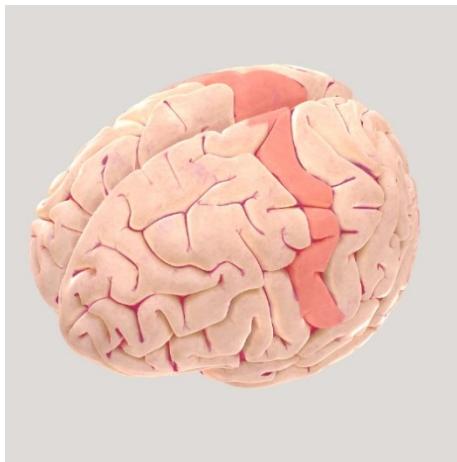
Terminology



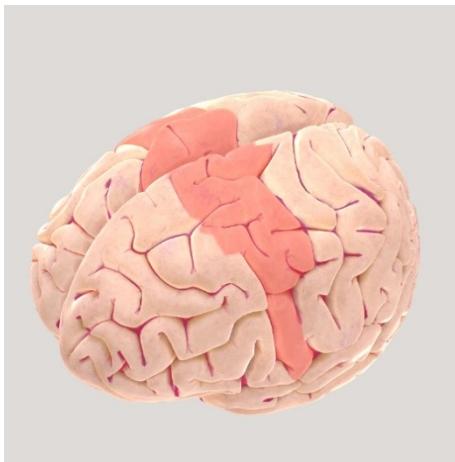
To correctly refers to all the brain's section we need to introduce a specific terminology. First, we introduce some terms to describe where we are looking inside the brain and then we introduce some anatomical brain subdivision.

- *Lateral*: Externals of the brain by looking on a coronal plane.
- *Medial*: Inner region of the brain by looking on a coronal plane.
- *Dorsal*: Upper region of the brain.
- *Ventral*: Lower region of the brain.

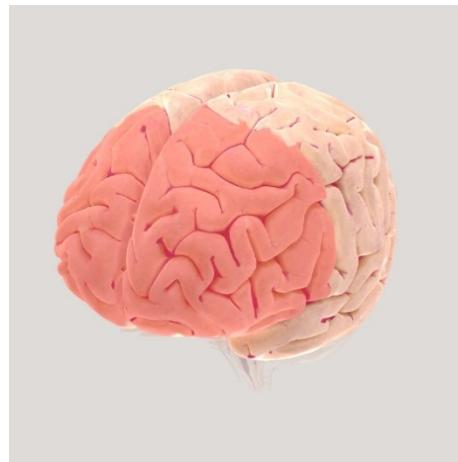
- *Anterior*: Frontal regions of the brain by looking on a sagittal section of it.
- *Posterior*: Back regions of the brain by looking on a sagittal section of it.



Motor Cortex



Premotor Cortex



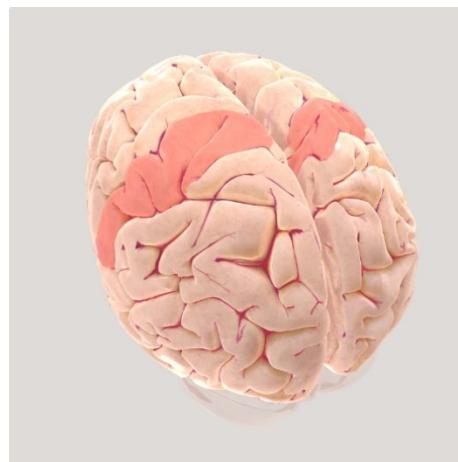
Prefrontal Cortex



Wernicke's Area



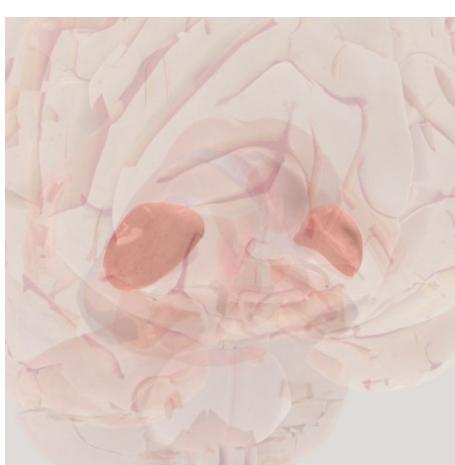
Primary Visual Cortex



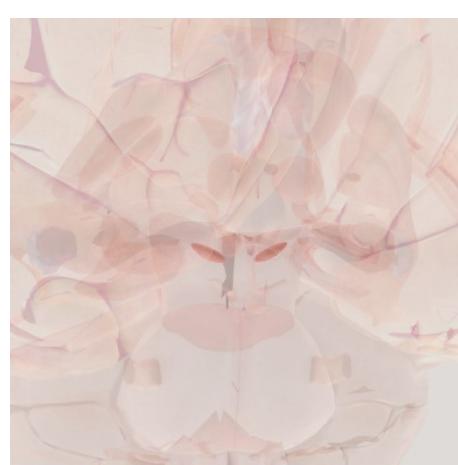
Somatosensory cortex



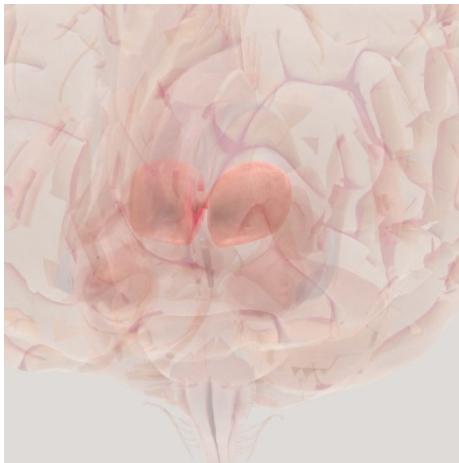
Primary Auditory Cortex



Putamen



Substantia Nigra



Thalamus



Hypothalamus

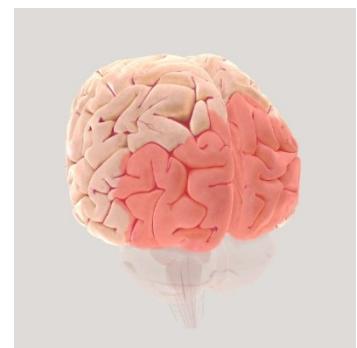
Other important brain's region studied most is the frontal lobe. To understand better the anatomical subdivision of the brain we can introduce the lobe subdivision and their anatomical position:

- Frontal Lobe

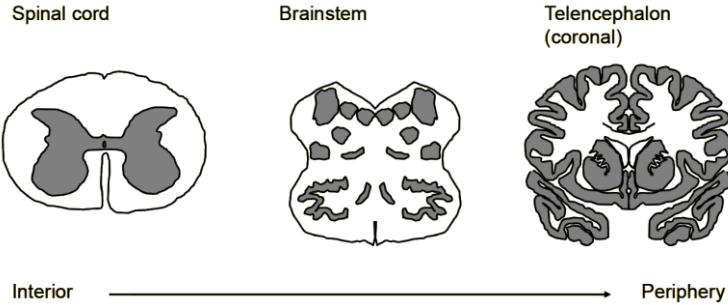
- Parietal Lobe

- Temporal Lobe

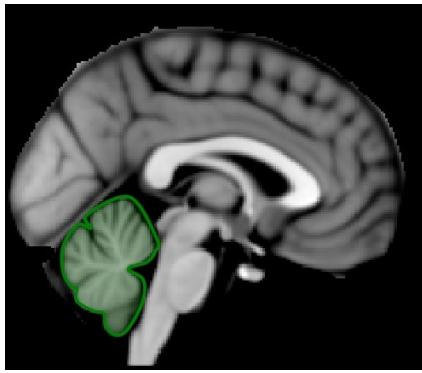
- Occipital Lobe



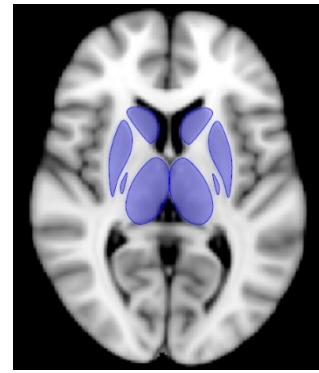
The *CNS (Central Nervous System)* relates to the rest of the body through the spinal cord. This last central nervous system region relates to the brain by the brain stem. The first section of the brain stem is the *Medulla oblongata*, then the *pons* and finally the *midbrain*. The main component of the *CNS* in cells terms are the neurons and the glia cells. The first ones are the basic cells that exchange information between each other. The second ones instead are defined as support cells and nowadays the research fields are still trying to understand their real aim, the only knowledge is that these cells are strictly implied in the major brain diseases, such as ALS (*Amyotrophic Lateral Sclerosis*). Generally, we could subdivide the glia cells in three categories majorly: the astrocytes, the oligodendrocytes, and the microglia cells. In the brain from the tissues point of view, we could define three types of cerebral tissues. The first one is also known as *grey matter*, the *white matter*, and the *cerebrospinal fluid (CSF)*. The grey matter is constituted from the soma of all the neurons while the white matter is made up of the axons and the glia cells that cover and give the lighter colour to all the axons. The cerebrospinal fluid or CSF is a liquid that separates the brain from the skull. Thanks to all the imaging techniques that are nowadays available researchers were able to define a specific white matter and grey matter pattern inside the three main CNS's structure (the *CNS*, the *brainstem*, and the *telencephalon*).



There are two main brain components that are involved in the major modulatory system, the cerebellum, and the basal ganglia. These two systems are the ones reported below:



Cerebellum



Basal Ganglia

Where the basal ganglia are made up of the caudate, the putamen, the globus pallidus and the thalamus. The main function that has been associated to the caudate besides the motor function are the procedural learning, the associative learning and play some role in the inhibitory control of action. The putamen instead is mainly involved in the activity of learning and activity of motor planning. The other structure connects different brain regions. One of the important roles of studying the basal ganglia are to understand the pharmacological diversity, by studying many neurotransmitters and neuromodulators. Moreover, by analysing these components researchers could also find some imbalance linked to psychiatric disorders, in fact, there are two antagonistic pathways the direct and the indirect one, and the imbalance activity between these two branches could lead to disorders of movement and cognition as the Parkinson disease and the Huntington disease.

Processing of the images

We change now perspective, and we try to look to the basic imaging processing techniques. Our aim is to merge results coming from different imaging techniques. To achieve this goal, we need some knowledge about some pre-processing techniques. These are the ones listed below:

- ***Co registration:*** with this term we define the process of overlay between the structural and functional images. By this process we can merge anatomical and functional information.
- ***Normalisation:*** the procedure can warp images to fit to a standard brain's template.
- ***Smoothing:*** elaboration used to increase the signal to noise ratio.

- *Segmentation*: Procedure to slice timing correction and unwarping images.

Co registration

The co registration is based on a simple idea. This one is to take an object and try to fit its physical dimensions to a target object. To achieve these results, we need first to apply some stretches to get object of the same dimension and we can also perform some images rotation. The process of stretches is very important because each subject involved in the analysis have different brain dimensions even if the structure of this organ should be the same. The specific definition of co registration is:

The co registration refers to the process of estimating an optimal spatial transformation that allows to overlap two or more images of the same person taken at different times and in different ways, highlighting different brain content.

Within the same subject when we could collect a T1 weighted image and a 18FDG image and we may not have the same dimension for these two. This could be due to movement of artifact of the person that undergo on the exams but also from the characteristic of the instrumentation used. There are many methods used to co – register which differs by some fundamental choices at an architectural and algorithmic level.

- *Type of geometric deformation between the images that the method tries to estimate.*
- *Similarity measure used to determine how close we are to the expected results.*
- *Interpolation techniques.*

The following subdivision into separate sub problems of co registration has therefore become standard:

- | | |
|---|---|
| <ul style="list-style-type: none"> - <i>Type of transformation</i> - <i>Distance measurement between images</i> | <ul style="list-style-type: none"> - <i>Interpolation</i> - <i>Optimization</i> |
|---|---|

The different co registration techniques are distinguished between those that use transformation based on rigid models (or rigid transformation) and those based on deformable models (or non – rigid transformation). Usually, we choose to consider one template image, and one subject image and we need to match them by just the concept of translation or stretches and translations. One important thing to set initially in our analysis is which images remain fixed and which images could instead change (due to all the previous techniques). In the following of the lesson, we will consider first the rigid transformation model and the non – rigid transformation ones.

Co registration: Rigid transformation

The first algorithm that we can think to apply to different images that comes from different bodies are some rigid transformations. The rigid body model includes just translation and rotation. This type of transformation can be used for example when we have two images of the same subject acquired with the same MR sequence but in two different times (pay attention that the time should not be too different otherwise we could find some brain atrophy artifact) or with different sequences or scanners. In these

transformations the brain does not change shape and size, but it can just rotate and translate. There are two main types of rigid transformation:

- *Translations*: in 2D object the translations could be directed on x or y axes and mathematically they could be summarized as:

$$x_1 = x_0 + t_x$$

$$y_1 = y_0 + t_y$$

In a 3D object instead, we have a third dimension called z and so we can generalize the translations as:

$$x_1 = x_0 + t_x$$

$$y_1 = y_0 + t_y$$

$$z_1 = z_0 + t_z$$

- *Rotations*: The definitions of the rotations are not exactly extendable from the bidimensional condition to the three – dimensional condition because the math operators that need to be used are a little bit more complex. While in the two dimensions we can just use simple sine and cosine linear combination of one angle, that represent the angle of rotation. Down below are reported the analysis separated between 3D and 2D images.

- *2D*: The mathematical expression in this case could be formalized as:

$$x_1 = x_0 \cos \alpha + y_0 \sin \alpha$$

$$y_1 = -x_0 \sin \alpha + y_0 \cos \alpha$$

- *3D yaw*: A yaw is a counterclockwise rotation of about the z – axis. The rotation matrix is given:

$$x_1 = x_0 \cos(\alpha) + y_0 \sin(\alpha) + z_0 0$$

$$y_1 = -x_0 \sin(\alpha) + y_0 \cos(\alpha) + z_0 0$$

$$z_1 = x_0 0 + y_0 0 + z_0$$

- *3D pitch*: A pitch is a counterclockwise rotation of y – axis. The rotation matrix could be obtained from the next equation's system.

$$x_1 = x_0 \cos(\beta) + 0y_0 + z_0 \sin(\beta)$$

$$y_1 = x_0 0 + y_0 + z_0 0$$

$$z_1 = -x_0 \sin(\beta) + y_0 0 + z_0 \cos(\beta)$$

- *3D roll*: This elaboration is a counterclockwise rotation of x – axis. The rotation matrix is given by the system of equations reported below:

$$x_1 = x_0 + y_0 0 + z_0 0$$

$$y_1 = x_0 0 + y_0 \cos(\gamma) - z_0 \sin(\gamma)$$

$$z_1 = x_0 0 + y_0 \sin(\gamma) + z_0 \cos(\gamma)$$

- *Zooms*: Transformation that maintain all the distance between different brain structure of the same image. The only thing that we have to keep in mind is that could be expressed as a scaling operation with the same scaling factor above all the dimensions.

The overall combination of translations and rotations cold be expressed by the usage of a simple three – dimensional vector matrix representation:

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \end{bmatrix} R_{yaw} R_{pitch} R_{roll} \begin{bmatrix} s_x & 0 & 0 & t_x \\ 0 & s_y & 0 & t_y \\ 0 & 0 & s_z & t_z \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}$$

In the most of toolbox implemented in the most quantity of programming languages this matrix is already implemented, and we need just to give all the necessary parameters to apply the transformation to a specific image. It is one of the fastest brain elaboration method. It is strictly use when we need to deal with motion correction.

Co registration: Affine transformation

If we want to match images that come from different subjects, we need to apply the affine transformations. These elaborations are defined as rotation, shear, translation and zoom. The translations, rotations and zooms are the same process applied on the image that we have already described before. Instead, the shear is the new transformation introduced with the affine methods.

- *Shear*: It is stretch and a scale of the image by different quantity in all the principal directions.

Affine transformation takes hours sometimes, and generally we can apply different types of algorithm. In order to solve this pitfall, we have to introduce the argument of optimization.

Co registration: Algorithm optimization

The Algorithm optimization is very important part of the engineering that try to reduce the time needed to complete a certain computational process. The optimization algorithm that are implemented to reduce the time to complete an affine transformation are based on the maximization of the mutual information. This concept will be defined better in the following of the lesson. Generally, this type of algorithm try to maximise the statistical correlation between two images taken in input. To do so, in the algorithm are implemented some cost function based on some estimator's techniques.

Mutual information

Mutual information measures the degree of statistical dependence of two random variables and can be used to measure how well one image explains the other. For two data set $X = x$ and $Y = y$, the mutual information is defined as:

$$I(X, Y) = \sum_{x,y} p(x,y) \ln \frac{p(x,y)}{p(x)p(y)}$$

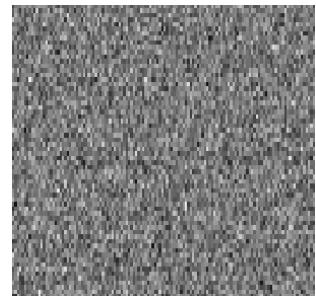
Where $p(x, y)$ is the joint distribution of the pair (x, y) , and $p(x)$ and $p(y)$ are the marginal distribution of x and y . This statistical concept has some properties, which are: it is always non-negative, and it is symmetric. Mutual information is related to the concept of Entropy. To understand well what the mutual information is we need to understand what entropy is.

Entropy

Entropy is a measure of uncertainty, higher the entropy is, more uncertain is the random variable take in consideration. Image entropy is a quantity which is used to describe the amount of information which must be coded for by a compression algorithm. Low entropy images, such as those containing a lot of black sky, have very little contrast and large runs of pixels with the same or similar values. An image that is perfectly flat will have an entropy of zero. Consequently, they can be compressed to a relatively small size. On the other hand, high entropy images such as an image of heavily cratered areas on the moon have a great deal of contrast from one pixel to the next and consequently cannot be compressed as much as low entropy images.



Low entropy



High entropy

Entropy could be used as biomarker instead of variance and standard deviation. If we apply the entropy calculation on a variable which is gaussian we could find out that is a measure strictly correlated with standard deviation. Entropy could be expressed with the next definition:

$$H[x] = - \sum_x p(x) \log_2 p(x)$$

Where the minimum value of $H[x] = 0$ and the maximum value of $H[x] = \log_2 M$ where M represent the number of the possible state of the variable x . The unit of measurements used for this concept is the bit. Consider a random variable x having 8 possible states, each of which is equally likely. The entropy of this variable x is given by:

$$H[x] = 8 \times \left(\frac{1}{8} \log_2 8 \right) = 3 \text{ bits}$$

Entropy can be interpreted as a measure of disorder, the maximum entropy configuration corresponds to an equal uniform distribution of probabilities across the possible states of the variable.

LESSON 4: CO REGISTRATION AND SEGMENTATION

Mutual information and entropy

There is a strong connection between the definition of entropy and mutual information. In fact, known that mutual information could be written as:

$$I(X, Y) = \sum_{x,y} p(x, y) \ln \left(\frac{p(x, y)}{p(x)p(y)} \right)$$

Where $p_{XY}(x, y)$ is the joint distribution of the intensity pair (x, y) and $p_X(x)$ and $p_Y(y)$ are the marginal distribution of x and y . Entropy is a measure instead of uncertainty, higher entropy is more uncertain is the prediction of a random variable. Generally, the mutual information I and the entropy H could mixed up as:

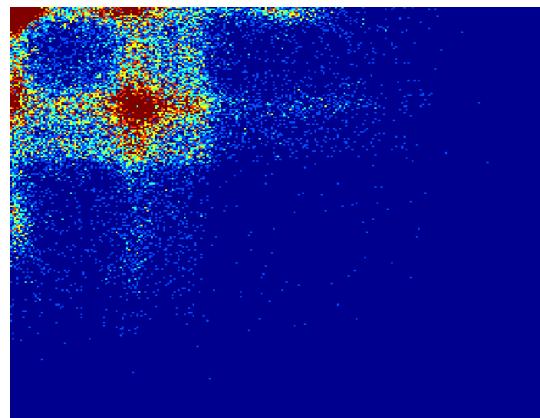
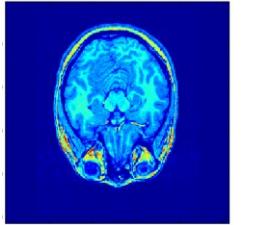
$$I(X, Y) = H(X) + H(Y) - H(X, Y)$$

Where $H(X, Y)$ represent the entropy between the two random variable and is also known as the joint entropy:

$$H(X, Y) = - \sum_x \sum_y p(x, y) \ln p(x, y)$$

Mutual information algorithm

We take into account two different images and we want to calculate the mutual information (I) of an image. Down below are reported the usual step that has to be implemented into an algorithm. The first two ones consist in retrieve the dimension of the image analysed and compute the joint intensity histogram. To do this last operation we need to allocate an $N_X \times N_Y$ array for evaluate this quantity correctly. Where the variable N_X and N_Y represents respectively the number of grey levels inside the image X and the image Y . The joint histogram represents a chart where the row indicates the level of grey inside the image X while the column is the respective but for the image Y . So, if we analyse an element of this matrix in position (x, y) where the first indicate the level of grey inside the X image and second the level of grey inside Y , it contains the number of voxel combination between the two images that have the level of the voxel of the first image equal to x and the second voxel image equal to y .



Therefore, the mutual information measures the amount of information that one image contains about the other:

- If X and Y are completely independent, $I(X, Y) = 0$.
- If X and Y are identical, $H(Y) = H(X)$ and $I(X, Y) = H(Y) = H(X)$ and assume the maximum value.

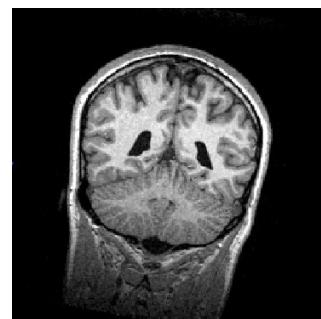
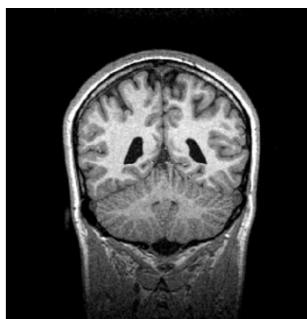
We may use the mutual information as measure to register images. This because we can think to use it to determine the similarity between the two images. For this reason, mutual information could be used as a cost function to apply in a specific case. In fact, in neuroscience, in the application of co registration technique it is largely exploited.

Kullback – Leibler divergence

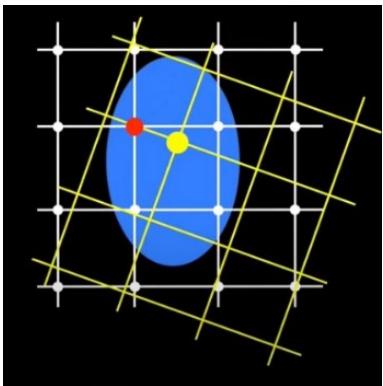
Divergence measures play an important role in measuring the distance between two probability distribution functions. It is really the expression of mutual information. Kullback – Leibler divergence function is a popular measure in this class. It is a measure of how much one probability distribution is different from another probability distribution. It can be exploited as a cost function in co registration algorithm. For a jointly discrete or jointly continuous pairs (X, Y) , mutual information is the Kullback – Liebler divergence of the product of the marginal distribution $p(x)$ and $p(y)$ from the joint distribution $p(x, y)$. To conclude, in general, the Kullback – Leibler divergence is a measure of how one probability distribution is different from a second, reference probability distribution.

Co registration: Interpolation

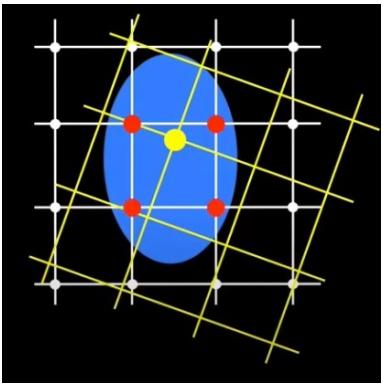
When we move an image from one position to another, we are referring to a process called interpolation. The interpolation process is a part of the co registration techniques. This step is very important because once we select the type of movement that has to undergo the image, we need to recalculate the value contained inside all the voxels. Down below is reported an example of image rotation of 10° respect the z – axis.



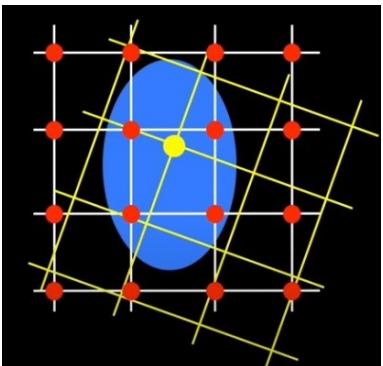
Once the transformation is applied, we need to recalculate respect the new grid the value contained inside each voxel of the image. There are different methods that could be applied to find out the correct results. To summarize all the concepts the interpolation step is needed to find the intensity values between grid points. Down below are reported the main methods usable in our case.



- *Nearest neighbour:* It is an algorithm that performs the interpolation based on the level of grey contained inside the nearest neighbour. In the image reported on the left side of the page we have a red dot and a yellow dot. Let's suppose that the yellow grid is the one obtained from the white one by rotating it. In this situation the yellow dot is the one that has to be estimated. Through the nearest neighbour algorithm, we need first to identify the nearest neighbour, highlighted with the red colour and then we need to copy its value in the yellow one. This algorithm is very fast but has some problems with the edges.

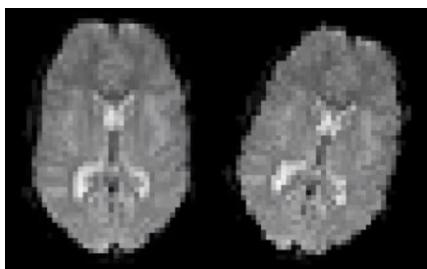


- *Trilinear:* In the trilinear algorithm or bilinear algorithm if we are in 2D, we identify the nearest neighbours of the new dot in the previous grid and we associate to it the weighted sum of all the values contained in these identified points. The weight depends on the distance, much more an old point is close to the new point much more is the incidence. The weighting function is a linear function, and this is the reason why the algorithm is called linear. It is fast but can blur the image a little bit.

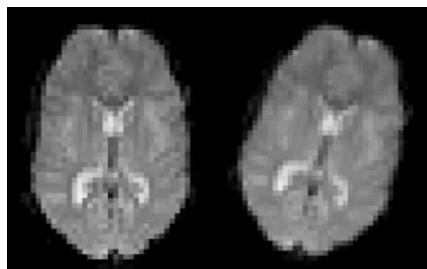


- *Spline, sinc and k – space methods:* These are other algorithms that involve in the elaboration the overall number of points inside the initial grid to calculate the new value of a certain point in the new grid. These algorithms create sharp images thanks to the analysis inside a very large set of points. The pitfalls of this technique are that they can create values outside the original range.

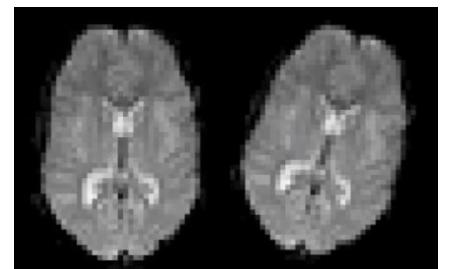
Down below are reported the results obtained from the application of the different algorithms.



Nearest neighbour: As already said before, it creates worsen edges because it is based on the closest point of the initial grid.



Trilinear: The trilinear algorithm blurs a little the image, this could be a con in those applications where we want to keep a certain level of details.



Spline: Spline is an algorithm less fast but generates a much more accurate result.

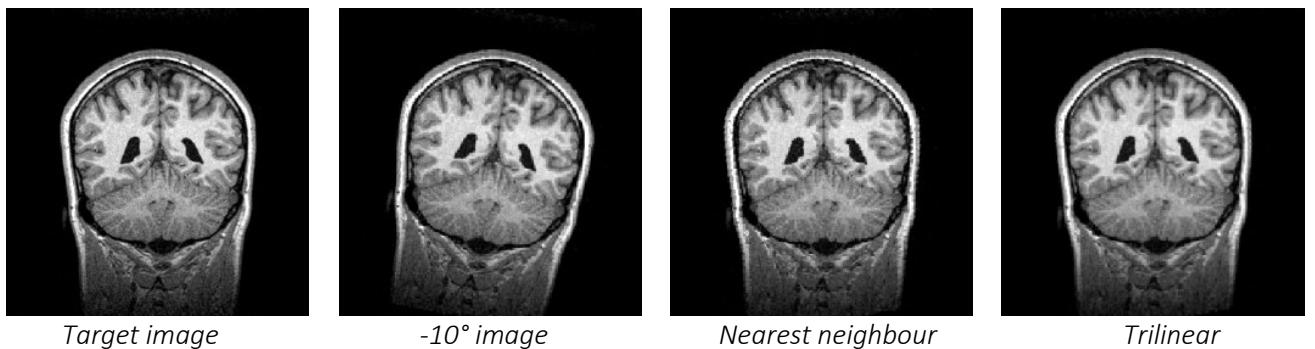
The algorithm that has to be applied depend on the type of analysis that we need to perform. One important concept that we need to hold in mind is that the algorithm can also reduce or amplify the size of artefacts. Mainly, nearest neighbour contains the value of artifacts inside a certain range while spline could also amplify it. Generally, to apply a transformation we need to follow some strict steps. These are:

- *Estimating the transformation that it is needed to apply on the image.* This step consists just in find the most appropriate transformation to apply in our situation and there is not any resampling stage.
- *Resampling.* The next step consists in apply the previous identify transformation to our image and thus creating a new modified image. For this step we need to apply the interpolation algorithm chosen. When this step is done, we have always to worry about the quality of the image in output.

To indicate the process of finding the transformation, estimate it and apply it we can use the term *registration, co – registration* or moreover *spatial normalisation*.

In order to avoid the loss of quality of the image we need to delay the resampling and one more important fact is that we need to avoid multiple resampling and so in the most of pipeline where are needed multiple transformation we combine them all together and just finally we perform a resampling step.

Now we come back on our analysis and we want to highlight the main differences between the results of resampling step by the application of different interpolation paradigm.



Spline algorithm for resampling stage is the largely used and is set as default in software like FSL and SPM.

Co registration: Aim

Just to summarize what we want to do we introduce the next set of steps. First, we need some measurements data, like MRI images, PET data or fMRI ones and we need also a key element defined as the template. This last concept is represented as the standard reference image to be able to compare the results obtain from the different analysis. Generally, we need to remember that the types of analysis performable are mainly two. The first one is where we have to take different measures from the same subject. In this case we need a reference which is usually chosen between a T1 weighted or a T2 weighted image. Once, selected the template, on all the other images, we have to be applied some co registration technique. Otherwise, in the second case, we may have a lot of data that come from different patients. In this situation we need to have a template which usually is represented by a default template called *MNI image* and apply a co – registration algorithm on all the data obtained. Obviously, one important thing to highlight that is always true is that closer is an image to the template easier is the transformation that is needed to be

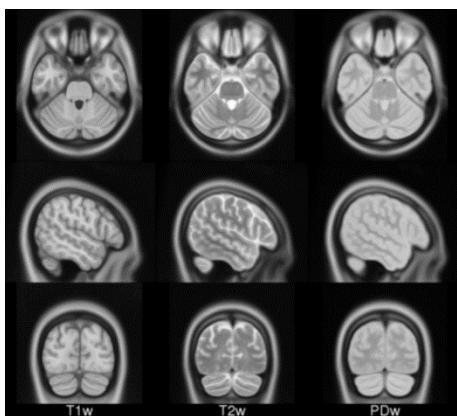
applied. Another important information is that all the software developed are specialized on specific transformation, and so depending on the type of these last ones we could use the most suitable software for our application.

Co registration: Talairach template

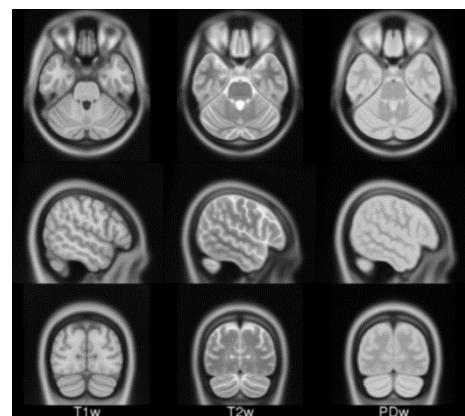
In the previous paragraph we have talked about template that need to be correctly selected to perform our analysis in some cases. The human brain has a morphology that differs between subjects. In mean, the brain of 2 subjects differs of about the 30%. Brain volumes of different subjects have to be normalized in the same 3D space, known also as stereotactic space or Atlas or Talairach. This last one represents coordinates, also known as Talairach space. It is a three – dimensional coordinate system (known as an *atlas*) of the human brain, which is used to map the location of brain structures independent from individual differences in the size and overall shape of the brain. Generally, the Talairach has a characteristic position in the brain. The normalization is based on the front commissure, where with the term commissure we refer to the section of fibres connecting the two hemispheres. There is the posterior and anterior commissure. The front one is the easiest to recognize. The values measure the distance in meters from the origin. The Talairach atlas has been obtained by dissecting and photographing the brain of a single subject (an elderly woman). This is why nowadays is not largely more used. The origin in the Talairach coordinate system is made up of one standard point as we already mentioned before.

Co registration: MNI template

Alternatively, we can use the atlas created by the Montreal Neurological Institute (*MNI template*). The first *MNI* atlas is called *MNI305* and it is made up on all right handed subjects. The most important thing of this model is that it is widely use instead of just the Talairach three – dimensional space. This reference is the results coming out from an analysis of 305 subjects, female and male of age between 19 and 30. Since, it was based on younger population, the brain model is larger than the 10 – 15% respect to the Talairach one. Currently the standard *MNI* template is the *ICBM152 2009b*. This is one of the most used, but it is not the only one nowadays available. Even when we talk about the *MNI* template we need to understand which template we have to apply on a specific situation. Down below are reported the two latest *MNI* template:



ICBM 2009b Nonlinear Asymmetric 0.5 x 0.5 x 0.5 mm template



ICBM 2009b Nonlinear Symmetric 0.5 x 0.5 x 0.5 mm template

We can see clearly the differences between the two models. Generally, we select the Symmetric model, but obviously depend on the types of analysis that we need to perform. When we deal with *fMRI* we choose the Asymmetric MNI template.

Co registration: Template advantages

The advantages of using a standard space are the one listed below:

- It allows the comparison between subjects; it facilitates comparison with results from other studies.
- It facilitates comparison with results from other studies.
- It allows to specify brain areas through simple coordinates.
- It allows medium operations between different subjects.

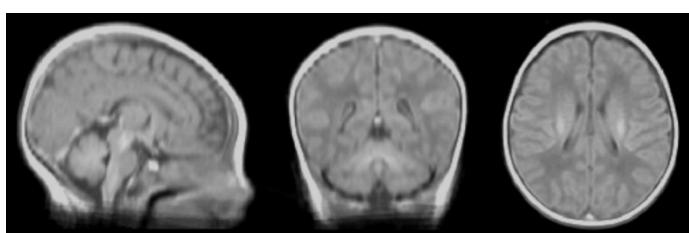
Co registration: Template disadvantages

The most common disadvantages instead of using a template are:

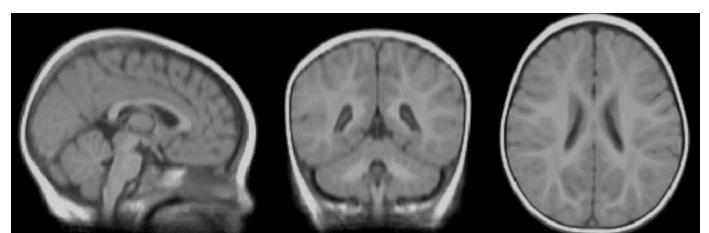
- We need some affine co registration technique to apply on the different results to compare them.
- Affine transformation is no longer needed, instead we need some non – linear transformation.
- It is really time consuming for the heavy elaboration that we need to perform.
- Depending on the characteristics of the brain to be normalized the algorithms used can give not optimal results.

Co registration: Ages and templates

It is very important also to analyse the differences between people's brain of different ages. In children the white matter is not surrounded by myeline and so we cannot recognize properly the white matter. In people with age between 8 and 11 the contrast increases thanks to the development of the myeline. Then by taking in consideration a template of an adult of age between 44 and 60 we can clearly see the subdivision in white and grey matter. In this last template the regions are well delineated. All the templates are based on T1w images because is the image that needs the least number of voxels, but the structures are still well defined.



0 – 2 Years



8 – 11 Years



44 – 60 Years

Segmentation

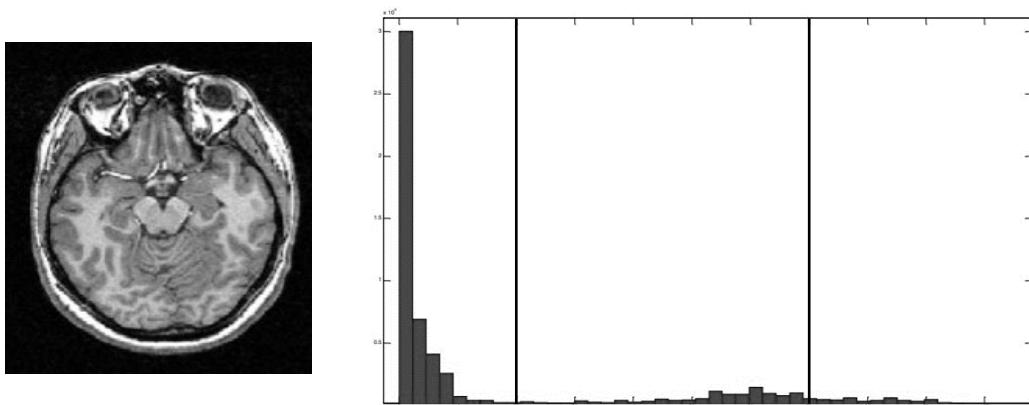
The main definition of segmentations is to divide up the image into patchwork of regions, each of which is homogeneous, that is the same in some sense as instance the intensity, the texture the colour etcetera. Generally, we can separate the information and from this separated image we want to create maps. It is an important step to integrate in the model and to enrich our statistical analysis. To correctly perform these procedures, we have some main methods to apply. These ones are:

- *Global information*, this step could be done by applying thresholding algorithm, cluster algorithm or classification algorithm.
- *Identification of edges*. The next step of having applied the segmentation technique, we need to apply different filtering phase, we may also apply some tracing and fit methods.
- *Local information*. From the previous analysis we may also perform some local information analysis such as split and merge algorithm or apply some watershed paradigm.
- *Deformable models*. To enrich our data, we may also create some models to determine the deformability of the brain structure considered.

In the general case we have more correct procedure to follow to obtain some good results when we want to apply segmentation.

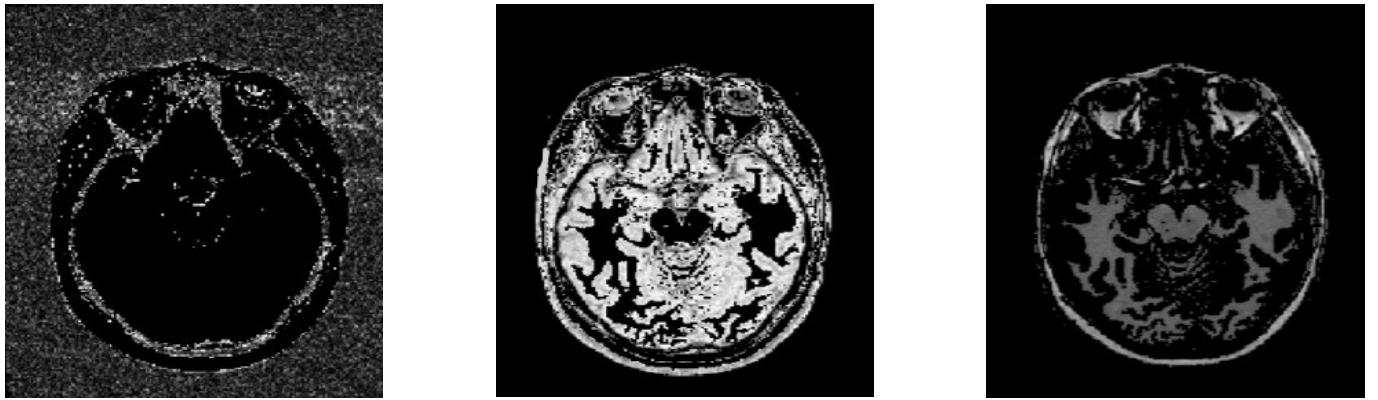
Segmentation: Concepts

We want to separate the information from a data matrix. If we want to achieve this result, we can use some thresholds T to separate the different tissue that made up the brain. By consider an example where we deal with a *T1w MRI* image we can represent the histogram of the information of the brain volume considered:



We can see from the image reported above that the *T1 MRI* image histogram present a sort of two distributions. The first one contains the voxels characterised of a low grey scale intensity while the second contains all the voxels defined with a high grey level. To separate the two distribution, we need to use the threshold T_1 and T_2 set up like in the image shows above.

The results obtained if we apply the thresholds selected are the next three images. The first one defines mainly the borders between the head and outside the head. Instead, the second one seems like it is representing the grey matter and finally, the image obtained with the grey level greater than the second threshold represents the white matter.

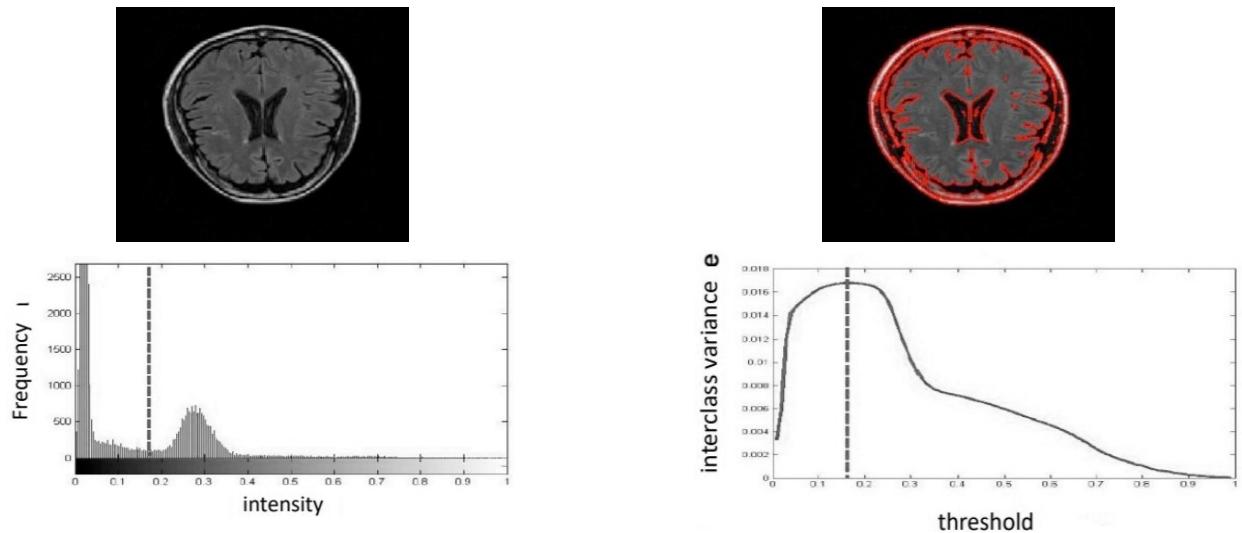


Segmentation: Thresholding

There is an algorithm that can find the optimal threshold. It assumes there are only two classes in the image. The optimal threshold for separating these two classes R_0 and R_1 can be obtained by minimizing intra-class variance σ_w^2 . This algorithm is the so-called *Otsu's method* first developed in the 1979. The equations on which is based this last algorithm are:

$$T_{OPT} = \max_I[\sigma_b^2] = \max_T[w_1 w_2 (\mu_1 - \mu_2)^2]$$

In the image reported below is reported the application of this algorithm in a real case. We can see from the images that it is not the best approach to use because in the brain we have a lot of classes to distinct between.



```
1 %
2 %     Title: Laboratory 2
3 %
4 %     Note:
5 %     Co registration - Segmentation - Dynamic PET activity visualization
6 %
7 % Author: Matteo Martin
8
9 % THEORY -----
10 %
11 % We will see in this lessons some application of segmentation and co
12 % registration. The data that we have in our disposition in this case are
13 % organized in the following structure:
14 %
15 %     - Atlas folder: Contain all the atlas information.
16 %         - atlante_2mm.nii : Nitfy image representative of brain regional
17 %                               subdivision.
18 %         - MNI152_T1_2mm.nii : MNI template 152 2009b
19 %
20 %     - DatiMR folder: Contains all the information about the magnetic
21 %                       resonance performed on the analysed subject.
22 %         - T1_2mm.nii : T1 weighted image of the subject
23 %
24 %     - DatiPET folder: It contain the PET data get from the patients.
25 %         - PET.img : 4D object of the patient PET exams
26 %         - sumPET.img : 3D object of the overall PET process
27 %
28 % Generally T1w has higher resolution than T2w. This last one are the
29 % principal image used for any antomical analysis. In this lab the main
30 % reference would be the T1w image.
31 %
32 %
33 % SEGMENTATION -----
34 %
35 % We start this lab by performing segmentation on T1w image. In literature
36 % there are several software that can perform segmentation. Our analysis
37 % will be done by the usage of SPM (Statistical Parametric Mapping).
38 % The data that we will use during the laboratory could be
39 % visualized also in FSLeyes which is a software developped from Oxford
40 % University and is quite more performant. Normally when we display
41 % anatomical images we choose to use the grey scale, functional images
42 % instead are represented with colored scale. To summarize the colormap
43 % used in different cases we could write:
44 %
45 %     - Anatomical map: Grey scale
46 %     - Functional map: Coloured scale
47 %
48 % ATLAS
49 % In this lab we will use also an Atlas provided from the MNI. This is
50 % a representation of a reference brain structure. It could also contain
51 % some regional subdivision. The reason of using an Atlas is because it is
52 % a common way to have a widely accept subdivision of all the brain area.
53 % In our analysis we will use this image to define the region of interest
54 % ROI. Usually Atlas is provided also with some PDF files that are aviable
```

55 % in the same folder of the Atlas data which defines all the brain regions
56 % presented inside the 3D model.
57 %
58 % SPM
59 % SPM is a software widely used for analysing fMRI data, PET and EEG.
60 % Even if this last data are largely analysed with EEGLAB, which is a
61 % specific software for EEG analysis. In literature there are not a lot of
62 % Toolbox to perform PET analysis, but SPM is one of this.
63 % To start with our elaboration and studies we need to invoke the command
64 % "spm" inside the command window and it will pop up a new window
65 % containing a GUI to interact with of SPM software. We then have to click on
66 % "fMRI" button and, in our case that we need to perform a segmentation
67 % analysis, we need to rightclick on "Segment". It will be popped up a new
68 % window where we need to specify the image that need to be segmented.
69 % There are a lot of prior information and are used a lot of template
70 % already pre installed to perform correctly this analysis.
71 % The prior information are very important to distinguish between
72 % different patients state. In fact, these are usually changed depending on
73 % the types of patient that we analyse. In the "Data" section we need to
74 % specify inside "Volume" field the image to segment. When all these
75 % parameters are correctly set up we can press the "run" button in the
76 % left upper corner of the window represented with a green start button.
77 % The process then starts and takes a while to complete correctly the
78 % elaboration. In the left lower panel it will be represented the
79 % likelihood function that is generally calculated from the program. This
80 % function represent the state of the process. In fact in the x axis is
81 % represented the number of iteration while in the y axis the likelihood.
82 % Once the process is ended all the data will be found inside the same
83 % selected file folder.
84 %
85 % RESULTS
86 % After the data analysis has been completed we inspect all the results.
87 % To display the results we can use the SPM "Display" button. Once clicked
88 % this button we could go inside and specify the image to visualize. To
89 % move forward we need to click done and it will appear a window containing
90 % the specified image. Generally, the image is 3D and so are represented
91 % the three main brain sections (Coronal, Sagittal, Axial).
92 % This visualization procedure is less usable than the one provided with
93 % FSL but it is still a very good alternative.
94 % Once the process has been completed we can move the cursor over the three
95 % images in order to find out the intensity of a specific voxel or region.
96 % The values found usually are contained inside the interval of [0,1].
97 % If the cursor is moved all the three views of brain section are updated.
98 % Down below are reported the results obtained from the segmentation and
99 % their content.
100 % - Image 1 - Gray matter
101 % - Image 2 - White matter
102 % - Image 3 - CSF
103 % - Image 4 - Scalp
104 % - Image 5 - Scalp with soft tissue not part of the brain
105 %
106 % PRE PROCESSING
107 % Every time we need to do some pre processing analysis, it is quite
108 % useful to perform a segmentation analysis because we can accept or not the

163 % of the file computed.
164 %
165 % PET, sumPET -> T1w
166 % The same procedure has to be followed for PET and sumPET data. We choose
167 % to pass as "Source image" sumPET and as "Other images" PET data.
168 % Generally when we deal with PET data we could achieve some good results
169 % and interpretation of them if and only if the studies of PET performed
170 % do not involve any receptor studies.
171 %
172 % CO REGISTRATION INFORMATION
173 % Additional information about the co registration algorithm could be
174 % found in the last section of the window. Deeply, we could change the
175 % basic cost function paradigm. The default one is based on the mutual
176 % information concept. One more parameter editable is the interpolation
177 % algorithm used. In this cases we will use B spline or nearest neighbour
178 % algorithm. To perform correctly the analysis and in general when we
179 % choose to move an atlas on another reference system we need to apply
180 % the nearest neighbour paradigm to keep the same number results of
181 % anatomical parcellation of the atlas.
182 %
183 % RESULTS INSPECTION
184 % The inspection of the results is a very important step. To see if the
185 % coregistration works well or not we need to use SPM also in this case. We
186 % can click on "Check registration" and it will pop up a window where we
187 % need to specify the images to represent at the same time. Obviously, as
188 % the other visualization methods, the 3D objects analysed are represented
189 % through an axial, a saggital and a coronal view. If we see that something
190 % gone wrong because the results obtained are not align we could change
191 % some algorithm parameters as the cost function or the interpolation
192 % methods.
193 %
194 %
195 % DATA MANIPULATION -----
196 %
197 % Now that we have all the possible information in the same space we could
198 % work with them. All the voxels refers to the same position thanks to the
199 % co registration techniques applied. One important results that we could
200 % retrieved are some information about the dynamics data coming from the
201 % PET in different brain region. To do so we need to do these following
202 % steps:
203 %
204 % - Import data:
205 % - rPET : Co registered PET
206 % - ratlas : Co registered Atlas
207 % - t1w : Reference structure
208 % - Select atlas region and save the results in a index vector.
209 % - Using the index vector we choose the anatomical region in the PET.
210 % - Summarize the PET activity in the region.
211 %
212 % This will let us be able to find some data that summarize the activity
213 % inside a specific brain region with a specific anatomical and
214 % physiological role.
215 %
216 % PROGRAM -----

```

217
218 %% CCC
219 % Clear all, close all and clear the comand window
220 clear all, close all, clc
221
222 %% LOAD - DATA
223 % Load the data inside the workspace of Matlab
224
225 dynPET      = load_untouch_nii('rPET.img');                      % PET single volume
226 PET         = double(dynPET.img);
227
228 dynATLAS    = load_untouch_nii('ratlante_2mm.nii');                % ATLAS
229 ATLAS       = double(dynATLAS.img);
230
231 dynT1w      = load_untouch_nii('T1_2mm.nii');                     % T1 Weighted
232 T1w         = double(dynT1w.img);
233
234 WMMask      = load_untouch_nii('c1T1_2mm.nii');                  % White matter
235 WMMask       = double(WMMask.img);
236 WMMask       = WMMask > 150;
237
238 GMMask      = load_untouch_nii('c2T1_2mm.nii');                  % Gray matter
239 GMMask       = double(GMMask.img);
240 GMMask       = GMMask > 150;
241
242 PET(PET < 0) = NaN;
243 ATLAS = repmat(ATLAS, [1,1,1,size(PET,4)]);
244 WMMask = repmat(WMMask, [1,1,1,size(PET,4)]);
245 GMMask = repmat(GMMask, [1,1,1,size(PET,4)]);
246
247 %% MANIPULATION - DATA (1)
248 % Manipulation of the data loaded. We have to estimate the dynamic activity
249 % in the following four ROIs:
250 % - 40 : Thalamus left
251 % - 41 : Thalamus right
252 % - 17 : Cerebellum right
253 % - 18 : Cerebellum left
254
255 ThalamusL = 40;           ThalamusR = 41;
256 CerebellumR = 17;        CerebellumL = 18;
257
258 iThalamusLW = (ATLAS == ThalamusL) .* WMMask;
259 iThalamusLG = (ATLAS == ThalamusL) .* GMMask;
260 iThalamusRW = (ATLAS == ThalamusR) .* WMMask;
261 iThalamusRG = (ATLAS == ThalamusR) .* GMMask;
262 iCerebellumRW = (ATLAS == CerebellumR) .* WMMask;
263 iCerebellumRG = (ATLAS == CerebellumR) .* GMMask;
264 iCerebellumLW = (ATLAS == CerebellumL) .* WMMask;
265 iCerebellumLG = (ATLAS == CerebellumL) .* GMMask;
266
267 aThalamusLW = squeeze(nanmean(PET.*iThalamusLW , [1,2,3]));
268 aThalamusLG = squeeze(nanmean(PET.*iThalamusLG , [1,2,3]));
269 aThalamusRW = squeeze(nanmean(PET.*iThalamusRW , [1,2,3]));
270 aThalamusRG = squeeze(nanmean(PET.*iThalamusRG , [1,2,3]));

```

```

271 aCerebellumRW = squeeze(nanmean(PET.*iCerebellumRW,[1,2,3]));
272 aCerebellumRG = squeeze(nanmean(PET.*iCerebellumRG,[1,2,3]));
273 aCerebellumLW = squeeze(nanmean(PET.*iCerebellumLW,[1,2,3]));
274 aCerebellumLG = squeeze(nanmean(PET.*iCerebellumLG,[1,2,3]));
275
276 tiledlayout(2,2,'TileSpacing','compact')
277
278 h(1) = nexttile;
279 hold on, p(1) = plot(aThalamusLW,'b');
280         p(2) = plot(aThalamusLG,'m'); hold off
281         title('Thalamus Left'),
282
283 h(2) = nexttile;
284 hold on, plot(aThalamusRW,'b');
285         plot(aThalamusRG,'m'); hold off
286         title('Thalamus Right'),
287
288 h(3) = nexttile;
289 hold on, plot(aCerebellumRW,'b');
290         plot(aCerebellumRG,'m'); hold off
291         title('Cerebellum Right')
292
293 h(4) = nexttile;
294 hold on, plot(aCerebellumLW,'b');
295         plot(aCerebellumLG,'m'); hold off
296         title('Cerebellum Left')
297
298 set(h, 'XGrid', 'on', 'YGrid', 'on', 'XMinorGrid','on','YMinorGrid','on')
299 set(h, 'Box', 'on'),
300 l = legend(nexttile(2),{'WM','GM'});
301 l.Location = 'NorthEastOutside';
302
303 %% MANIPULATION - DATA (2)
304 % Manipulation of the data loaded. We have to estimate the dynamic activity
305 % in the following four ROIs:
306 % - 28 : Middle frontal Gyrus Left
307 % - 29 : Middle frontal Gyrus Right
308 % - 64 : Lingual gyrus Left
309 % - 65 : Lingual gyrus Right
310 % - 17 : Cerebellum right
311 % - 18 : Cerebellum left
312
313 MiddleFrontGyrusL = 28; MiddleFrontGyrusR = 29;
314 LingualGyrusL = 64;      LingualGyrusR = 65;
315 CerebellumR = 17;       CerebellumL = 18;
316
317 iMiddleFrontGyrusLW = (ATLAS == MiddleFrontGyrusL) .* WMMask;
318 iMiddleFrontGyrusLG = (ATLAS == MiddleFrontGyrusL) .* GMMask;
319 iMiddleFrontGyrusRW = (ATLAS == MiddleFrontGyrusR) .* WMMask;
320 iMiddleFrontGyrusRG = (ATLAS == MiddleFrontGyrusR) .* GMMask;
321 iLingualGyrusLW = (ATLAS == LingualGyrusL) .* WMMask;
322 iLingualGyrusLG = (ATLAS == LingualGyrusL) .* GMMask;
323 iLingualGyrusRW = (ATLAS == LingualGyrusR) .* WMMask;
324 iLingualGyrusRG = (ATLAS == LingualGyrusR) .* GMMask;

```

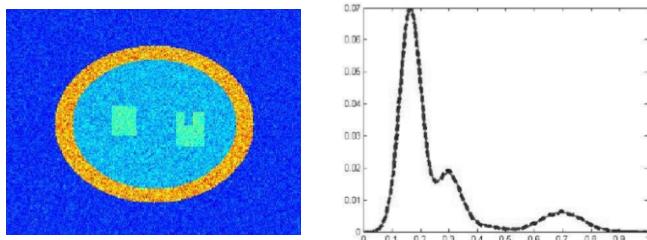
```
325 iCerebellumRW = (ATLAS == CerebellumR) .* WMMask;
326 iCerebellumRG = (ATLAS == CerebellumR) .* GMMask;
327 iCerebellumLW = (ATLAS == CerebellumL) .* WMMask;
328 iCerebellumLG = (ATLAS == CerebellumL) .* GMMask;
329
330 aMiddleFrontGyrusLW = squeeze(nanmean(PET.*iMiddleFrontGyrusLW,[1,2,3]));
331 aMiddleFrontGyrusLG = squeeze(nanmean(PET.*iMiddleFrontGyrusLG,[1,2,3]));
332 aMiddleFrontGyrusRW = squeeze(nanmean(PET.*iMiddleFrontGyrusRW,[1,2,3]));
333 aMiddleFrontGyrusRG = squeeze(nanmean(PET.*iMiddleFrontGyrusRG,[1,2,3]));
334 aLingualGyrusLW = squeeze(nanmean(PET.*iLingualGyrusLW,[1,2,3]));
335 aLingualGyrusLG = squeeze(nanmean(PET.*iLingualGyrusLG,[1,2,3]));
336 aLingualGyrusRW = squeeze(nanmean(PET.*iLingualGyrusRW,[1,2,3]));
337 aLingualGyrusRG = squeeze(nanmean(PET.*iLingualGyrusRG,[1,2,3]));
338 aCerebellumRW = squeeze(nanmean(PET.*iCerebellumRW,[1,2,3]));
339 aCerebellumRG = squeeze(nanmean(PET.*iCerebellumRG,[1,2,3]));
340 aCerebellumLW = squeeze(nanmean(PET.*iCerebellumLW,[1,2,3]));
341 aCerebellumLG = squeeze(nanmean(PET.*iCerebellumLG,[1,2,3]));
342
343 tiledlayout(3,2,'TileSpacing','compact')
344
345 h(1) = nexttile;
346 hold on, p(1) = plot(aMiddleFrontGyrusLW,'b');
347         p(2) = plot(aMiddleFrontGyrusLG,'m'); hold off
348         title('Middle frontal Gyrus Left'),
349
350 h(2) = nexttile;
351 hold on, plot(aMiddleFrontGyrusRW,'b');
352         plot(aMiddleFrontGyrusRG,'m'); hold off
353         title('Middle frontal Gyrus Right'),
354
355 h(3) = nexttile;
356 hold on, plot(aLingualGyrusLW,'b');
357         plot(aLingualGyrusLG,'m'); hold off
358         title('Lingual gyrus Left')
359
360 h(4) = nexttile;
361 hold on, plot(aLingualGyrusRW,'b');
362         plot(aLingualGyrusRG,'m'); hold off
363         title('Lingual gyrus Right')
364
365 h(5) = nexttile;
366 hold on, plot(aCerebellumLW,'b');
367         plot(aCerebellumLG,'m'); hold off
368         title('Cerebellum Left')
369
370 h(6) = nexttile;
371 hold on, plot(aCerebellumRW,'b');
372         plot(aCerebellumRG,'m'); hold off
373         title('Cerebellum Right')
374
375 set(h, 'XGrid', 'on', 'YGrid', 'on', 'XMinorGrid','on','YMinorGrid','on')
376 set(h, 'Box', 'on'),
377 l = legend(nexttile(2),{'WM','GM'});
378 l.Location = 'NorthEastOutside';
```

379
380
381
382

LESSON 5: SEGMENTATION, STATIC AND DYNAMIC PET DATA

Segmentation: Gaussian mixture model

In general, it is not used the Otsu algorithm and a better way to approach on segmentation is the *Gaussian mixture model*. It exploits the possibility to represent and deal with the image's histogram. The hypothesis on the base of this technique is that every part of the image has intensity coming from a gaussian probability density function. In the chart reported below, where there is represented the histogram of the image, we could see that it may be represented by the sum of gaussian's probability. To correctly separate the results, we need to define specifically the number of classes which are needed to be recognized. In a mathematical way we can express our image's histogram as:

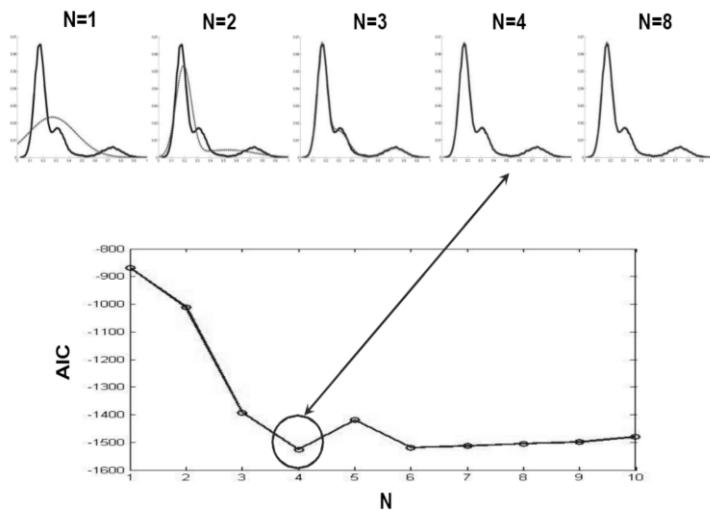


$$h(g) \approx GM(g) = \sum_{i=1}^N w_i e^{-0.5 \left(\frac{\mu_i - g}{\sigma_i} \right)^2}$$

Starting from the histogram of the images we need to estimate all the possible parameters of the previous sum of exponential model by the usage of the appropriate estimator. After having done this step, we can identify and choose all the possible gaussian distribution. We need to correctly set the threshold which will be placed in the position of minimum classification error (correspondently on the x position where it can be found the intersection between two different gaussian distribution).

One of the biggest problems of this method is that we do not know the exact number of gaussian distribution we need to use. So first we can set the number of gaussian distribution, then we need to calculate the histogram of the image by sampling the grey levels. After do that we need to define the analytical model of Gaussian mixtures using $h(g)$ previously defined. Finally, we can estimate the model parameters by different methodology as instance the non - linear weighted or not least square.

The choose of the number of tissues is very crucial. By increasing the number of models, we could increase the accuracy of the prediction, but we will obtain models of increasingly complexity. To select the best one, we need to use some parsimony criterion to balance between parsimony and goodness of fit exempli gratia the Akaike information criterion.



The image reported above represent the fit of the previous image's example by the usage of increasing complexity model. Higher is the model complexity better is the estimation of the curve but also the number of parameters to estimate increases. The model selected is as expected not the one with the highest order but the one with order four.

Segmentation: Expectation maximization approach

Another approach usable to compute correctly all the parameters' values is the *expectation maximization approach*. This method is a good alternative of the nonlinear weighted or non – weighted least square because this paradigm has some troubles. First it could be sensitive to initial conditions, it needs of computation of gradient and finally, the model is optimized on an approximate histogram. The *expectation maximization algorithm* is made up of two steps.

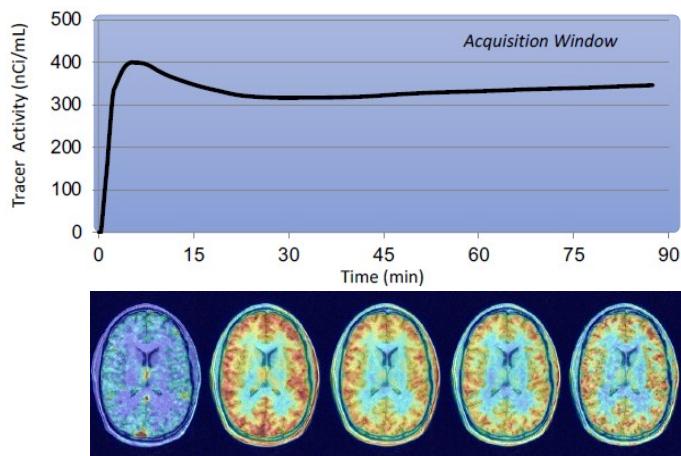
The first consist in the next one: we start from fixed values for the mean and the variance of the density probability distribution and so we can compute the probability density function of all the data. Then second step considering the results regenerate the mean and variance to rebuild the probability density function. These two steps are iterated until the solution converge. The algorithm is usually ended when two subsequent estimation are close than a threshold.

In SPM, the number of gaussian function is settled to a certain value pre – installed. In SPM this is not the algorithm which is implemented to correctly perform the segmentation but is a very big approximation. In those specific cases when we must deal with people with some diagnosis, in the most of software we need to change all the template that needed to be regenerated to correctly deal with these patients.

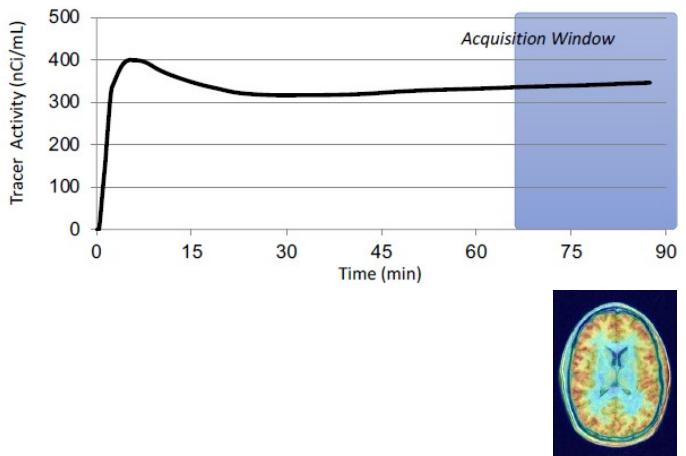
PET: Static data & Dynamic data

Pet is a very wide measurement system because we have a lot of tracers that we could use depending on the type of studies to perform. Even in research and clinical area of medicine PET is largely used. There are two paradigms to obtain and measure PET data, the static and the dynamic PET imaging. Even if the paradigm of measurement is different the PET data are always obtained with the same principle. This consists in inject a certain dose of radiotracers via venous and then measure the decay activity of the radionuclide. Generally, the interval of

time where are performed the measurements identify the two different paradigms. In the static one we analyse the PET data just for a brief interval of time usually after a certain delay, while with the dynamic PET data we observe the overall decay process. In the first paradigm we need to use the scanner for a briefer amount of time and usually in the instant of dose injection the patient does not need to be in the scanner's bed. While, instead, with dynamic data analysis, when the injection is performed the patient need to be lay down on the scanner's bed and the measurement activity starts. In the most of clinical analysis the dynamical data are not needed because of the major field of application of PET which is oncology. In this cases doctors need to evaluate the metabolism activity of the tumour inside the body and detect its changes. To get this information long dynamic data are not needed and so the major analysis retrieves just static data. There is one more other reason because of the PET data are not inspected dynamical in clinical application and is because of the high cost of the technology. We can find a more dynamical use of PET instrumentation when we are dealing in research field.



Dynamic PET acquisition



Static PET acquisition

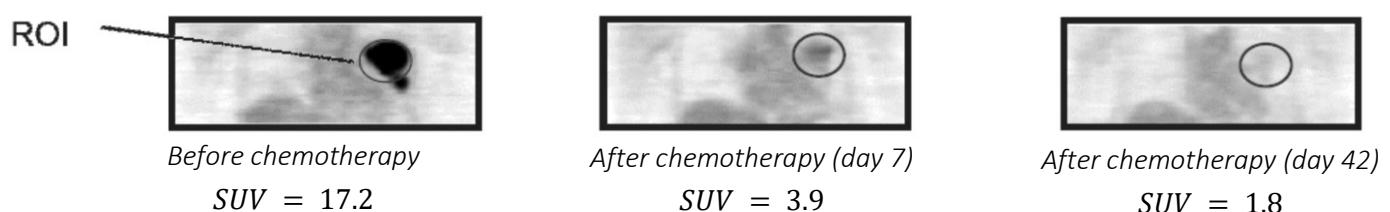
We have to hold in mind that the radiotracers injected are radioactive and so they can damage the body structures. Radiotracers after a certain time interval start decay and decrease the possibility to detect the annihilation process. To compensate this pitfall, we obtain an image by considering wider window of time. For instance, when we want to acquire an image by the usage of this instrumentation, we need to consider interval of time depending on the time passed from the dose injection. Usually, the first couple minutes from the injection are characterised by rapid changes (dynamic of the curve) and after an interval of time usually defined as settling time the value of the curve remain almost stable. So, to represent correctly in one voxel the PET data we need to adopt shorter window for the first couple of minutes to describe better the dynamic behaviour of the data and wider window for the steady state condition. The measure which is done through all the detectors are the count for each voxel of the scintillation phenomenon detected. After that, thanks to the knowledge of what does a scintillation phenomenon means and the knowledge about the interval of time adopt it is possible to reconstruct the concentration of radiotracers in that voxel.

PET: Static PET imaging

The information that we obtain from static PET imaging is the *Standardized Uptake Value (SUV)*. It represents the radiotracers concentration over the tracer injected dose. It could be normalized with the body weight or with the body surface area or also with the lean body mass. Usually, we select as tracer the ^{18}F – FDG the normalization is performed over the body surface area, while for most tracers the normalization is over the body weight. The SUV is well correlated with the physiological information that we want to obtain, id est the glucose metabolism for ^{18}F – FDG, serotonin transporter density for $[^{11}\text{C}]$ DASB etcetera.

The advantages of using this metric are that it is easy to calculate and is mainly and largely used for clinical studies. The disadvantages, instead, consist in the non-validation of this metric for several PET tracers.

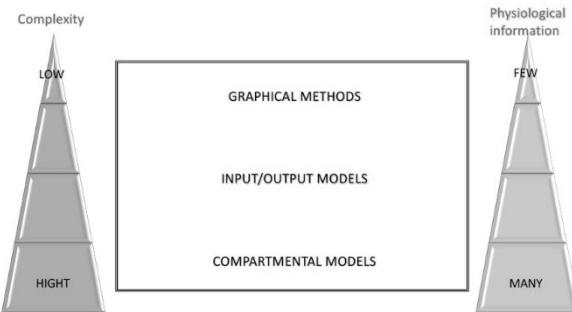
One possible field of application of the SUV is that it is a semiquantitative index often used in oncology. In the image reported below we could see the evaluation of the response to chemotherapy of a tumour.



We can use this metric to understand if the subject is answering correctly to the therapy. In the image reported above we could see a very high change in the tumour size and distinguishability in different days.

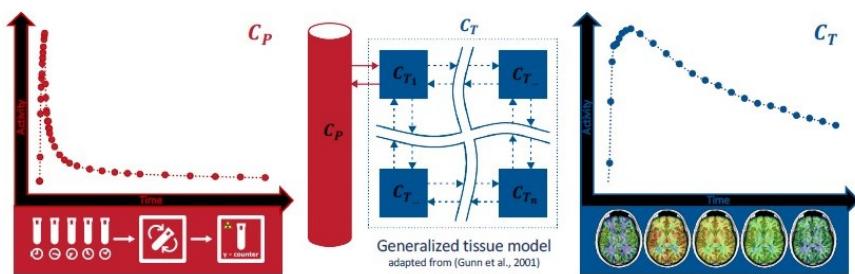
PET: Quantitative functional imaging

The next concept that we need to introduce is the quantitative functional imaging. We will start with compartmental models to describe the PET kinetics. From this analysis we can derive many physiological information. The downside of using this approach is the high complexity in dealing with compartments and requires a lot of priori information. Other studies methodology are the ones that uses input/output models and graphical models. These latest techniques even if they present a very low level of complexity, they do not provide a lot of physiological information.



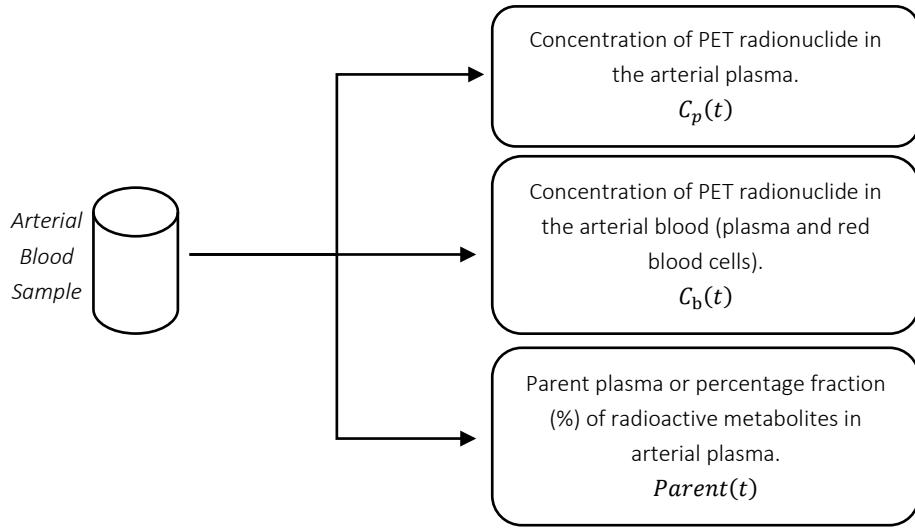
PET: Dynamic PET imaging

When we have dynamic PET data called time activity curve or with the acronym TAC the appropriate model is applied to the tracer under examination and the parameters of the model are estimated by weighted nonlinear least squares. As already said before we can access to different information by choosing the right radiotracer. The data that we need to have to correctly estimate all the parameters of the model are an invasive sampling of the blood from an artery (it is crucial to have blood samples coming from an artery and not from a venous) and the activity from the PET image of each voxel in the body region under analysis. By merging this two information, and with some a priori knowledge about the model structure we could perform an estimation process and find out some very meaningful information such as blood flow, glucose uptake, potential binding, and other more important physiological information not retrievable with simple static PET technique.



PET: Dynamic PET imaging and blood radiotracer concentration

Kinetic analysis of dynamic PET images is based on a model of the tissue. To invert the model and thus obtain estimates of physiologically meaningful parameters, it is necessary to also know the exact input of the model. For this reason, we have always to remember the different state of the radiotracer inside the blood.



In our compartmental model one of the most important part is the presence of a structure that regulate the exchange of tissue between the arterial blood flow and the inner zones of the tissue. To understand how this interface works and which are the fractional transfer coefficient that describe the exchange of material between the two previous mentioned sites we need to detect the real amount of radiotracer that is available in the blood. To detect this quantity, we need to introduce the concepts explained in the chart reported above. First, the C_p represent the plasma concentration of the radionuclide. We need to remember that the radionuclide is just the radioactive structure inside the radiotracers. For this reason, just have some information about this concentration is not enough. Because mainly, when the substance is injected into our body it is processed and undergo on two main type of degradation of its properties:

- *Red blood cells metabolism*, blood is made up also from a cellular fraction. In this part of the blood, we could find also the red blood cells. These last ones capture the radiotracers and metabolise them or keep them connected to the cell membrane. This deactivates the radiotracers metabolism properties but keep the radioactivity properties alive.
- *Liver or other organ metabolism*. This process is quite common with all the possible radiotracers. This because of all the radiotracers are made up with some proteins or some carbohydrates, which are involved in some body function, marked with specific radionuclides. This is a very significant properties because all the biological structures could be metabolized from different organs and this led to the loss of the functional role but keep the radionuclide unaltered and free in the plasma.

So, after having understand the high quantity of radiotracers that are deactivate in the blood sample, we want to measure just the concentration of still active PET tracers and so we need to clean from the blood sample also from the metabolized radiotracers molecules. From the arterial blood samples, we could measure the C_p concentration that we have already mentioned previously and other important concentration and fraction such as C_b and $\text{Parent}(t)$. The first one represents the fraction of the PET radionuclide in the arterial blood (in the cellular and plasma fraction) while the parent fractions represent the fraction of radioactive metabolites in arterial plasma. These data are very meaningful and usually when all the measurement process are complete, we receive our data from the laboratory. The first step now is to try to estimate with different model the parent and then apply it to multiply the concentration C_p and C_b to find just the concentration in plasma and overall, in the blood of radioactive tracers available. These new concentrations will be then used in the model and will be considered as noise free signal. This is a very

important assumption because let us be able to apply a forcing function method to estimate all the parameters of the complete model.

PET: Quantification of PET images

Now that we have the blood samples and the PET dynamic data, we need some quantification of PET images at *ROI level* or *Voxel Level*. These last two are different methods to inspect the PET images results. Until 10 year ago all PET studies were performed on ROI level while nowadays all the analysis are performed on Voxel level.

- *ROI level*: This type of analysis is the less used nowadays but still have a high signal to noise ratio compare to the others. We can use the weighted nonlinear least square, the method is computationally demanding but the ROIs are few, instead as downside with defining different ROIs we lose the original spatial resolution of the images.
- *Voxel level*: In this analysis the signal to noise ratio is very low because we are investigating at the voxel level, another downside of this paradigm is that we have many data and so we need for fast algorithm for identification. One important result, instead, is the possibility to keep the original spatial resolution of the images.

When we perform the estimation process at voxel level, the kinetic parameters can be calculated for each of them: the maps that display for each slice the various kinetic parameters are called parametric maps.

PET: Model identification

Now, considering the invasive artery information and all the PET data, TAC for each voxel if we apply a voxel level analysis or a TAC for each ROIs if we apply a ROI level analysis, we could estimate the parameters of our model at voxel level and we could get some important physiological information about the process. One trivial but important thing is that we have several parametric maps equal to the number of parameters begin used in our model.

PET: Receptor's kinetics quantitative studies

The PET is a nuclear medicine technology, and it is strongly related with radiation decay. We can use PET to perform a several blood flows analyses by the application of specific tracers. One the most recent and important discovery of the PET application is to measure the receptors density. The last 20 years of PET application has been dominated in receptors analysis. We will deal just on the schematic view of synapse in the brain and the process of secreting the neurotransmitter. To understand this part, we subdivide the process in steps. In the first step the neurotransmitter is contained inside vesicles, when the stimulus arrives to the axonal button the calcium channels open and vesicles excite the neurotransmitter in the synaptic cleft. Neurotransmitters reach the next neuron dendrite or target cells and connect to specific receptors and activate specific answer in the target tissues. By the usage of the appropriate radiotracers, we could estimate the number of vesicles, the number of presynaptic receptors, the quantity of neurotransmitters delivered in the synaptic cleft and finally the number of postsynaptic receptors. Usually, when we deal with this type of studies, we have some important physics quantity to access:

- B_{max} : It is also known as receptor concentration. It is expressed in [nM] or [nmol/l].

- $\frac{1}{K_d}$: It represents the binding affinity [$nmol/l$]. It is the representation of the interaction of most ligands with their binding sites. If the affinity of the PET tracer is high for a particular receptor, then only a low concentration is needed to bind many of the receptors.
- $BP = \frac{B_{max}}{K_d}$: It represents the binding potential. It is obviously dimensionless.
- V_d : In general, is called as volume of distribution and it is measured in [ml/cm^3] or [ml/g^{-1}]. It is equal to the ratio of tissue and plasma PET tracer concentration at the equilibrium. It is related to the BP value, higher is this quantity higher the volume of distribution is.

One important thing to remember is that if we perform a voxel analysis with the PET technique, we have to remember that a voxel represents a lot of neurons, glia cells, microglia, astrocytes, oligodendrocytes and also some capillary. So, the single voxel is a complete mixture of different tissues and so all the data coming from a voxel represent an average intensity.

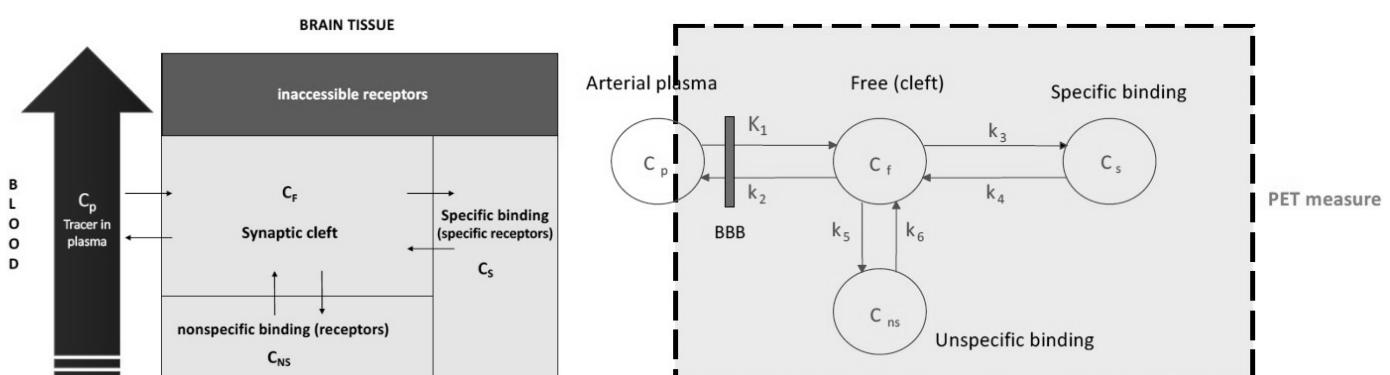
Some examples of application of in vivo imaging technique are the one reported below. Generally, to study the brain receptor system we could use two different paradigms.

- In the first one we use only the PET tracer. In this case we could find out some information about the permeability of the brain blood barrier to the tracer or ligand, we could calculate the distribution volume and finally we can measure the binding potential.
- In the second one instead, we use a PET tracer and a ligand (displacement) and we can calculate the receptor occupancy of the ligand and also measure the affinity and receptor density in a specific region. By using the PET tracer and a ligand we could calculate the receptor occupancy of the ligand and to measure the affinity and receptor density in a specific region.

PET nowadays is largely used by big pharmacology companies to analyse and understand how the body interact with drug delivered. Compartmental model in parallel to these types of analysis are very important because they can produce some important additional knowledge.

PET: Compartmental model

We want to start now with compartmental model. We need to study which compartmental model we could use to understand the movement of the radiotracer. From the scheme reported below we could see the exchange of molecules between the blood flow and the bidimensional voxel. Once the tracer is inside synaptic cleft it could bound with specific receptors and nonspecific receptors. We need to consider all these type of situation inside the brain tissue to describe correctly on what is happening inside a voxel. The image below is a physiological representation of what really happens. It is already a compartmental model and so we can move from the schematic representation to the specific for the compartmental model.



We could describe the dynamic of each compartment with the next system of equation. There are just three compartment and not four because we have already all the information about the concentration inside the arterial plasma that it is supposed as a noise free signal.

$$\begin{aligned}\dot{C}_f(t) &= K_1 C_p(t) - (k_2 + k_3 + k_5) C_f(t) + k_4 C_s(t) + k_6 C_{ns}(t) & C_f(0) &= 0 \\ \dot{C}_s(t) &= k_3 C_f(t) - k_4 C_s(t) & C_s(0) &= 0 \\ \dot{C}_{ns} &= k_5 C_f(t) - k_6 C_{ns}(t) & C_{ns}(0) &= 0\end{aligned}$$

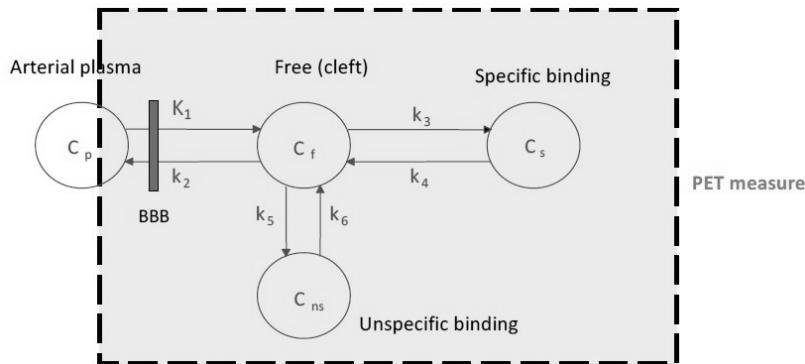
The initial values are put to zero because the radioactive tracers are set to be zero initially. There is an addition information about the PET measure, which is:

$$PET(t) = (1 - V_b)(C_f + C_s + C_{ns}) + V_b C_{blood}(t)$$

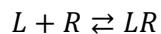
It represents the overall activity detectable and strictly connected to the voxel activity. In general, the parameter V_b represent a fraction and it is possible to estimate it. Usually is equal to the 0.05 and this means that the PET measure's are affected for just the 5% of information that come from the blood.

LESSON 6: PET COMPARTMENTAL MODEL AND GRAPHICAL METHODS

PET: Compartmental model and parameters



K_1 is a capital k in fact the unit of measurement is different from the other fractional transfer coefficient assumed constant during the tracer experiment. PET is an exam as we have already said that can specify a measure related with physiology. Now we introduce the meaning of the parameters in the model. Before doing so, we need to remark a chemical reaction between the free radiotracer inside the synaptic cleft and the specific binding site. This last interaction could be treated as a biomolecular reaction. This has the analytical representation reported below:



Where the forward reaction is mediated by the rate constant k_{on} . While the inverse reaction is characterized from a rate constant k_{off} . The first one represents the rate constant at which the binding site and the radiotracer interact, while the second represent the inverse activity or the rate at which the degradation of the connection between the radiotracer and the binding site works. We may measure in vivo the value of k_{on} . By knowing this term and the k_{off} term we can establish some important relation between the parameters of the previous compartmental model presented and the rates constant k_{on} and k_{off} .

$$k_3 = B_{max}k_{on}$$

$$k_4 = k_{off}$$

$$K_D = \frac{k_{on}}{k_{off}}$$

Where the meaning of these variables is:

- B_{max} : Represent the total concentration of receptors. Usually is measured in nM.
- K_D : It identifies the equilibrium dissociation constant. It is usually adimensional. The inverse of this variable and so $1/K_D$ represent the affinity.

From a pathological point of view these parameters may lead to a better distinguish between healthy patients and nonhealthy ones. Another relevant parameter that could be found is the binding potential. We indicate this quantity as BP and we may express it as:

$$BP = \frac{B_{max}}{K_d}$$

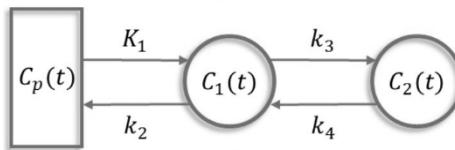
As we can derive from the previous relation if we fix the affinity $1/K_d$ and B_{max} , instead, increases also the binding potential increases too. When we hold fix the total concentration of receptors or B_{max} and we vary the affinity $K_d = 1/K_D$, if this last term increases (and so the equilibrium dissociation constant decreases) the binding potential BP increases too. To conclude this analysis, the binding potential is inverse related with the equilibrium dissociation constant but is proportional to the total concentration of receptors.

PET: Compartmental model identifiability

The compartmental model reported in the beginning of this lesson is the full version of the compartmental model. This one suffers from some identifiability issues. In fact, this specific model is not a priori identifiable (this means that the number of parameters that can be estimated is much higher than the possible measure that we can take from the model). This could be demonstrated, since we have assumed constant the different fractional transfer coefficient (k_i), by the usage of Laplace or analytical methods. If we neglect this fact and we try to solve the system, we will obtain two symmetric solution for k_3 and k_4 and k_5 and k_6 .

PET: 2TCM

To solve this issue, we need to simplify the compartmental model and the best way to achieve the best results is by the usage of the compartmental model reported below:



This compartmental structure is also known as: 2TCM or two tissue compartment model. It is largely used for most of PET receptors studies. In this case the compartment representative of the concentration of the radiotracers inside the synaptic cleft and the compartment representative of the concentration of the radiotracers bind with non – specific receptor are merge. The last compartment represents still the concentration of tracer specifically bound to the target receptor. The k_3 and k_4 are still representative of the same quantity. The only measure in this case which is adimensional is the binding potential that in this case we could express as:

$$BP = \frac{k_3}{k_4}$$

This quantity is unitless. Other important parameters are the volume of distribution known also as V_d . This represents the volume occupied by the PET tracer. Usually, it could be measured in ml/cm^3 . It is calculated by a formula that involves all the parameters of the model.

$$V_d = \frac{K_1}{k_2} \left(1 + \frac{k_3}{k_4} \right)$$

The model could be easily described from differential equations easily derivable from the model structure. These are:

$$\begin{aligned}\dot{C}_{f+ns}(t) &= K_1 C_p(t) - (k_2 + k_3) C_{f+ns}(t) + k_4 C_s(t) & C_{f+ns}(0) &= 0 \\ \dot{C}_s(t) &= k_3 C_{f+ns}(t) - k_4 C_s(t) & C_s(0) &= 0\end{aligned}$$

The measured tracer concentration in each voxel, $C_{voxel}(t)$, is described by the percentage of the tissue component and the fraction of the blood component within the voxel. This percentage is represented with the symbol V_b . The equation of the concentration in the voxel is:

$$C_{voxel}(t) = (1 + V_b) (C_{f+ns}(t) + C_s(t)) + V_b C_{blood}(t)$$

From this model we can derive a lot of macro parameters. These are the one listed in the table below.

<i>Equilibrium measure</i>	<i>Rate constant expression</i>	<i>Parameter name and description</i>
C_{ns}/C_p	K_1/k_2	V_{ns} is defined as the volume of distribution of non – displaceable compartment.
C_s/C_p	$(K_1 k_3)/(k_2 k_4)$	V_s represent the distribution volume of specific binding compartment, also called BP_p .
C_T/C_p	$K_1/k_2(1 + k_3/k_4)$	V_T defines the total distribution volume.
C_s/C_{ns}	k_3/k_4	This is the specific non – displaceable equilibrium partition coefficient, also equal to BP_{ns} .

The meaning of each term in the compartment model 2TCM, its equivalent measure and the unit of measurement are summarized in the table below.

<i>Rate constant</i>	<i>Biophysical equivalent</i>	<i>Unit of measurement</i>
K_1	$CBF(1 - e^{PS/CBF})$	$ml \text{ cm}^{-3} \text{ min}^{-1}$
k_2	K_1/V_{ns}	min^{-1}
k_3	$f_{ND} k_{on} B_{max}$	min^{-1}
k_4	k_{off}	min^{-1}

In this table is highlight how each rate constant are characterize with a different unit of measurement. As we have already seen at the beginning of the lesson K_1 is capital and not small because have a different unit of measurement. Moreover, is important to underline that CBF is an acronym and stand for *cerebral perfusion pressure*. We need to highlight the k_3 parameters value and its biophysical equivalent. It is represented from the product of the fraction of free and non – specifically bound tracer, f_{ND} , k_{on} and B_{max} . There is an expression which represent the f_{ND} term and its definition:

$$f_{ND} = C_f/C_{f+ns}$$

Generally, K_1 is a direct expression of the *Blood Brain Barrier* (BBB) effectiveness. By using compartmental model, we can estimate parameters that are directly connected with biophysical processes. Obviously, this has a cost, and this is represented by the model complexity. Moreover, by comparing value of patient with diagnosis of some neuropathology and healthy ones we can establish where the pathology works and develops. This leads to a better understanding of the mechanism of the pathology. Other important binding potential are:

$$BP_F = \frac{B_{max}}{K_d}$$

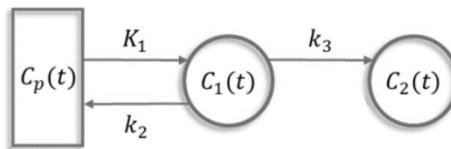
$$BP_{ND} = f_{ND} \frac{B_{max}}{K_d}$$

To highlight the unit of measurement in PET we introduce the next table to sum up all the possible unit of measurement.

<i>Parameter</i>	<i>Unit of measurement</i>
K_1	[ml/cm ³ /min] [ml/g/min] [ml _{plasma} /ml _{tissue} /min]
k_2, k_3, k_4	[1/min]
$C_p(t)$	[kBq/ml _{plasma}]
$C_{f+ns}(t), C_s(t)$	[kBq/cm ³] [kBq/g] [kBq/ml _{tissue}]

PET: 2TCM Simplification

The model is different when it is close to zero the rate constant of degradation of the connection between the specific receptor and the tracer. In this case the model is characterized by the next structure.



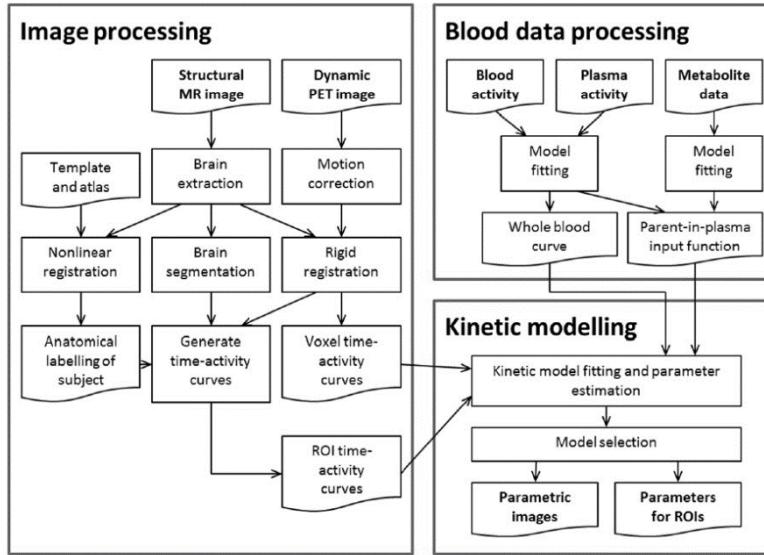
The k_2 is the outflow rate constant and quantifies how much material goes out from the brain. k_3 is referring to a chemical process, or the rate constant of the transformation of the radiotracer from its initial state to the metabolized version of the radiotracers. The macro parameters in this case could be represented as:

$$K_i = \frac{K_1 k_3}{k_2 + k_3}$$

PET: Compartmental model parameter's estimation

We have now all the possible information to identify our model. Suppose to analyse the two compartmental model with the value of k_4 different from zero. Now we need to compare the possible estimators to correctly identify the parameter's values. The gold standard in this case is to identify the model for the quantitative analysis for all the tracers. We can estimate all the possible parameters by the usage of NLWLS technique. This provides a precise and accurate estimate. By the usage of this method, we have some problem as instance the initial estimates selection, its onerous computational effort and finally it is a method that may not converge if the noise level is too high. So, we need to distinguish when we could apply the WNLLS and

when not. When we perform ROI analysis the WNLLS works well but not when we perform a voxel approach. To conclude we need to find out some other methods applicable to reduce the complexity of the voxel level analysis and find out still good results. Obviously, all the techniques that we will present have some pros and some cons. The pros are that they are computationally not expensive and often the quality of the estimates is very similar to that of the WNLLS. The cons, instead, are that sometimes the accuracy of estimates is lost and there are any preferable methods to apply in a situation instead of another and the choice partially lies with the person that conduct the analysis. Before going deeper inside other methods we need to resume how the process reach this point.



In the image represented above we could see all the possible pre – processing steps to apply on PET and blood data to reach then the kinetic modelling phase where we can extract some physiologically relevant information. The kinetic model fitting stage could consider different techniques depending on the tracer used to perform the experiment. When the modelling step is finished, we could get some parametric images which are images that represents the value of a parameter of the model all over a slice of our head.

PET: Parametric map generation

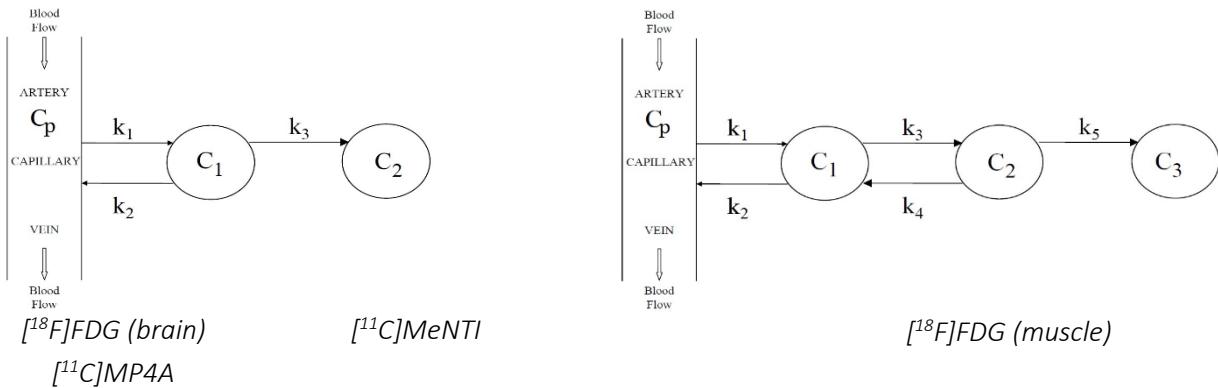
The principal methods for the generation of parametric maps are the one listed below:

- *Graphical method (Patlak and Logan)*
- *Basis Function Method BFM (Gunn)*
- Generalized Linear Least Squares GLLS (Feng)
- *Linear Ridge Regression with Spatial Constraints, simplified or generalized (Zhou)*
- *Non – Linear Ridge Regression NLRRSC (Zhou)*

The first three methods can be used also at ROI level, instead the fourth method since it uses the concept of spatial proximity can be used only at voxel level. We will deal just with the first two methods.

PET: Parametric map generation – Patlak Graphical Method

The Patlak Graphical Method starts by considering again the compartmental model. It fixes just some requirements. This last requirement is that there is only one irreversible compartment and the tracer leaves the reversible compartments through the plasma or the irreversible compartment. By comparing the compartmental model made up of two compartments (2TCM) the irreversible one is represented by the compartment majorly positioned on the left side. While the reversible one is represented by the middle compartment. There are other important assumptions to do in order to apply correctly these results. These are that the kinetics of the tracer is described by linear differential equations, of the first order with constant coefficients. If the biological system metabolizes the tracer, its metabolic products are measurable and finally the initial concentration of tracer in the tissue is zero. The Patlak Graphical Method could be applied on the models represented with the following schematic structure.



Right down below the images are reported the types of tracers that could be used to correctly perform the experiment. With this graphical approach we could just estimate the macro parameters and not the single k_i component. To find out some information through the Patlak graphical method we do not need to know the exact number of reversible compartments, but we need just to know that the structure of the model is irreversible.

We could estimate just K_i and we could not separate the different information that are used to compute this index. The meaning of this coefficient does not depend on the model structure but is tracer specific. In fact, if we deal with FDG or MP4A tracers this item represents the *fractional rate of irreversible metabolism* of the tracer. This index quantifies how many millilitres of tracer present in the plasma are metabolized per gram of tissue every minute. Instead, if we perform an analysis with MeNTI as tracer we need to use the next definition for the index. This is the *fractional rate of irreversible tracer binding to specific receptor sites*. It quantifies how many millilitres of tracer present in the plasma bind to the specific sites per gram of tissue every minute.

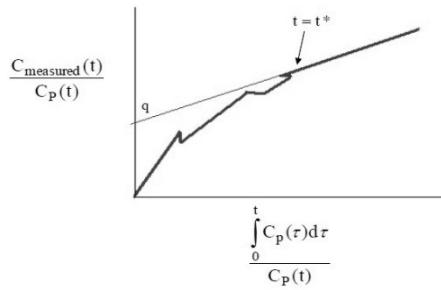
To complete the analysis correctly with Patlak graphical method we need some additional hypothesis. Suppose there is an instant t^* such that for $t > t^*$, all the reversible compartments are in equilibrium with the plasma concentration, in other words:

$$\frac{C_i(t)}{C_p(t)} = \text{costant}$$

For $t > t^*$ if C_i is a reversible compartment we can show that if we use the next coordinate system defined with the coordinates relations:

$$x = \frac{\int_0^t C_p(\tau) d\tau}{C_p(t)} \quad y = \frac{C_{measured}(t)}{C_p(t)}$$

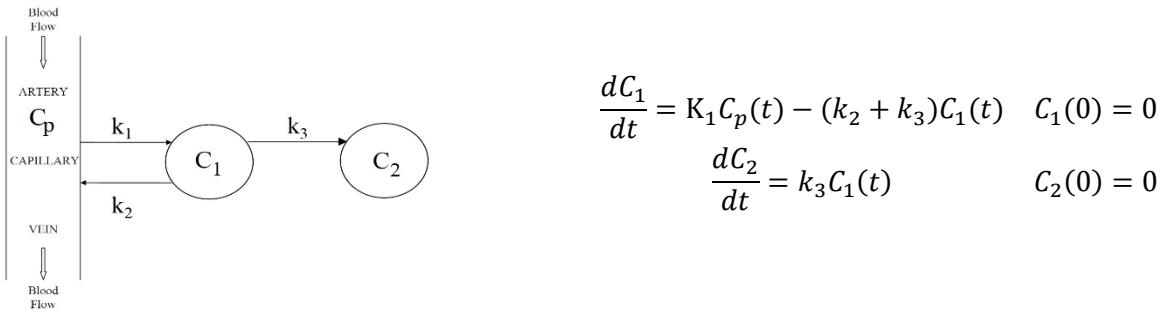
We can build a new graph and find out that the curve plotted after the instant of time t^* the ratio between the concentrations is proportional and so in the chart the curve is following a straight – line trend. The slope of this line is equal to K_i defined before.



This is a very powerful methods because we don't need to use any WNLSS, but we need only the information that comes from the measurement of the concentration in the blood and the concentration measured with PET. Obviously, in this analysis we can clearly see that the Patlak's method implicitly assumes that there is no appreciable presence of blood volume within the voxel. This assumption is acceptable at the voxel level, less acceptable at the region of interest level. Moreover, we do not need to know which the structure of the model is as initially said, but if we have it, we could determine exactly how K_i is linked to the microparameters of the model.

PET: Parametric Map generation – Patlak Graphical Method – Example of linearization

Let consider now the model structure reported below and its relative differential equations:



We can handle these equations in this way, if we hypothesize the blood contribution to the measurement PET $C(t)$ in a voxel is made up as:

$$C(t) = C_1(t) + C_2(t)$$

Our objective of this analysis is to find the analytical mathematical linearization after the instant of time t^* that let us be able to estimate the coefficient K_i . To do so, we have to start from the previous equation and substitute each member of the equation by its integral form. In this way we can write:

$$C_1(t) = K_1 \int_0^t C_p(\tau) d\tau - (k_2 + k_3) \int_0^t C_1(\tau) d\tau \quad C_2(t) = k_3 \int_0^t C_1(\tau) d\tau$$

From the first equation we could write:

$$\int_0^t C_1(\tau) d\tau = -\frac{1}{k_2 + k_3} C_1(t) + \frac{K_1}{k_2 + k_3} \int_0^t C_p(\tau) d\tau$$

We can substitute this expression in the second relation and obtain:

$$C_2(t) = k_3 \int_0^t C_1(\tau) d\tau = -\frac{k_3}{k_2 + k_3} C_1(t) + \frac{K_1 k_3}{k_2 + k_3} \int_0^t C_p(\tau) d\tau$$

Now we merge all the information, and we substitute the expression for $C_2(t)$ in the total concentration one:

$$C(t) = \frac{K_1 k_3}{k_2 + k_3} \int_0^t C_p(\tau) d\tau - \frac{k_3}{k_2 + k_3} C_1(t) + C_1(t) \Leftrightarrow C(t) = \frac{K_1 k_3}{k_2 + k_3} \int_0^t C_p(\tau) d\tau + \frac{k_2}{k_2 + k_3} C_1(t)$$

If there is a time t^* from which the compartment C_1 is in equilibrium with the plasma compartment, id est such that, for $t > t^*$ we have:

$$\frac{C_1(t)}{C_p(t)} = \text{constant} = \alpha$$

If we divide the last equation by the concentration inside the plasma we may obtain:

$$\frac{C(t)}{C_p(t)} = \frac{K_1 k_3}{k_2 + k_3} \frac{\int_0^t C_p(\tau) d\tau}{C_p(t)} + \frac{k_2}{k_2 + k_3} \frac{C_1(t)}{C_p(t)}$$

We obtain for $t > t^*$ the next structure:

$$\frac{C(t)}{C_p(t)} = \frac{K_1 k_3}{k_2 + k_3} \frac{\int_0^t C_p(\tau) d\tau}{C_p(t)} + \frac{k_2}{k_2 + k_3} \alpha \Leftrightarrow y = mx + q$$

This is a linear relationship whose angular coefficient is:

$$m = \frac{K_1 k_3}{k_2 + k_3} = K_i$$

PET: Parametric Map generation – Patlak Graphical Method – K_i extraction

In the previous page we have seen how the linearization after an instant of time t^* works. To let this analysis be clearer as possible we subdivide this process in steps.

- For each voxel and for each instant of time where a new image is acquired, we calculate the coordinate as previous seen.

$$x_i = \frac{\int_0^{t_i} C_p(\tau) d\tau}{C_p(t_i)} \quad y_i = \frac{C_{measured}(t_i)}{C_p(t_i)}$$

- The time t^* is estimated based on the observation of the $x - y$ plot. The instant of time t^* is chosen by select that temporal instant from which the curves (x_i, y_i) shows a linear trend. This choice could be performed by visual inspection or by automatic algorithm.
- We extract only the values of (x_i, y_i) that satisfy this condition and we perform a regression. Thanks to the fact that the model that has to be applied in this situation is a model of data and it is linear in the parameter we can use a linear weighted least square method to fit the curve.

$$\hat{p}_{WLS} = (G^T \Sigma_v^{-1} G)^{-1} G^T \Sigma_v^{-1} z$$

As we know in this case, we need to select the right covariance error matrix. The best way to do this is by using an identity matrix and assuming no error or no weight in the estimation or more in general we can weight each measure with a covariance error matrix diagonal where the $i - th$ element in the diagonal follow the sequent rule:

$$\frac{C_{measured}(t_i)}{\Delta(t_i)}$$

Where in the numerator we find the concentration measured, while in the denominator we found the interval of time needed to acquire correctly the $i - th$ slice.

Obviously, this methodology has some pros and some cons. These are listed down below. Patlak graphical method is very fast, and we do not need any type of non – linear estimators. It allows to estimate only a macroparameter and so we cannot get any information on the individual kinetic parameters. It applicable only on irreversible models. The choice of t^* can be non – trivial and influence the results.

PET: Parametric MAP generation – Logan Graphical method

In the literature exists also the so – called Logan Graphical method. It is an alternative to the Patlak graphical method. We can use it if and only if the PET tracer is described by a reversible model. It is the evolution of Patlak and reevaluate the constraint assumed by the same. In this case, instead of the other, we are assuming that the model is made up only with reversible models. It is a methodology only used with receptors studies. We can estimate only one macroparameter known as volume distribution indicate with the symbol V_d . Its unit of measurement is $V_d [ml_{plasma}/g_{tessuto}]$. Its meaning, independently from the model structure represent

the *distribution volume of the tracer*. This is the ratio between the tracer concentration in the tissue and the steady – state plasma concentration:

$$V_d = \frac{C_{measured}(t)}{C_p(t)}|_{ss}$$

There is an additional hypothesis for the Logan graphical model too. There we can write that for $t > t^*$, all the compartments are in balance with the plasma, in other words works:

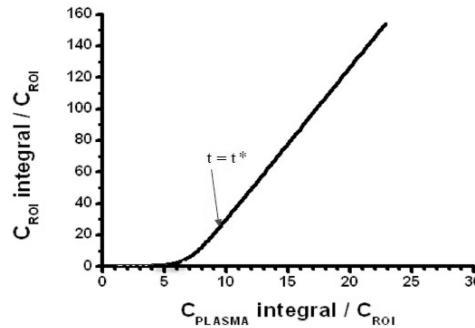
$$\frac{C_i(t)}{C_p(t)} = \text{constant} = \alpha$$

For $t > t^*$ the curve plotted by computing the point's coordinate define as:

$$x = \frac{\int_0^t C_p(\tau) d\tau}{C_{measured}(t)}$$

$$y = \frac{\int_0^t C_{measured}(\tau) d\tau}{C_{measured}(t)}$$

The trend of the curve if the time's inequality is verified follow the one of a straight line. This could be seen from the chart reported below. The slope of the line in this case is equal to the volume of distribution V_d .



As seen previously we can say that if we know the model structure, we could find out an exact relation between the micro and macroparameters. The model linearization and estimation of V_d is quite similar to the one seen for Patlak graphical methods.

```
1 %
2 % Title: Laboratory 3
3 %
4 % Note:
5 % Processing of PET images, Logan plot
6 %
7 % Author: Matteo Martin
8
9 % THEORY -----
10 %
11 % PET DATA
12 % First of all one important remark is that to quantify all the information
13 % retrievable in PET we need to use the compartmental modelling. We have
14 % seen in theory part that we could define different types of
15 % compartmental model based system. But someone, which describes the
16 % physiological concentration with a very intensive meaning are not a
17 % priori identifiable. For this reason researchers have defined some
18 % simplification of these systems which lose something in physiological
19 % information but gain in identifiability. The methods that could be used
20 % to analyse these systems could be by the usage of graphical or other
21 % method. In this first category we could find the Logan graphical method
22 % and the Patlak graphical method.
23 %
24 % GRAPHICAL METHOD - LOGAN PLOT
25 % In this laboratory we will see the Logan graphical method which is based
26 % on a compartmental model based on just reversible structure (relaxation
27 % of Patlak constraint). With this method it is possible to estimate the
28 % Volume of distribution (Vd) and the binding potential (BP) of the i-th
29 % voxel analysed.
30 % To estimate properly these quantities we need to inspect the Logan plot.
31 % This is a chart where in the X axis and Y axis respectively we could find
32 % these down below reported terms:
33 %
34 % 
$$X = \frac{\int_{t=0}^{\infty} C_{PLASMA}(t) dt}{C_{ROI}(t^\circ)}$$

35 %
36 %
37 %
38 % 
$$Y = \frac{\int_{t=0}^{\infty} C_{ROI}(t) dt}{C_{ROI}(t^\circ)}$$

39 %
40 %
41 %
42 % With this representation after a certain instant of time usually define
43 % as  $t^*$  the curve becomes a straight line. We need first to identify this
44 %  $t^*$  and associate to it the index after that the straight line condition
45 % in the chart is satisfied. We have to pay attention in choosing the right
46 % amount of sample where we need to fit the line. This because sometimes in
47 % particular when the number of voxel involved in the analysis is not too
48 % high we could not find a real  $t^*$  that satisfies the previous condition.
49 % For this reason, in these cases, we need to hold information about an
50 % higher number of samples even if the straight line condition in the
51 % overall considered samples is not satisfied.
52 %
53 % ANATOMICAL REGION
54 % The anatomical region where this laboratory is based are:
```

```
55 %
56 % - Cerebellum : It is much smaller than the cerebrum but plays an
57 % important role in motor control.
58 % - Raphe Nuclei: These nuclei act like an autoreceptors in the brain and
59 % regulate the decrease of release of serotonin. It is
60 % located in the latest section of the brainstem.
61 %
62 %
63 % DATA -----
64 %
65 % EXPERIMENTAL SETTINGS
66 % The folder DatIDASB contains PET data (collected at PET Center University
67 % of Pittsburgh, PA, USA) obtained with the [11C]DASB tracer. This tracer
68 % allows to study the serotonin re-uptake foci functionality. These
69 % locations are involved in the regulation of brain serotonin at synaptic
70 % level. Their malfunction has been associated to a wide variety of central
71 % nervous system pathologies such as Schizophrenia, Alzheimer, Parkinson,
72 % anxiety, mood and eating disorders. Antidepressants drugs act on the
73 % serotonin system blocking the serotonin re-uptake over the rafe nuclei,
74 % resulting in a local serotonin concentration enhancement.
75 %
76 % AVAILABLE DATA
77 % The data that are largely used are the one reported below:
78 %
79 % - dynPET_3228 : Dynamical PET data
80 %                   - duration: PET scan ranges [min]
81 %                   - plane17 : 17-th plane axial PET scan [ $\mu$  Ci/cm3]
82 %                               Time x Voxel x Voxel [34 x 128 x 128]
83 %                   - plane28 : 28-th plane axial PET scan [ $\mu$  Ci/cm3]
84 %                               Time x Voxel x Voxel [34 x 128 x 128]
85 % - cer : Cerebellum mask delineating the most important
86 %          cerebellum sections.
87 % - raph : Raphe Nuclei mask delineating the raphe Nulcei
88 %          anatomical region.
89 % - arterial_3228 : Arterial information data
90 %                   - corrected : Metabolite - corrected plasma tracer
91 %                               [ $\mu$  Ci/ml]
92 %                   - uncorrected : Non - metabolite - corrected plasma
93 %                               tracer concetration [ $\mu$  Ci/ml]
94 %                   - t_plasma : Sampling instant of time
95 %
96 % PROGRAM -----
97 %
98 % The program is subdivided in the following steps:
99 %
100 % - DATA - LOAD : Load of all the data inside the
101 %                   workspace
102 %
103 % - TIME - GENERATION : Generation of hemi scan times
104 %
105 % - ROI - SELECTION (X) : Generation of a mask usable to
106 %                           select correctly the X antomical region
107 %
108 % - VOXEL TIME SERIES - : Extraction of all the possible
```

```
109 %      SELECTION (X)                      time series associated to X.  
110 %  
111 %      - LOGAN SPACE - VISUALIZATION : Logan plot of the X anatomical  
112 %          (X)                                region.  
113 %  
114 %      - VOLUME OF DISTRIBUTION -          : Estimation of the volume of  
115 %          ESTIMATION (X)                  distribution for each voxel  
116 %                                              in the X anatomical region.  
117 %  
118 %      - VOLUME OF DISTRIBUTION -          : Estimation of the reference volume of  
119 %          REFERENCE (X)                  distribution for the X anatomical  
120 %                                              the X anatomical region.  
121 %  
122 %      - BINDING POTENTIAL -              : Evalutation of the distribution  
123 %          DISTRIBUTION (X)                in the X anatomical region of the BP.  
124 %  
125 %      - BINDING POTENTIAL -              : Estimation of the reference BP for  
126 %          REFERENCE (X)                  the X anatomical region.  
127 %  
128 %      - BINDING POTENTIAL -              : Estimation of the BP by linear  
129 %          ESTIMATION (X)                regression model.  
130  
131 %% CCC  
132 % Clear all, close all, clear the comand window  
133  
134 close all, clear all, clc  
135  
136 %% DATA - LOAD  
137 % Load of all the possible data necessary for this Laboratory  
138  
139 load("arterial_3228.mat"), load("dynPET_3228.mat"),  
140 load("cer.mat"),           load("raphe.mat")  
141  
142 %% TIME - GENERATION  
143 % Time vector generation, we use the information contained inside the  
144 % vector called duration, we decrease this quantity of duration half.  
145 %  
146 % duration   :      Vector that quantify the duration of the measure  
147 %                   by the usage of PET technology.  
148 % time       :      Vector of instant of time. The instant of time is  
149 %                   always referred as the middle instant of time of  
150 %                   the time interval analysed.  
151  
152 time_PET = cumsum(duration) - duration/2;  
153  
154  
155 %% CEREBELLUM -----  
156 % Cerebellum elaboration section  
157  
158 %% ROI - SELECTION (CEREBELLUM)  
159 % Cerebellum selection. To acquire correctly the cerebellum as the ROI of  
160 % our analysis we need to inspect the histogram of the image and choose the  
161 % right threshold to select just the lighter structure in the image. The  
162 % methodology provide for choose the Cerebellum are those one:
```

```
163 %
164 % - Visual inspection
165 % - Otsu algorithm
166 % - Roipoly
167 %
168 % To enhance the structure of the cerebellum we implement some preliminary
169 % enhancement algorithm.
170
171 SP17 = squeeze(sum(plane17,1)); % Temporal summation
172 SP17(SP17<0) = 0; % Negative activity deactivation
173
174 GG = fspecial('gaussian'); % Gaussian mask
175 GL = fspecial('laplacian'); % Laplacian mask
176 SP17F = imfilter(SP17 ,GG); % Gaussian filtering
177 SP17F = imfilter(SP17F,GL); % Laplacian filtering
178 SP17E = 2*SP17 + SP17F; % Enhance imaging
179
180 %% VISUALIZATION
181 % Visualization of the previous procedure to enhance the cerebellum
182
183 figure, subplot(1,2,1), imagesc(SP17E),
184         title('IMAGE')
185         colormap 'hot'; colorbar
186         subplot(1,2,2), histogram(SP17E),
187         title('HISTOGRAM'), grid minor, grid on,
188         xlabel('P.I.'), xlabel('A.F.')
189 sgttitle('View: Axial - Slice: 17 - Cerebellum')
190
191 %% ROI - SELECTION (CEREBELLUM)
192 % We apply the threshold estimated by visual inspection of the previous
193 % chart and we see if activate just the cerebellum structure and we see if
194 % also in the image histogram the threshold choose separate two distinct
195 % distribution.
196
197 TH = 15; MK17 = SP17E > TH;
198 figure, subplot(1,2,1), imagesc(SP17E.*MK17),
199         title('IMAGE THRESHED')
200         colormap 'hot'; colorbar,
201         subplot(1,2,2), histogram(SP17E), xline(TH,'r--','LineWidth',2)
202         title('HISTOGRAM'), grid minor, grid on,
203         xlabel('P.I.'), ylabel('A.F.')
204 sgttitle('View: Axial - Slice: 17 - Cerebellum')
205
206 % ALTERNATIVE:
207 %
208 % An alternative to find the region of the cerebellum is to use a first
209 % threshold and then the successive lines of code.
210 %
211 % CC = bwconncomp(SP17 > TH);
212 % numPixel = cellfun(@numel,CC,PixelIdxList);
213 % [biggest, idx] = max(numPixel);
214
215 %% VOXEL TIME SERIES - SELECTION (CEREBELLUM)
216 % Cerebellum voxel time series extraction and visualization. Extraction of
```

```
217 % all the possible time series associated to voxel belonging to cerebellum.
218
219 nTIME = size(plane17,1); % Number of instant of time
220 IPTS17 = find(MK17); % Active positions
221 PTS17 = zeros(length(IPTS17),nTIME); % Initialization
222
223 for idT = 1:1:nTIME % Activity extraction cycle
224     temp = squeeze(plane17(idT,:,:));
225     PTS17(:,idT) = temp(IPTS17);
226 end
227
228 figure, hold on,
229     for idV = 1:1:length(IPTS17), plot(time_PET, PTS17(idV,:)), end
230     plot(time_PET, mean(PTS17, 1), 'ro-', 'LineWidth',2),
231     hold off, grid on, grid minor, box on,
232     title('View: Axial - Slice: 17 - Cerebellum - Voxel Activity'),
233     xlabel('Time [ms]'), ylabel('Concentration [\muCi/ cm^3]')
234     xlim([time_PET(1), time_PET(end)]),
235     yline(0,'k-','LineWidth',2)
236
237 % All the reconstruction algorithm during the process of data could be
238 % affected from noise and they develop values which are strictly negative
239 % which have no significant physiological meaning. So, when we will perform
240 % some analysis with compartmental model we need to neglect and
241 % not considered them during the estimation process.
242
243 %% LOGAN SPACE - VISUALIZATION (CEREBELLUM)
244 % Visualization of the Logan curve to inspect for which instant of time the
245 % curve obtained by plotting the ratio between the integral of
246 % concentration in the ROI and the concentration in the ROI and the
247 % integral of concentration in plasma over the concentration in the ROI is
248 % subsequently linear.
249
250 Cp = interp1(t_plasma, corrected, time_PET);
251 Cm17 = mean(PTS17,1)';
252 ICp = cumtrapz(time_PET,Cp);
253 ICm17 = cumtrapz(time_PET,Cm17);
254
255 figure, plot(ICp./Cm17,ICm17./Cm17,'ob-','LineWidth',1),
256     grid on, grid minor, box on,
257     title('LOGAN SPACE - Cerebellum'),
258     xlabel('C_{PLASMA}/C_{ROI}'),
259     ylabel('C_{ROI}/C_{ROI}')
260     axis([0, max(ICp./Cm17), 0, max(ICm17./Cm17)])
261
262 T17 = 23; % Position
263
264 %% VOLUME OF DISTRIBUTION - ESTIMATION (CEREBELLUM)
265 % Estimation of the volume of distribution inside the cerebellum
266
267 nVOXEL = size(PTS17,1);
268 p17 = zeros(2,nVOXEL);
269
270 for idV = 1:1:nVOXEL % Activity extraction cycle
```

```
271     Cm = PTS17(idV,:)';
272     ICm = cumtrapz(time_PET,Cm);
273
274     X = ICp./Cm; Y = ICm./Cm;
275     G = [X(T17:end), ones(size(X(T17:end),1),1)];
276     p17(:,idV) = ((G'*G)^(-1)*G'*Y(T17:end));
277 end
278
279 Vd17 = nan(size(SP17));
280 Vd17(IPTS17) = p17(1,:);
281
282 figure, imagesc(Vd17), colormap 'hot'; colorbar
283 title('View: Axial - Slice: 17 - Cerebellum - V_d distribution [ml/cm^3]')
284
285 %% VOLUME OF DISTRIBUTION - REFERENCE (CEREBELLUM)
286 % Evaluataion of the volume of distribution reference by the usage of the
287 % cer.mat file which contains the mask for cerebellum areas.
288
289 nVOXEL = sum(cer,'all');
290 VdREF = nansum(Vd17.*cer,'all')/nVOXEL;
291
292 %% BINDING POTENTIAL - DISTRIUTION (CEREBELLUM)
293 % Evaluation of the Binding Pontential of the image.
294
295 BP17P = (Vd17 - VdREF.*MK17) / VdREF;
296 BP17P(BP17P < 0) = NaN;
297
298 figure, imagesc(BP17P), colormap 'jet'; colorbar
299 title('View: Axial - Slice: 17 - Cerebellum - BP distribution []')
300
301 %% BINDING POTENTIAL - REFERENCE (CEREBELLUM)
302 % Evalutaion of the reference BP of the Cerebellum
303
304 nVOXEL = sum(MK17,'all');
305 BPREF = nansum(BP17P,'all')/nVOXEL;
306
307 %% VISUALIZATION
308 % Visualization of all the indexes estimated in previous section
309
310 fprintf('+++++++\n')
311 fprintf('CEREBELLUM: \n')
312 fprintf(' - Volume Distribution Ref: %f [ml/cm^3]\n',VdREF)
313 fprintf(' - Binding Potential Ref : %f []\n',BPREF)
314 fprintf('+++++++\n')
315
316
317
318 %% RAPHE NUCLEI -----
319 % Raphe Nuclei elaboration section
320
321 %% ROI - SELECTION (RAPHE NUCLEI)
322 % We need to select the Raphe nuclei. To do so we use the roipoly method
323 % beacuse it is the only one that let use define a region close to the
324 % center of the image. The selection so is done manually.
```

```
325
326 SP28 = squeeze(sum(plane28)); % Temporal summation
327 MK28 = SP28 > 7; % Thresholding
328
329 figure, subplot(1,2,1), imagesc(SP28.* (MK28)),
330 colormap 'jet'; colorbar, title('IMAGE'),
331 %raphe = roipoly;
332 subplot(1,2,2), title('IMAGE - THRESH')
333 imagesc(raphe.*SP28), colormap 'jet'; colorbar,
334 sgttitle('View: Axial - Slice: 28 - Raphe Nuclei')
335
336 % Pay attention that if we plot alone the sum PET of the slice 28 there are
337 % negative value and also a blue border. This are not physiological and
338 % usually is due to the motion correction algorithm.
339
340 %% VOXEL TIME SERIES - SELECTION (RAPHE NUCLEI)
341 % Raphe Nuclei voxel time serires extraction and visualization. Extraction
342 % of all the possible time series associated to voxel belonging to
343 % cerebellum.
344
345 nTIME = size(plane28,1); % Number of instant of time
346 IPTS28 = find(MK28); % Active positions Mask
347 PTS28 = zeros(length(IPTS28),nTIME); % Initialization
348
349 for idT = 1:1:nTIME % Activity extraction cycle
350     temp = squeeze(plane28(idT,:,:));
351     PTS28(:,idT) = temp(IPTS28);
352 end
353
354 figure, hold on,
355     for idV = 1:1:length(IPTS28), plot(time_PET, PTS28(idV,:)), end
356     plot(time_PET, mean(PTS28,1), 'ro-','LineWidth',2),
357     hold off, grid on, grid minor, box on,
358     title('View: Axial - Slice: 28 - Raphe Nuclei - Voxel Activity'),
359     xlabel('Time [ms]'), ylabel('Concentration [\muCi/ cm^3]')
360     xlim([time_PET(1), time_PET(end)]),
361     yline(0,'k-','LineWidth',2)
362
363 % Due to the lower number of voxel averaged over instant of time we obtain
364 % a worse behaviour of the concentration curve. We have to remember that
365 % the average processing delete the eventual noise overlay to the signal.
366
367 %% LOGAN SPACE - VISUALIZATION (RAPHE NUCLEI)
368 % Visualization of the Logan curve to inspect for which instant of time the
369 % curve obtained by plotting the ratio between the integral of
370 % concentration in the ROI and the concentration in the ROI and the
371 % integral of concentration in plasma over the concentration in the ROI is
372 % subsequently linear.
373
374 Cp = interp1(t_plasma, corrected, time_PET);
375 Cm28 = mean(PTS28,1)';
376 ICP = cumtrapz(time_PET,Cp);
377 ICm28 = cumtrapz(time_PET,Cm28);
378
```

```
379 figure, plot(ICp./Cm28,ICm28./Cm28,'ob-','LineWidth',1),  
380     grid on, grid minor, box on,  
381     title('LOGAN SPACE - Raphe Nuclei'),  
382     xlabel('C_{PLASMA}/C_{ROI}'),  
383     ylabel('C_{ROI}/C_{ROI}')  
384     axis([0, max(ICp./Cm28), 0, max(ICm28./Cm28)])  
385  
386 T28 = 26;                                % Position  
387  
388 %% VOLUME OF DISTRIBUTION - ESTIMATION (RAPHE NUCLEI)  
389 % Estimation of the volume of distribution inside the Raphe Nuclei  
390  
391 nVOXEL = size(PTS28,1);  
392 p28 = zeros(2,nVOXEL);  
393  
394 for idV = 1:1:nVOXEL                  % Activity extraction cycle  
395     Cm = PTS28(idV,:);  
396     ICm = cumtrapz(time_PET,Cm);  
397  
398     X = ICp./Cm; Y = ICm./Cm;  
399     G = [X(T28:end), ones(size(X(T28:end),1),1)];  
400     p28(:,idV) = ((G'*G)^(-1)*G'*Y(T28:end));  
401 end  
402  
403 Vd28 = nan(size(SP28));  
404 Vd28(IPTS28) = p28(1,:);  
405  
406 figure, imagesc(Vd28), colormap 'hot'; colorbar  
407 title('View: Axial - Slice: 28 - Raphe Nuclei - V_d distribution [ml/cm^3]')  
408  
409 %% VOLUME OF DISTRIBUTION - REFERENCE (RAFE NUCLEI)  
410 % % Evaluation of the volume of distribution reference by the usage of the  
411 % % cer.mat file which contains the mask for cerebellum areas.  
412 %  
413 % nVOXEL = length(IPTS28);  
414 % VdREF = sum(Vd28,'all')/nVOXEL;  
415  
416 %% BINDING POTENTIAL - DISTRIBUTION (RAFE NUCLEI)  
417 % Evaluation of the Binding Potential of the image.  
418  
419 BP28P = (Vd28.*MK28 - VdREF.*MK28) / VdREF;  
420 BP28P(BP28P < 0) = NaN;  
421  
422 figure, imagesc(BP28P), colormap 'jet'; colorbar  
423 title('View: Axial - Slice: 28 - Raphe Nuclei - BP distribution []')  
424  
425 %% BINDING POTENTIAL - REFERENCE (RAFE NUCLEI)  
426 % Evalutaion of the reference BP of the Raphe Nuclei  
427  
428 nVOXEL = length(IPTS28);  
429 BPREF = nansum(BP28P,'all')/nVOXEL;  
430  
431 %% VISUALIZATION  
432 % Visualization of all the indexes estimated in previous section
```

```
433
434 fprintf('+++++++\n')
435 fprintf('RAPHE NUCLEI: \n')
436 fprintf(' - Volume Distribution Ref: %f [ml/cm^3]\n',VdREF)
437 fprintf(' - Binding Potential Ref : %f []\n',BPREF)
438 fprintf('+++++++\n')
439
440
441 %% VISUALIZATION
442 % Visualization of the results in the same window
443
444 figure(), h(1) = subplot(1,2,1); imagesc(BP17P), colormap 'jet'; colorbar
445 title('BP - Slice 17'),
446 h(2) = subplot(1,2,2); imagesc(BP28P), colormap 'jet'; colorbar
447 title('BP - Slice 28'),
448 set(h,'Clim',[0 3.0643])
449
450
451 %% BINDING POTENTIAL - ESTIMATION (CEREBELLUM)
452 % Estimation of the binding potential inside the cerebellum
453
454 nVOXEL = size(PTS17,1);
455 p17    = zeros(2,nVOXEL);
456
457 T17 = 26;
458
459 for idV = 1:1:nVOXEL                         % Activity extraction cycle
460     Cm = PTS17(idV,:)';
461     ICm = cumtrapz(time_PET,Cm);
462
463     X = ICm28./Cm; Y = ICm./Cm;
464     G = [X(T17:end), ones(size(X(T17:end),1),1)];
465     p17(:,idV) = ((G'*G)^(-1)*G'*Y(T17:end));
466 end
467
468 BP17R = nan(size(SP17));
469 BP17R(IPTS17) = p17(1,:) - ones(1,size(p17,2));
470 BP17R(BP17R < 0) = NaN;
471
472 figure, imagesc(BP17R), colormap 'jet'; colorbar
473 title('View: Axial - Slice: 17 - Cerebellum - BP distribution model []')
474
475 %% BINDING POTENTIAL - ESTIMATION (RAPHE NUCLEI)
476 % Estimation of the binding potential inside the Raphe Nuclei
477
478 nVOXEL = size(PTS28,1);
479 p28    = zeros(2,nVOXEL);
480
481 for idV = 1:1:nVOXEL                         % Activity extraction cycle
482     Cm = PTS28(idV,:)';
483     ICm = cumtrapz(time_PET,Cm);
484
485     X = ICm28./Cm; Y = ICm./Cm;
486     G = [X(T28:end), ones(size(X(T28:end),1),1)];
```

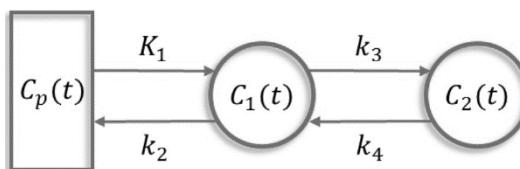
```
487 p28(:,idV) = ((G'*G)^(-1)*G'*Y(T28:end));
488 end
489
490 BP28R = nan(size(SP28));
491 BP28R(IPTS28) = p28(1,:) - ones(1,size(p28,2));
492 BP28R(BP28R < 0) = NaN;
493
494 figure, imagesc(BP28R), colormap 'jet'; colorbar
495 title('View: Axial - Slice: 28 - Raphe Nuclei - BP distribution model []')
496
497 %% VISUALIZATION
498 % Visualization of the results in the same window
499
500 figure(), h(1) = subplot(1,2,1); imagesc(BP17R), colormap 'jet'; colorbar
501 title('BP - Slice 17'),
502 h(2) = subplot(1,2,2); imagesc(BP28R), colormap 'jet'; colorbar
503 title('BP - Slice 28'),
504 set(h,'Clim',[0 max(BP28(:))])
505
506
507 %% VISUALIZATION
508 % Complete visualization of all the results obtained.
509
510 M = max([max(BP17R(:)) max(BP17P(:)) max(BP28R(:)) max(BP28P(:))]);
511 m = min([min(BP17R(:)) min(BP17P(:)) min(BP28R(:)) min(BP28P(:))]);
512
513 figure(), h(1) = subplot(2,3,1); imagesc(BP17P), colormap 'jet'; colorbar
514 h(2) = subplot(2,3,2); imagesc(BP17R), colormap 'jet'; colorbar
515 d(1) = subplot(2,3,3); imagesc((BP17R - BP17P)./BP17P)*100
516 colormap 'jet'; colorbar,
517 h(3) = subplot(2,3,4); imagesc(BP28P), colormap 'jet'; colorbar
518 h(4) = subplot(2,3,5); imagesc(BP28R), colormap 'jet'; colorbar
519 d(2) = subplot(2,3,6); imagesc((BP28R - BP28P)./BP28P)*100
520 colormap 'jet'; colorbar,
521
522 set(h,'Clim',[m M])
523
524 % Much more simplification we do much more bias we introduce in our
525 % analysis.
526
```

LESSON 8: PET BASIS FUNCTION METHOD, SPECTRAL ANALYSIS

Graphical method: Patlak and Logan general consideration

In PET every parameter has a direct link to the physiology. In MR, the things are a little bit different because the information retrieved are not directly linked with physiology but more with anatomy. In the last lesson we were talking about the two types of graphical method, the Logan and the Patlak. These techniques let be able to estimate just macroparameter represented by the binding potential and the distribution volume respectively. Consequently, when we look to PET data with graphical methods, we always refer to macroparameter and microparameter. The first are composition and mathematical expression of the seconds and usually are the only ones that are calculable.

There is a very big difference between Patlak and Logan graphical method. This could be seen in the coordinate used to describe the curve. In fact, in Patlak at the denominator of each coordinate x and y there is the blood concentration which is assumed noise free. Obviously, this could be taken as true only if the estimation process of the blood sample is done correctly at the beginning of the analysis. Logan instead, present in denominator the measured concentration by PET techniques which is always affected by noise. Knowing also exactly the structure of the model we can find the exact relation between the micro and macro parameter. By referring to the next structure for compartmental model we could write the equations:



$$V_d = \frac{K_1}{k_2} \left(1 + \frac{k_3}{k_4} \right)$$

$$BP = \frac{k_3}{k_4}$$

$$V_d^{ref} = \frac{K_1}{k_2}$$

$$BP = \frac{V_d - V_d^{ref}}{V_d^{ref}}$$

Where the reference volume distribution is computed over the reference region which is tracer specific and usually does not present the second compartment. The model of this region is represented by a single compartment model because the region does not present the specific receptor tracer. Moreover, the distribution volume obtained, which analytical expression is reported above, is assumed to be function of the fractional transfer coefficient K_1 and k_2 which are assumed to be equal in the reference and analysed region. The concept of reference region is not just a concept of receptor investigation, but it is used in general in PET. Reference region is a location in the brain that is well defined and delimited (sometimes also anatomically) without specific receptors exempli gratia cerebellum, which is devoid of serotonergic receptors, and therefore it is taken as a reference in the studies on the serotonergic system. Pay attention that assuming K_1 and k_2 equal in the reference and in the analysed region is not a very strong constraint, because the value of these two fractional transfer coefficients depends on the exchange process between the blood flow and the brain region. Usually, over all the voxel these two items do not change too much and the only situation where the value of K_1 could be altered is due to some pathological condition of the patient which is a priori known.

Graphical method: Patlak and Logan pros and cons

Even for this type of analysis exists some pros and some cons. Starting with the first ones we could say that these two techniques are very fast because based on simple principle and linear estimator that does not require additional results from different process. Moreover, they do not require a specific compartmental model to represent a specific area of the brain and thanks to the presence and the definition of the reference region make their use very convenient. As cons instead, we have a such strong simplification of the brain region analysed and we can obtain just the value of the macro parameters and not the one for the micro parameters. Moreover, Patlak does not show a bias in the estimates while Logan of course and the estimates, for both two methods, depends heavily on the choice of t^* .

Simplified reference tissue model

There are other approaches that we can consider for the quantification of the PET images instead of just the two graphical methods that we have seen before. To introduce the new techniques, we need to do some additional hypothesis. These are:

- Exists a reference region in which the kinetic could be described with a mono compartmental model.
- Other regions and voxels have a kinetic described with the classic two – compartmental model.
- Finally, the value of $R_1 = \frac{K_1}{K'_1}$ and $\frac{K_1}{k_2} = \frac{K'_1}{k'_2}$

This model is based on estimate the binding potential (BP) and exploits only the tissue information (without using C_p). The relation that describes the concentration in the tissue is expressed with the next equation:

$$C_t(t) = R_1 C_r(t) + \left[k_2 - \frac{k_2 R_1}{1 + BP} \right] C_r(t) \otimes e^{-\frac{k_2}{1+BP}t}$$

Where $C_t(t)$ represent the concentration of the region that we want to describe. The $C_r(t)$ instead represent the concentration in the reference region. This model that is derived when the previous three assumption are true and the analytical expression allows us to find out the parameters R_1 , k_2 and BP . Obviously, the most logical answer to estimate all the parameters at once is by using the *NLWLS*, but due to the high computation cost of the technique and the initial value solution dependency, we may use different way to solve and estimate all the parameters. The two different strategies that could be implemented are the basis function method and the linearization of the model. Let's see the first techniques.

Simplified reference tissue model: Basic function method

By taking in consideration the previous expression we can define three main parameters which could simplify a lot the analytical expression and more important they are function of the three parameters of the model.

$$C_t(t) = R_1 C_r(t) + \left[k_2 - \frac{k_2 R_1}{1 + BP} \right] C_r(t) \otimes e^{-\frac{k_2}{1+BP}t} \Rightarrow C_t(t) = \theta_1 C_r(t) + \theta_2 C_r(t) \otimes e^{-\theta_3 t}$$

Subsequently the definition of θ_1 , θ_2 and θ_3 are respectively:

$$\theta_1 = R_1 \quad \theta_2 = k_2 - \frac{k_2 R_1}{1 + BP} \quad \theta_3 = \frac{k_2}{1 + BP}$$

The equation that is reported before, is linear in the parameters θ_1 and θ_2 but it is still not linear in the parameter θ_3 . The simplification of this methods now leads to a fruitful result and it is not a wrong thing to perform to get some appreciable results. The basic concept relay on the definition of a grid of N values for the parameter θ_3 . If these is settled to a specific value, the shape and behaviour of the exponential is completely known. Hence, we can write:

$$\theta_{3i} \Rightarrow \theta_3^{min} \leq \theta_{3i} \leq \theta_3^{max} \quad \forall i \in \{1, \dots, N\}$$

The bounds are set by some in vitro and in vivo experiment and needs to satisfy these constraints:

$$\theta_3^{min} \leq \frac{k_2^{min}}{1 + BP_{MAX}} \quad \theta_3^{max} \geq k_2^{max}$$

At this point, once define N component for the value of θ_3 we can also define N possible function, also called *basis function*, which differ one from the other over the shape of the exponential of θ_3 . Thus:

$$B_i(t) = C_r(t) \otimes e^{-\theta_{3i} t} \quad \forall i \in \{1, \dots, N\}$$

By substituting this to the main equation we can write:

$$C_t(t) = \theta_1 C_r(t) + \theta_2 B_i(t) \quad \forall i \in \{1, \dots, N\}$$

The equation now is liner for θ_1 and θ_2 and we can use a linear least square estimators to determine the value of each parameter. Thanks to the fact that this problem is translated into a linear problem we can write the previous equation by a matrix vector representation:

$$A_i = \begin{bmatrix} C_r(t_1) & B_i(t_1) \\ C_r(t_2) & B_i(t_2) \\ \vdots & \vdots \\ C_r(t_M) & B_i(t_M) \end{bmatrix} \in R^{M \times 2} \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \in R^{2 \times 1} \quad C_t = \begin{bmatrix} C_r(t_1) \\ C_r(t_2) \\ \vdots \\ C_r(t_M) \end{bmatrix} \in R^{M \times 1}$$

Where M indicates the number of scans acquired during the PET analysis. With the previous notation the equations could be written in the next way:

$$C_t = A_i \theta$$

Thanks to the matric notation and the linear structure of the model we can easily use a linear estimator and find the right value to attribute to the vector of parameter θ .

$$\hat{\theta} = [A_i^T W A_i]^{-1} A_i^T W C_t$$

Where W is a diagonal matrix and belongs to the space $R^{M \times M}$ and represents the weights of each measurements. By associating to it a specific meaning it represent the covariance matrix of the error. The algorithm is now very simple because we start with an initial value for θ_3 and we estimate all the parameter of the model or θ_1 and θ_2 . Once this step is completed, we evaluate the WRSS which is a good element that represent the goodness of the fit with that specific combination of parameter. This procedure is done and performed for all the values of θ_3 that initially have decided to investigate. To choose the best answer we need to minimize the cost function WRSS previously defined and calculated for all the inspected value of θ_3 . After having found the best set of parameters θ_1, θ_2 and θ_3 through the previous paradigm we can apply the inverse relation and estimate the possible micro and macroparameter of the model by the following expressions:

$$\hat{R}_1 = \hat{\theta}_{1i} \quad \hat{k}_2 = \hat{\theta}_{2i} + \hat{R}_1 \hat{\theta}_{3i} \quad \widehat{BP} = \frac{\hat{k}_2}{\hat{\theta}_{3i}} - 1$$

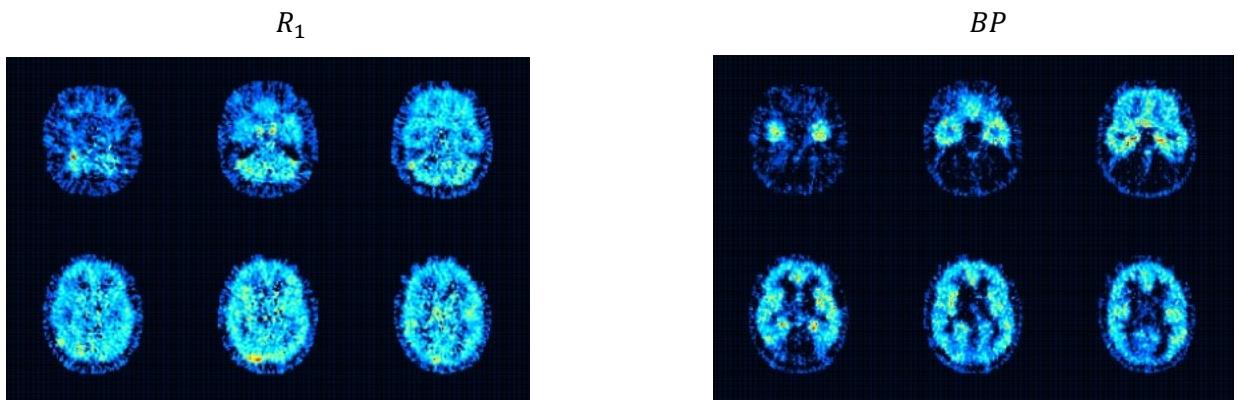
A reasonable number of initial values of θ_3 to try are at least 100 or 200 if we perform an analysis at the voxel level. If we consider a higher number as instance 1000 or more, we are not able to compute them all correctly because the computational effort for an analysis at the voxel level is too high. One other important observation is that the complexity of the model is not very strong and for this reason we will not fall in overestimation phenomenon even if we are searching for the combination of parameters that minimize the WRSS cost function. Hence, the weighted residuals would never show an overfitting trend.

Simplified reference tissue model: Basic function method pros and cons

With this technique to describe the compartmental model we have a less bias in estimation of the BP instead of Logan graphical method. The constraint of the value of θ_3 are much more not reliable. It is characterized by computational stability and finite time elaboration. The BFM estimates are such equal to the weighted nonlinear quantization estimates. As cons of this technique is the necessity of some a priori knowledge of k_2 and the binding potential to define the grid with N values for θ_3 . The final estimates depend on the discretization grid of N values for θ_3 . Finally, the precision of estimates θ_1 and θ_2 are biased.

Simplified reference tissue model: Basic function method parametric map

These are only some examples of parametric map of serotoninergic tracer using the reversible reference tissue model. From the image we can clearly see the presence of value at maximum equal to 1 and all the spikes in the figure are always related with blood flow. In this case each component are unitless.



Simplified reference tissue model: Model linearization

This method relies on the linearization of the previous defined model. In fact, the output equation used in the basis function method is the expression of the measurement made with PET. This equation could be simplified and linearized by the usage of the basic differential equation describing the simplified reference tissue model. These are:

$$\begin{aligned}\frac{dC_T(t)}{dt} &= K_{1T}C_p(t) - k_{2a}C_T(t) \\ \frac{dC_R(t)}{dt} &= K_{1R}C_p(t) - k_2C_R(t)\end{aligned}$$

Where the subscript T represents the voxel with the specific receptor, while R subscript letter represents the reference region. The equation of the concentration of tracer inside the tissue C_t could be derived from the two previous equations making the two further assumptions.

$$\frac{K_{1T}}{k_{2T}} = \frac{K_{1R}}{k_{2R}} \quad k_{2a} = \frac{k_{2T}}{1 + BP}$$

By integrating the previous differential equations, we can obtain the expression:

$$\begin{aligned}C_T(t) &= K_{1T} \int_0^t C_p(\tau) d\tau - k_{2a} \int_0^t C_T(\tau) d\tau \\ C_R(t) &= K_{1R} \int_0^t C_p(\tau) d\tau - k_{2R} \int_0^t C_R(\tau) d\tau\end{aligned}$$

We can substitute the integral of the C_p obtained from the second equation in the first one and we can obtain:

$$C_T(t) = \frac{K_{1T}}{K_{1R}} C_R(t) + \frac{K_{1T}k_{2R}}{K_{1R}} \int_0^t C_R(\tau) d\tau - k_{2a} \int_0^t C_T(\tau) d\tau$$

Defined the R as and exploiting on the assumptions made in the model:

$$R = \frac{K_{1T}}{K_{1R}} \quad \frac{K_{1T}}{k_{2T}} = \frac{K_{1R}}{k_{2R}} \Rightarrow k_{2T} = \frac{K_{1T}k_{2R}}{K_{1R}}$$

The previous integral equation could be written as:

$$C_T(t) = RC_R(t) + k_{2T} \int_0^t C_R(\tau) d\tau - k_{2a} \int_0^t C_T(\tau) d\tau$$

This last equation can be used to estimate the parameter values associated to R , k_{2T} and finally k_{2a} . This is just because of the linearization of the model. By estimating these parameters, we could also calculate the value of the binding potential BP with the next analytical relation:

$$BP = \frac{k_{2T}}{k_{2a}} - 1$$

The previous integral equation is linear in the parameter and therefore admit closed form solution. However, one of the columns on the X matrix (3rd) contains the integral of PET measures and therefore it is affected by noise. The consequence is that the closed form solution of linearized models provides biased parameter. The level of the bias depends on the level of the noise. If this one increase also the level of the bias increases too. This bias is negligible at the ROI level but could be significant at the voxel level. This problem could be solved by performing first the analysis at the ROI level and after performing the analysis at the voxel level by using as priori information the estimates coming from the analysis performed and mentioned before.

Spectral analysis

This method is very powerful because could find out some important information. It was first proposed by Cunningham in 1993 and nowadays this is a widely used instrument specifically here in Padova. It is very useful in PET quantification and enhanced the most informative data driven approaches. This is not a fully mathematically descriptive as the model – based methods are, but still give us such informative results. This uses a sort of black box modelling approach instead of the grey box situation used in compartmental model approaches. In this case, as it is considered a black box based, we don't know any information about the model structure, but we know just the input and the output from that specific model. We consider the input function as the concentration in the plasma and the output as the concentration in the voxel. The aim is to find the value of the parameters that let us be able to define the answer of our system to the unit impulse. This process and way of thinking is very well defined in control system theory and is denominated as forcing function method. This is such a well – defined problem because just with the information coming from the input and the output of the system, we want to estimate the transfer function of our control system. To correctly perform the analysis, we have to make an additional assumption that the input signal is noise free. Now that we have set up the background, we can assume also that we are dealing with compartmental model. This is an important information because let us be able to define the transfer function of the model as sum of exponential functions:

$$h(t) = \sum_{i=1}^M \alpha_i e^{-\beta_i t} \quad \beta_i \geq 0$$

The output of the system could be defined with the next analytical description:

$$C_{voxel}(t) = C_p(t) \otimes h(t) = \sum_{i=1}^M \alpha_i \int_0^t C_p(\tau) e^{-\beta_i(t-\tau)} d\tau \quad \beta_i \geq 0$$

To estimate correctly all the possible parameter of the model we should estimate the value of M and the correspondent values a_i and b_i for each i in the interval between $1, \dots, M$. Thus, the first step to solve the problem is to set the value for M . Usually a high value is chosen as instance 100 or maybe even 200. Then we create a grid of M values of plausible for β_i .

$$\beta_1 = 0, \beta_2 = \frac{1}{3T_{final}} = \frac{1}{\tau_2}, \dots \beta_i = \frac{1}{\tau_i} \dots \beta_M = \frac{3}{T_{first\ scan}} = \frac{1}{\tau_M} \quad \forall i \in 3, \dots, M - 1$$

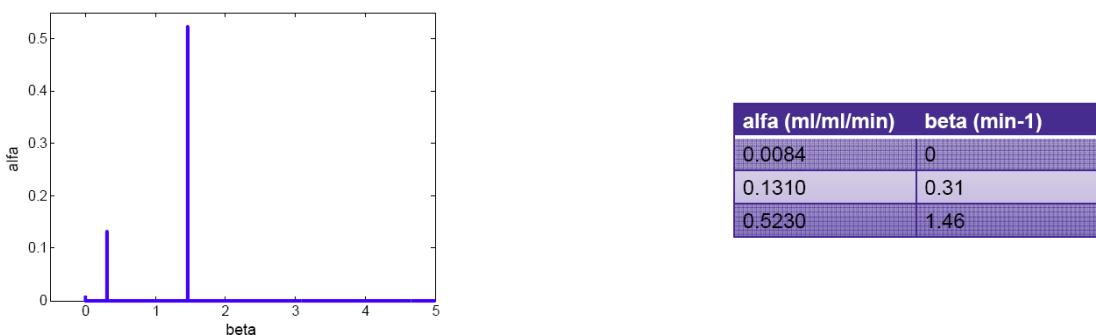
The first β is equal to zero because represent the possibility that the system is irreversible. The expression of the generic τ_i could be written as:

$$\tau_i = \tau_{i-1} \left[\frac{T_{final}}{T_{first\ scan}} \right]^{[1/(1-M)]}$$

After having created the most suitable β grid to the previous application we execute the M convolution integrals where we know the value of $B_i(t)$ (which is the convolution between the concentration in plasma measured and the i -th exponential) and the value of the measured PET concentration in the voxel. The expression is the next one:

$$C_{voxel}(t) = \sum_{i=1}^M \alpha_i B_i(t)$$

Where if we apply a multiple linear regression estimation, we will not obtain good values. This because there are a lot of degree of freedom, and this leads to non-optimal solution and overfitting problems. To induce sparsity researchers have the brilliant idea to put some physiological constraints. These methods are also known as *overcomplete basis*. These assert that all the possible α_i values are always positive semidefinite. This means that they could only assume value greater or equal to zero. Another assumption done is to have considered the measurement error, additive with a gaussian distribution, samples mutually uncorrelated with zero mean and variance define as σ^2 . To estimate correctly all the parameters, we need to apply the non-negative least square method. This constraint makes sense from the compartmental model point of view. Even use this methodology a lot of α has a module which is very close to zero. What we obtain after the estimation process is a spectrum hence the name of the technique.



We could clearly see the presence of spike for certain value of β . If the β are equal to zero but some of the α are not (the one that has no decay), this means that the tracer has at least one irreversible compartment model. Usually, this chart represents and highlight important information. In fact, the high frequency components are related with fast rhythms as the blood flow, while the low frequency component relate to metabolic event. Another important concept to correctly read a spectrum in output from this analysis it is to know that if we found β components which are very close one each other is a phenomenon called doubling and it is due to the discretization of β grid. This could be solved by increasing the specificity of the steps (do smaller steps) or by using posteriori some clustering analysis.

This result could be then translated into a compartmental model. In fact, the number of spectral lines correspond to the number of compartments, and in the situation where a β is equal to zero and the correspondent α is different from zero there is an irreversible compartment. Instead, the β associated to subscript $i = 2, \dots, M$ are always associated to reversible compartments.

We could calculate some macroparameters of the model or the K_i , the V_d and also K_1 . The analytical written:

$$K_i = \alpha_1$$

$$V_d = \sum_{i=2}^M \frac{\alpha_i}{\beta_i}$$

In spectral analysis we could also obtain some information about the V_b , or the fraction of blood volume that interfere with the concentration measure by PET instrumentation. Padova has been an active developer of spectral analysis method (*SAKE tool*), especially apply spectral analysis at the voxel level.

This method is the only one that let us be able to find the right number of the compartment to use to correctly analyse the model. In bioengineering we are asked to understand which is the compartmental model best to use in a specific situation with a specific tracer. Takes at least one year of work.

Bayesian estimator

During the generation of parametric maps there is one possible problem which is common in the most of cases. Due to the high complexity of the initial model structure and the subsequently linearization we induce a lot of bias in the solution. Thus, to solve this problem we need to use the Bayesian estimator that involved the knowledge of some possible means and population values for the vector of parameters. This is based on the Bayes rule that holds:

$$P(\theta|y, M) = \frac{P(y|\theta, M)P(\theta|M)}{P(y|M)}$$

Where all the terms inside this written has a specific meaning. All these last ones are listed down below:

- $P(\theta|y, M)$: Represent the probability that the parameter vector assumes certain value given the value of the observations and the model structure respectively denoted with y and M . It is also called the posterior probability.
- $P(y|\theta, M)$: Represents the probability that the observation vector assumes certain value given the parameter vector and the model structure. This term represents the likelihood.

- $P(\theta|M)$: It is also known as the a priori information and represent the probability that given the model structure M the parameter vector θ assumes certain value.
- $P(y|M)$: This represents the probability of the model evidence.

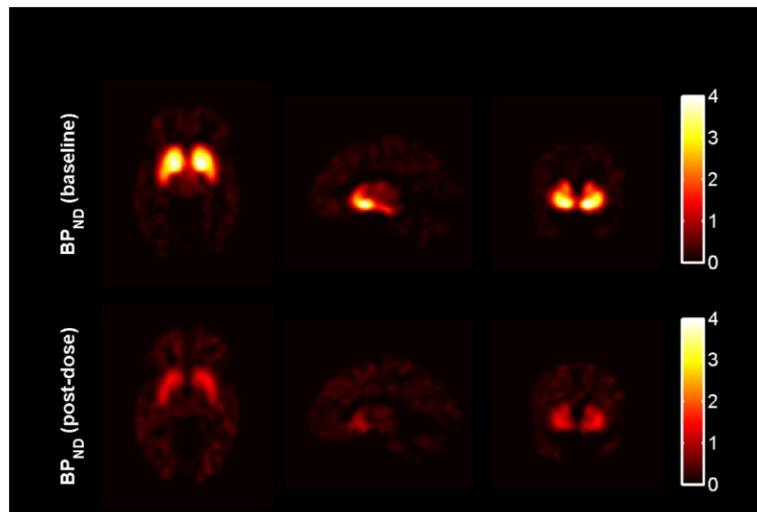
Thus, to main problem move to how implement it analytically. This could be simply done by the usage of the next analytical expression.

$$\hat{p} = \begin{bmatrix} \hat{k}_1 \\ \hat{k}_2 \end{bmatrix} = (G^T \Sigma_v^{-1} G + \Sigma_p^{-1})^{-1} (G^T \Sigma_v^{-1} z + \Sigma_p^{-1} \mu_p)$$

Now the problem is how to estimate the value of Σ_p and μ_p . To understand these two terms we need to consider a 4D image. Perform a clustering analysis as instance k – means and find in each cluster a metric of mean and covariance matrix of the parameter once estimated through a linear least square approach.

Drug receptor occupancy

When we use tracer, we need to remember that it could be used to develop drugs. In the image down below, we could see the presence of different situation of drug placement. We want to understand how the drugs act and interact with all the possible target receptor. This means that we need first to understand if the drug works or not and then we need to do two separate steps. In the first we deliver a certain amount of tracer to investigate how the PET tracer is distributed and what is the body activity before introducing the drug. This tracer needs to be specific for the receptor target of the drug too. After having seen the activity before introducing the drug we need to administer a small dose of it to the patients to modify the cerebral or the targeted activity. The drug is introduced with the tracer in order to enhance the activity and where the drug will connect.



To enhance the modification of the cerebral activity we can use a relative difference and weighting from the initial situation. This is done by the next analytical relation:

$$RO\% = \frac{BP - BP'}{BP}$$

LESSON 9: FUNCTIONAL MAGNETIC RESONANCE AND HDR

Functional Magnetic Resonance Imaging

In this lesson we will talk about the fMRI technique in neuroscience application. This technique is a very wide filed because depending on the contrast agent used, we can highlight different pathways and characteristics. The magnetic resonance imaging community is made up of different specialist of different filed, like physicist, engineer, biologist and finally doctors. There is also an annual meeting organized by the *ISMRM*, which is the acronym for *International Society for Magnetic Resonance in Medicine*, where each year different subjects try to present some innovation and also some news about this technique. This meeting is held worldwide and is such rich of content that every year the amount of information available increases.

fMRI is the acronym of *Functional Magnetic Resonance Imaging*. This technique is based on the BOLD signal mainly, but we can use different types of contrast agent. If this one is a physiological element, and it is contained in the blood it might be the haemoglobin which presents some specific magnetic properties. If we deal with haemoglobin the magnetic resonance image obtained from the scanner is based on the BOLD signal, as previously said. Usually, images strictly connected with BOLD signal are called BOLD images. This acronym stands for *Blood Oxygenation Level Dependent*. The blood oxygenation could be seen by inspecting the haemoglobin concentration near the site of analysis. This biomolecule, in fact, has the properties to carry the oxygen all over our body. It presents some specific magnetic properties which may change as the biomolecule brings oxygen or not. In the first state the haemoglobin is also called oxygenated haemoglobin (HbO) while in the second case it is called deoxygenated haemoglobin (Hb). In the first situation it behaves as a diamagnetic substance which produces a null magnetic momentum while in the second situation or in the state where it does not carry more oxygen it is characterized from paramagnetic properties and so generates a magnetic momentum different from zero. This lead to change the magnetic properties of the entire blood near the site. In fact, the magnetic susceptibility of totally deoxygenated blood increases of about the 20% than magnetic susceptibility of fully oxygenated blood.

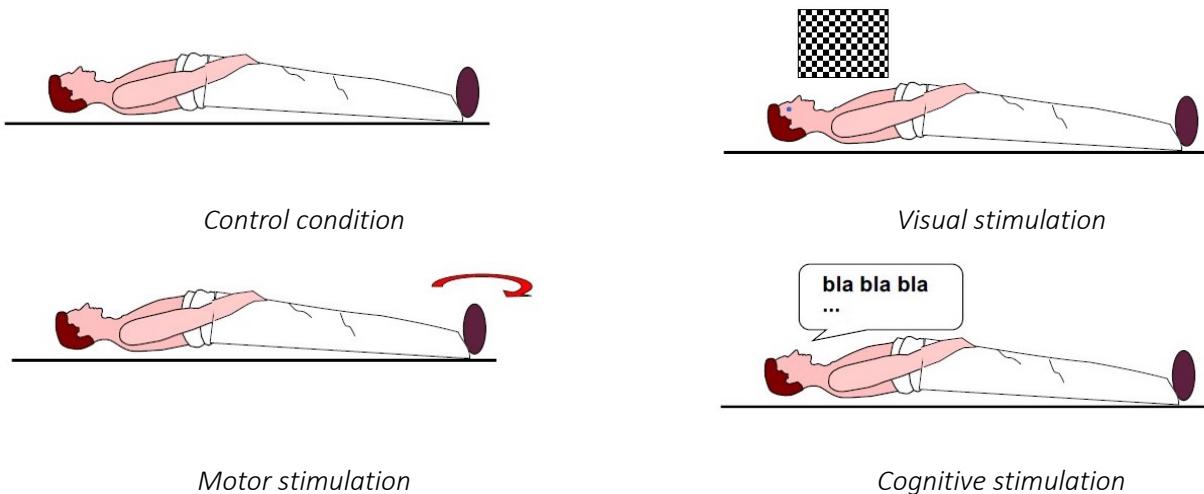
fMRI: Haemoglobin

Thanks to its properties the haemoglobin could be used as an endogenous contrast element. This principle was first observed in the beginning of the 80s at the Bell Laboratories (USA) from the professor Seiji Ogawa. At the time the experiments on animals showed that with magnetic field $B_0 \geq 1.5T$ and $T2^*$ weighted images it is possible to measure a signal that depends on deoxygenated haemoglobin level of a cerebral region. It was observed that just with certain intensity of the static magnetic field applied in the scanner was possible to investigate the properties of the haemoglobin. Nowadays we know that the static magnetic field need to reach at least 1.5T but even the most non – modern scanner today at work reaches this level. This is the reason why this methodology is worldwide used, and every year scientist and engineers keep working on this. We have to keep in mind that it is not the haemoglobin that produces the signal analysed and inspected with MR but are the hydrogen nuclei contained in the water molecule that produces the MR signal. The haemoglobin close to the water molecule changes the magnetic field capture from the nuclei of hydrogen when this substance is deoxygenated. This alteration of the local magnetic field creates all the possible inhomogeneities that generate spin dephasing. This phenomenon decreases the $T2^*$ relaxation time and it could be detected from the MR instrumentation.

The signal detected is mainly related with the altered activity of water molecules from level of blood oxygenation. It means that the signal is indirectly connected with the blood oxygenation, but this last one is strictly connected with the demand of oxygen from a specific brain region. Higher it is the demand higher would be the deoxygenation of the blood in the local site. This leads to the conclusion that the level of the blood oxygenation is strictly connected with cerebral activity. When we have an activity of neurons inside a specific brain area, to maintain the activities and thus generate the energy to keep the activity the blood flow increases in the region involved in the processes. This is due to some cellular process in which are involved the neurons. When these cells start generates action potentials to establish again the normal membrane potential are activated all the possible sodium potassium pumps. These ones are proteins that uses ATP to move inside and outside respectively potassium and sodium ions. To generate sufficiently ATP to correctly conclude this process it is needed glucose and oxygen close to the mitochondria. To answer to the increase demand of oxygen, the blood flow increases disproportionately, and more oxygen is delivered than is needed. Another mechanism that is involved when a certain brain area start working is the vasodilation of new capillaries close to the site.

fMRI: Activation study

We can in this way try to perform activation study. Down below are reported the main control condition and the related event detectable with the fMRI.



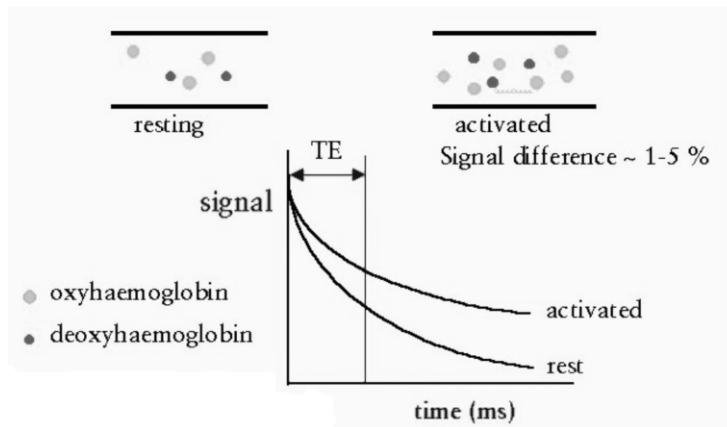
By inspecting through the usage of fMRI of BOLD signal and the activated region we can connect a certain brain region to a specific task. This is something that seems like the PET because we are indirectly inspecting the metabolism of the brain, but this technique presents a higher spatial and temporal resolution. Moreover, the fMRI exposes the patients to some electromagnetic radiofrequency stimulation that is always less toxic than administer to a subject a radioactive drug.

The fMRI activation study are based on a cascade of events. The stimulation increases the rCBF (*Regional Cerebral Blood Flow*) and this would lead to an higher oxygen extraction. While this action is occurring the unbalance between the Hb and HbO raises and to answer to this variation of the ratio blood flow increases

with also vasodilation of capillary close to site. All this cascade of events produces the $T2^*$ weighted images that we record. Down below in the scheme is reported briefly what happens when a region activates.

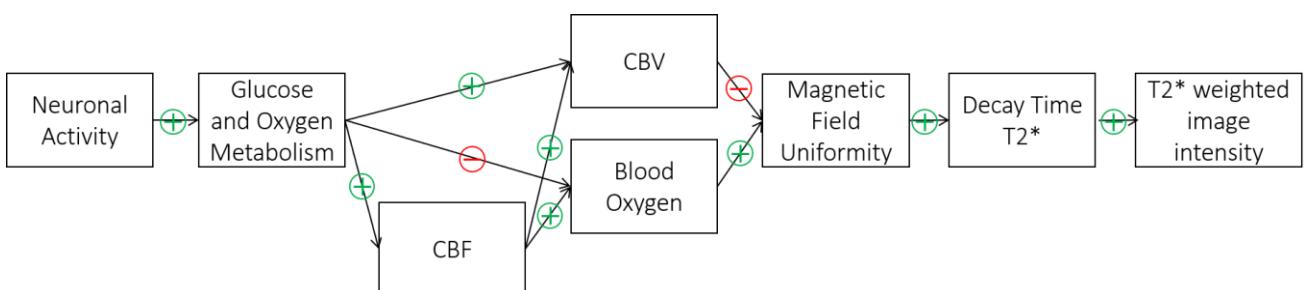


Down below in the scheme is reported what happens in the blood vessels close to the activated site. In this case we have a much more amount of HbO and a higher amount of Hb. This last decreases with time because is substituted with new HbO. After the echo time during a $T2^*$ weighted imaging acquisition the signal detected in the rest and activated tissue are very different and thus we can recognize which area has been activated.



This type of studies as already said before were mainly performed with PET inspection by tracing the water molecule and the glucose with some unstable radioactive atoms. The first BOLD fMRI experiment was performed in 1992 and this open a wide area of analysis in fact, the future years was based on developing and studying different area of the brain and link them to a specific function.

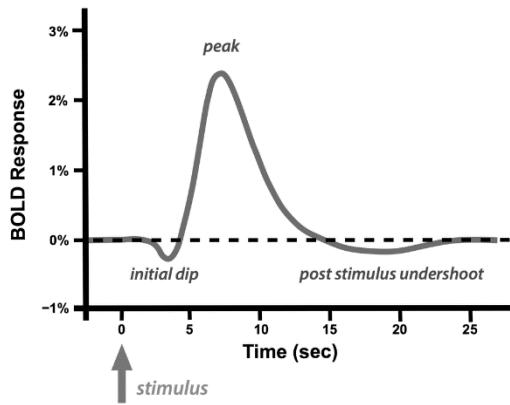
The activity of the neuron and all the biological signals and feedback and the acquisition of the fMRI signal could be described with the down below reported scheme.



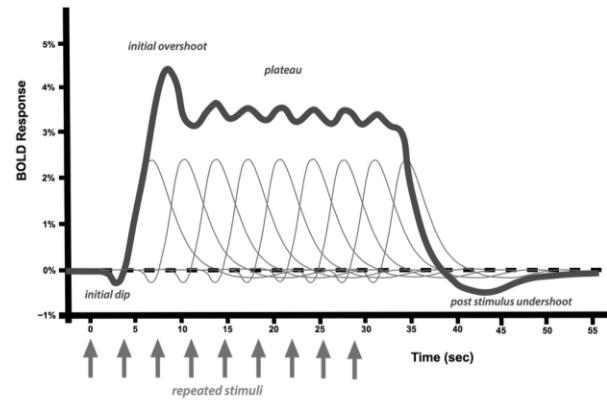
In this representation we have some acronym that need explanation, the *CBV* or *Cerebral Blood Volume* and the *CBF* or *Cerebral Blood Flow*.

fMRI: Haemodynamic Response

The changes of the signal due to the neuronal activity is called the *Hemodynamic Response* and could be summarized with the acronym *HDR*. One important thing to know and remember is that the shape of the *HDR* changes with the stimulus properties.

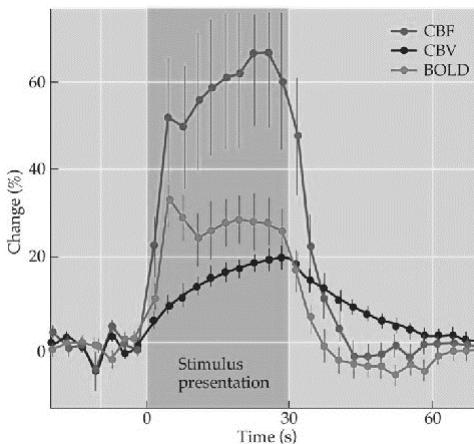


Single stimulus



Multiple stimulus

In the first case we have a single stimulus, and we can see that after a certain interval of time we can see the activation a hemodynamic response (HDR) and depending on if we keep administering the stimulus or not this response change its duration. In the second situation in fact, we have a plateau phase which is due to the spatial summation of different HDR. As we can clearly see from the two responses, we have a post stimulus undershoot and to understand why this happens we need to investigate the pattern of CBV and the CBF. These show the behaviour reported below:



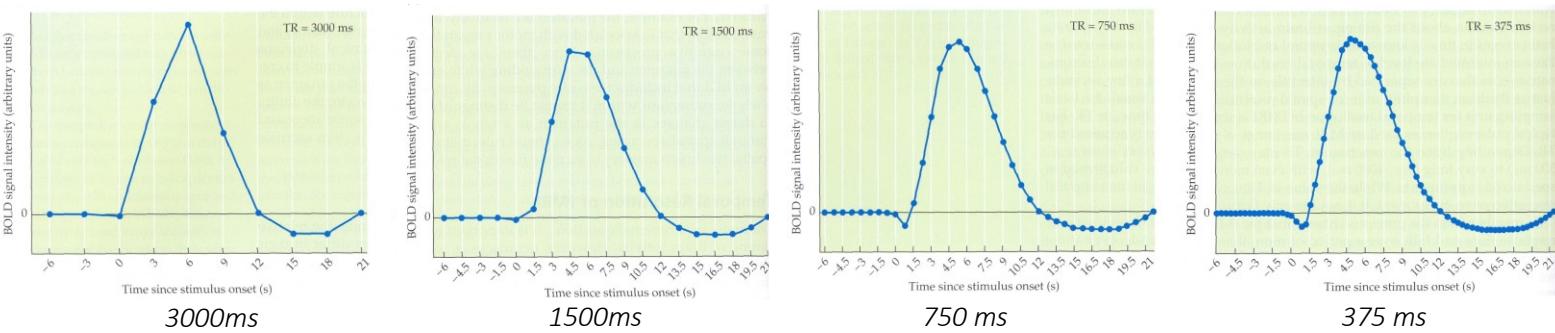
From this image we can see that the CBF is much quicker than the CBV and this explain why we can see the negative response or the negative peak ones the stimulus is terminated. In fact, the faster trend of CBF will lead to a higher concentration of deoxyhaemoglobin than the basal condition.

fMRI: Settings and technologies requirements

Now we need to take care about all the possible characteristic of the BOLD signal. Usually, the spatial resolution of an MR scanner is of about $3 \times 3 \times 3$ mm and it strongly depends on instrumentation and field of view (*FOW*). Bigger is the voxel, higher is the signal to noise ratio we can obtain. As side effect to increase

a lot the dimension of one voxel we could activate a lot of negative feedback mechanism as the partial volume effect. This is due to the presence inside the same voxel of different tissue.

We have to remember that with MR we are trying to characterize the grey matter of the brain and not the white one. This because, in the white one we do not have any soma of neurons but just different types of glia cells and axons of neurons. The brain regions where the metabolic activity of the neurons is concentrated is in the grey matter. This because we can find the major quantity of soma of the neurons and a lot of glia cells too. This matter is perfused from a lot of vessel that provide oxygen and other nutrients to all the cells. Moreover, we have to consider that the thickness of the grey matter is of about 2 or 3 mm. The temporal resolution, in the fMRI, is the repetition time (TR). At minimum, this value could be equal to 0.5 second with the best scanners in the market and at maximum 4 second with the most ancient one. With a conventional 3 T scanner we could find at maximum a TR equal to 2 s. Obviously, this has a lot of implication. To understand the sample frequency to get a good behaviour and reconstruct faithfully the pattern of BOLD signal we need to have a repetition time (which determine the sample frequency of the BOLD signal) lower as possible. In the most usual clinical application, the HDR is estimated with a repetition time of about 1.5 s or 3 s. Obviously the activity reconstruct is not the best one but it is enough in clinical practice to get a good visualization of the possible pathology condition of the patient.



We can see that by decreasing the repetition time (TR) we increase the precision of the pattern description. Of course, MR is a field where there are a lot of trade off. In fact, TR cannot be drastically decrease, if $TR > 2$ the flip angle is of about 90°. But if the TR is lower the flip angle must decrease too and, in this case, the measured signal drops. On the other hand, low TR increase the size of the volume that can be acquired. The technological requirements of the instrumentation that needed to be used are the next ones:

- High magnetic field, almost greater than 1.5 T.
- Fast acquisition such as the *Fast Gradient Echo* and the *Echoplanar Imaging*

While in the conventional we have that the acquisition time is almost greater than the product between the TR and the size of the matrix in the phase direction with the echo planar sequences the total amount of time needed to acquire an entire volume is TR. The different methods that are used to investigate the BOLD signal activation could detect different types of activity:

- **BOLD Methods:** GE – EPI, SE – EPI, RARE, (b)SSFP. They are more sensitive to the intravascular and extravascular effects.
- **Blood Volume Methods:** VASO, Contrast agents. They are more sensitive to the intravascular effects.
- **Blood Flow Methods:** ASL. They are more sensitive to the intravascular effects.

fMRI: BOLD Physiology

Regarding the physiology the neuronal activity is strictly related to the electrical activity. This is always due to ions concentration distributed over the membrane. During the resting state condition, the concentration of ions inside and outside the cell is not balance and this leads to the creation of the resting state membrane potential which is slightly negative (-70 mV). The electrical activity is propagated through the axons of the neuron and when it reaches the axonal button the neurotransmitter, which is contained in vesicles, are released in the synaptic cleft towards to the target cell. The usual neurotransmitter is glutamate, which in the CNS (*Central Nervous System*) is prominent and the GABA. The first one is an excitatory neurotransmitter while the second the most common inhibitory one.

The neurons are very far from the electrochemical equilibrium in fact there is always active an ions gradient that tends to move all the species and moreover, the establish again the equilibrium it is needed a lot of energy. In general, the brain at rest consumes the 25% of the glucose consumption and the 20% of oxygen consumption pumped from the heart every second. The brain does not have a store for these compounds and thus it needs to receive every moment a constant supply of these substrates. In fact, it receives the 15% of the cardiac output. The fresh blood is always guarantee from the complex of arteries and capillaries that sprinkles the grey matter. The system of venous clears the deoxygenated blood, waste products and heat. The major quantity of oxygen which is bring to the brain is contained in the red blood cells which contains the haemoglobin that carries the 98% of the oxygen.

When we look of what going on during the cerebral activity, we could see that the neurons, glial cells, and the microvasculature form a neurovascular unit. This is such important because means that these three components could exchange information between each other. Thus, if we take a close look of what is going on the arteries and penetrating arterioles are surrounded by smooth muscle that activate response to activation of cerebral region. In contrast to that capillaries lack smooth muscle and veins have sparse muscle cell coverage and this led to the passive response of the vein which is driven from the increase of upstream pressure.

fMRI: Model for BOLD signal

Before introducing how we model the oxygen supply and demand we need to introduce some definition:

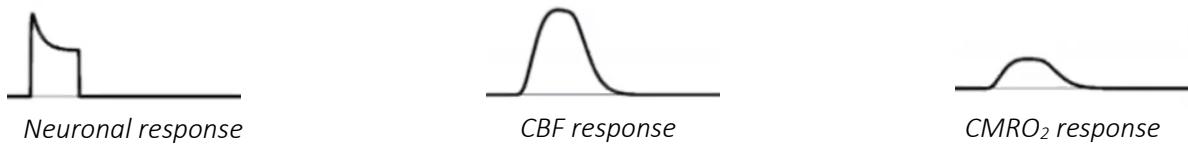
- *Cerebral Blood Volume (CBV)*: It represent the fraction of the blood volume inside the tissue volume or tissue voxel volume. In grey matter the fraction of blood volume is of about 4 to 5 %.
- *Cerebral Blood Flow (CBF)*: It is also known with the name of perfusion and represent the blood flow into capillaries. In the grey matter the blood flow is of about 50 ml / 100 ml of tissue / minute.
- *Cerebral Metabolic Rate of Oxygen (CMRO₂)*: It represents the rate of oxygen consumed by the tissue and in grey matter it assumes the value of 1.6 $\mu\text{mol O}_2$ / 100 ml tissue / minute.
- *Oxygen Extraction Fraction (OEF)*: It represents the relative O₂ extracted from the blood. Usually across the brain this value is of about 40 – 45 %. This parameter is very important to understand the BOLD effect as it is related proportionately to the concentration of haemoglobin.

Finally, all the possible terms reported above could lead us to define the Fick's Principle that connect all the possible values that we have described and introduce before. The law represents the mass conservation written for the previous situation:

$$CMRO_2 = [O_2]_a \times CBF \times OEF$$

This is one of the most important equation to understand the model of the BOLD signal. Because it may help us to understand how much oxygen remains in the vein after the extraction by the brain region activated. In fact, increasing only of the cerebral metabolic rate of oxygen it means that we have an increase of the fraction of extracted oxygen, while if we have an increase only in the perfusion, we have a decrease of extracted oxygen and this means that we have a higher amount of oxygen inside the blood.

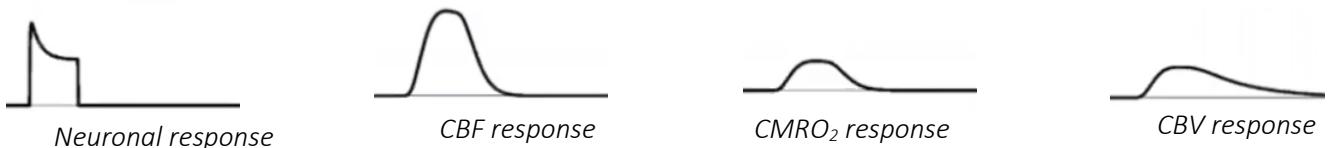
It has been shown that the increase of neuronal activity leads always to increase the glucose consumption. Moreover, researchers have found that the glucose consumption and the cerebral blood flow increase linearly. This could be studied with PET by using the appropriate tracer. The same studies that shown the linear relation between glucose consumption and the blood flow shows that there still a linear dynamic between the blood flow and the oxygen consumption, but the ratio is much lower than the one seen and detected for glucose.



The CBV and CBF are coupled together and in the steady state the behaviour could be described by the power law shown down below, where the exponent is called the *Grubb exponent*:

$$\frac{CBV}{CBV_0} = \left(\frac{CBF}{CBF_0} \right)^\alpha$$

The blood volume is not linear related to the blood flow, but changes as the power of the blood flow changes. For a laminar flow the exponent of Grubb is equal to $\alpha = 0.5$, but in experiments where this component was measured shown that the most correct value is about $\alpha = 0.38$ and this is due to the non – perfectly cylindric structure of the vessels. Small changes in blood volume could enable large changes in blood flow.



What links all to MR is the haemoglobin. Because we have this large overshoot in blood flow, but we have a low oxygen consumption we end up in an oversupply of the oxygen to the tissue which results in a decrease of deoxygenated haemoglobin and thanks to this we enhance the magnetic field and this lead to the HRF pattern.



Hb response



BOLD response

The blood volume oxygenation level – dependent signal could be expressed proportionally to:

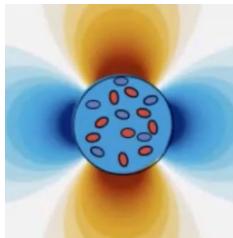
$$BOLD \propto \frac{CBF}{CBV CMRO_2}$$

From this relation we can see that if the blood flow and the blood consumption increase of the same amount this will have no effect of the BOLD effect.

To represent the BOLD effects, we use the hemodynamic response function known also as HRF. First it presents first an increase or more specifically a peak and then it returns to the steady state condition. It represents and capture the neuronal activity when it is activated. In general, it presents some variability between subjects, brain regions, brain state and pathology.

fMRI: Magnetic field

In terms of magnetic changes, as the venous blood is much more deoxygenated than the arterial one we have a pattern that is much higher a better defined in the veins, as could be seen from the images below.



Venous sO₂ ≈ 60%



Arterial sO₂ ≈ 98%

The venous will have a lower T_2^* decay. What we are really interested in is the:

$$R_2^* = \frac{1}{T_2^*} \propto [Hb]$$

During the activation of a cerebral area the amount of saturation of oxygen in the venous increase by the 10% from the previous state and therefore, the BOLD signal is venous weighted. Obviously, the above reported situation does not represent the reality or a voxel because in those cases we have a higher amount of venous inside the same voxel. Another important compartment to analyse is the variation of the magnetic field created at the molecular level. Thus, we have two variation effect, the extra vascular one and the intra vascular one. The percentage of contribution to the BOLD signal from the intravascular and extravascular compartment depends on the intensity of the static magnetic field B_0 . The intravascular contribution decreases by increasing the intensity of the field. The different MR sequence could capture different

characteristic. In fact, in the gradient echo sequence we can be more sensitive to pial veins, while by the usage of a spin echo technique we can find more information about the intracortical veins. The spin echo in practice is not used too much because the gradient echo gives us much more reliable results.

fMRI: Activation study – Experimental design

The experimental designs in this field of fMRI are such important and could be subdivided into mainly three categories:

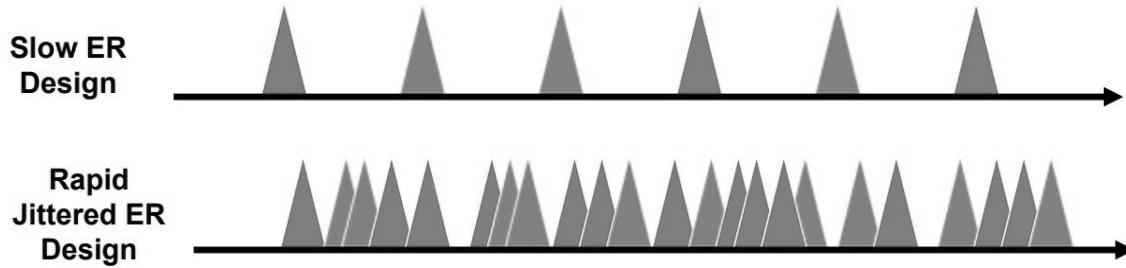
- *Blocked*
- *Event – Related*
- *Mixed*

The first that we consider is the blocked experimental design. It could be represented by the next scheme. In the image reported below we can see that we the patients need to perform a resting state condition and then we have an activity period. This pattern is repeated periodically and so the resting and the activity phases are separated from respectively activity and resting phases. With these experimental settings we first analyse the cerebral activity during a do – nothing condition and then we compare this state to the activity phases where instead we should record some HDR signal in determine brain region task – related.



This type of analysis and design of the experiment could involve different type of activity block and may also interval just two classes of action and not just one and resting state. This complicates a lot the design and the results are much more difficult to compute and analyse.

The second experimental design that we talk of if the event related. This category could be divided in two again, the periodic single trial and the jittered single trial. In the first category we have different stimuli that are shown with a low frequency. The inter stimulus interval is quite high in these cases. On the other hand, the second category the stimuli have different inter stimulus interval and are presented such randomly. The two representation of the experimental design are the one reported below.



The third category is the one that mix up the two previous methodologies. We intersperse resting state phase and task phases. During the task phases we may implement one of the two paradigms saw before for the event related experimental design. This experimental setting is the one show below.

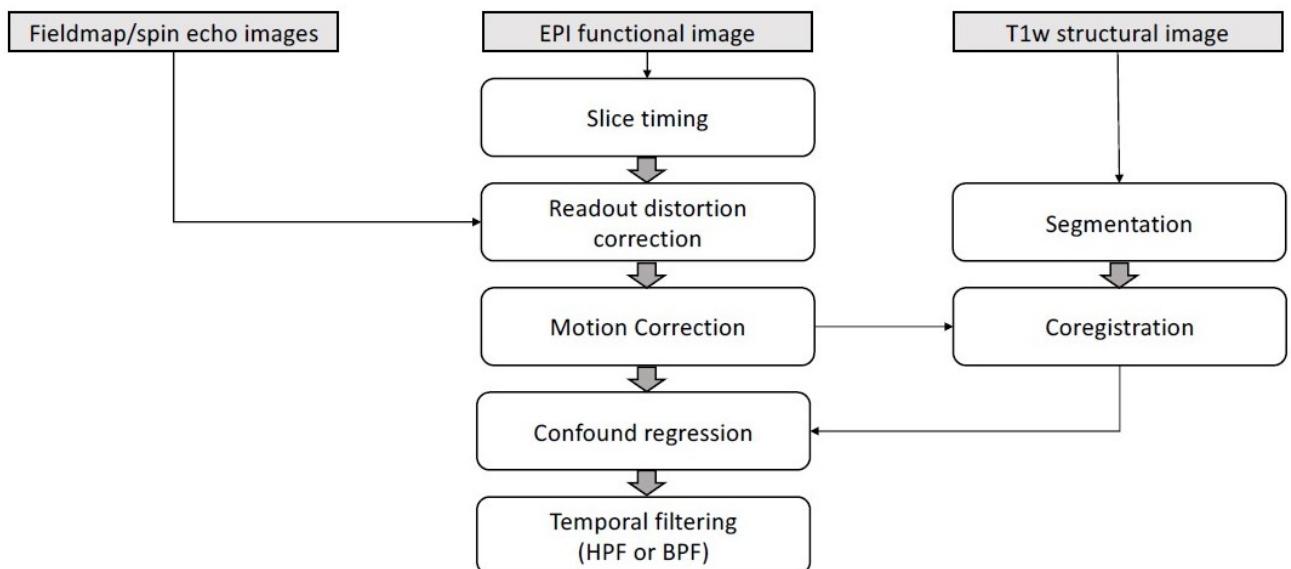


There are some pros and some cons in all the experimental design settings. The blocked design is less sensitive to head movement but not all the possible investigable tasks could be performed with a blocked experimental setting. There is the difficulty to pinpoint the HDR but as a pro this experimental setting presents a high statistical power. On the other hand, event-related has a higher dependency of the ISI value but could accurately describe the HDR (latency, shape and others characteristics).

There is in general a good consideration to hold in mind that answer to the question why it is so important to use an experimental design. The answer is easy too. During a specific activity or resting state period the subject involved in the analysis is requested to perform a specific task as instance a do-nothing condition or the thinking of a motor action. In all these cases when it is first proposed this stimulus to the subject it activates certain brain regions which are the only ones demanded to answer correctly to that stimulus. But if the stimulus and the specific activity is requested to a long time, the subject learns how to perform the task and there is an adaptation of the answer. This leads to the activation of different brain circuitry and regions. This is the reason why to the subject the maximum period of a certain activity is 30–40 seconds.

fMRI: Pre processing

Another important part of the analysis of an fMRI data is the pre-processing step. This is very important because let the data be prepared to be then elaborated and processed. The usual steps that are done are reported in the scheme down below.

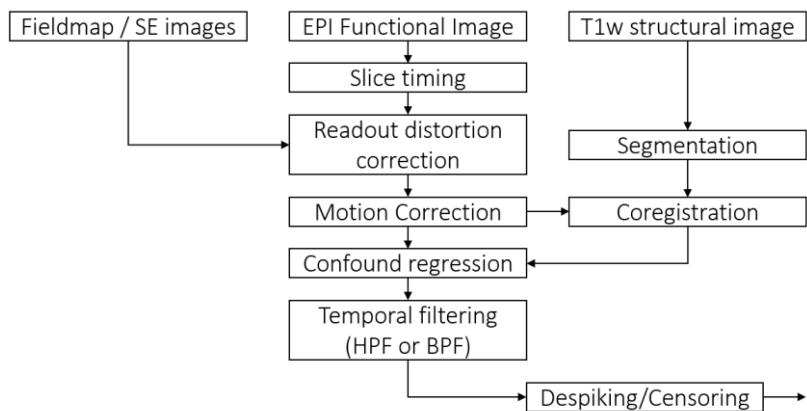


There is only one pre-processing step that is not considered in the previous scheme. This one is despiking and censoring. In the literature there are available different software to perform these analyses as instance SPM or FSL.

LESSON 10: FUNCTIONAL MAGNETIC RESONANCE PRE PROCESSING AND STATISTICAL

fMRI: Pre – processing step

The pre – processing steps are a lot and usually are applied to detect and decrease the impact of specific artifacts such as inhomogeneities in the magnetic field. The usual procedure used to acquire an MRI image is the EPI sequence (*EchoPlanar Imaging*). The pre – processing steps are done because we have to use the data to extract some physiological and informative parameter to improve an eventual pathology diagnosis made from the medical staff. Moreover, sometimes, the fMRI data are widely used to build up functional connectivity maps and to enrich and confirm the results obtained from these analyses we also need some structural maps. These are obtained from MR images. To correctly perform the study, we have to align all the data on a same reference. To do so, the correct process to apply is some a co – registration techniques. In general, we can summarize all the possible pre processing steps with the next scheme:



In the following of the lesson, we will look to all the possible previous block and their meaning.

fMRI: Slice timing

The slice timing is a very important concept and describe how the 3D volume is acquired. Before, introducing the possible methodologies we have to remark that the fMRI data are acquired slice per slice. Thus, a 3D volume is made up of different slices acquired in different instant of time. Vice versa, the PET is acquired as already a 3D object. The slice acquisition during an fMRI analysis could belong to three different categories.

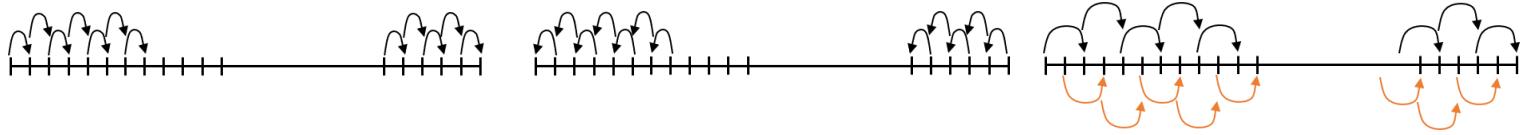
- *Ascending*

- *Descending*

- *Interleave*

In the first one, the slices are acquired one after another with a specific direction, during the descending phase instead, the acquisition follows a backward principle, respect to the previous acquisition progression. The third one is the most important methodology. In this case, in fact, the acquisition follows the next paradigm. Given the slice's index or $i_{slice} = 1, \dots, N$, where N is the total number of slices, during the first steps, the scanner acquired only the slices with odd index and then are acquired the slices with even index. This sounds strange without considering all the possible inhomogeneities of the magnetic fields applied. Therefore, the reason why this method is largely used is because our aim is to reduce as much as possible the influence of the excitation with a radiofrequency stimulus of the slices close to the one analysed before

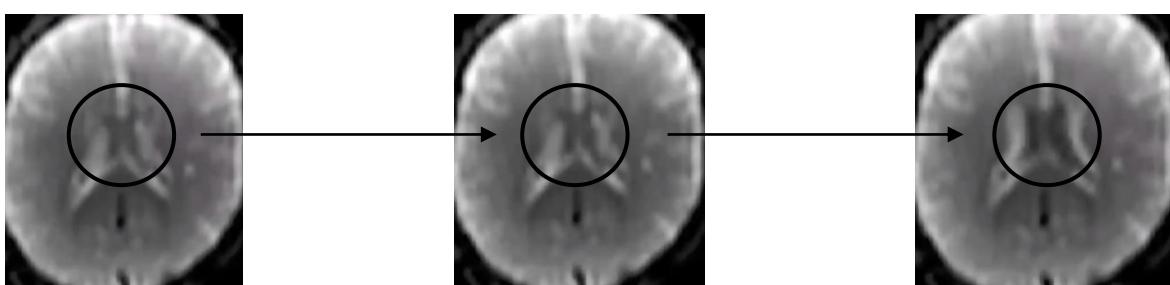
they are inspected. The name of this artifact is *cross talk*, and as we can understand this is very misleading during an ascending or a descending acquisition methodology. Down below are reported schematically the previous cited methodologies.



The order of slice acquisition influences the results obtained. Due to the non – same time analysis we have to perform a slice timing correction that consist in re – reference all the acquisition as all the slices would have been acquired in the same instant of time. To reduce as much as possible the artifacts due to the non – same time acquisition this process has to be as fast as possible. In general, when we use these three main methodologies to acquire an entire volume, we have to remember that the overall acquisition is associated to an instant of time. But to makes the measures in the different voxels be statistical meaningful, after the acquisition, we have to realign temporally all the possible signals. This process is called slice timing. To do so a temporal interpolation is needed to make the HDR appears like it is though. Only after this process all the possible hemodynamic responses are aligned and could be used. Obviously, the interpolation works well just in those cases where the sampling rate (time between slices) is higher than the changes in the data. Generally, is more effective for short *TR* (*Repetition Time*), 1 – 2 ms, than longer *TR* greater than 3 second. Some temporal realignment algorithms have been developed from Siemens and other important industries.

fMRI: Motion correction

The second step of the pre – processing phase consists in the motion correction also known as realignment phase. This process is very important because in the most of cases some movement artifacts are always present in MR images. Usually these are due to the high duration of the exams. To understand how much sensible the MR technology is, we have to look on the resolution of the scanner. Usually, just a head movement of 4 – 5 mm causes errors in the measured signal and could bring to misleading results. By taking in consideration this, researchers and clinician built up many fMRI experimental design to reduce the subject's fatigue of holding the same position for the entire procedure. The usual paradigms are divided in several sub experiments lasting 4 to 6 minutes tops. During the pause it is requested and force, if it is possible, to move and stretch the head a little. This experimental design could lead to a better result in the data obtained and even on the side of the patient, because the medical staff could explain to he or she what is happening. Sometimes, even applying the most energy less consuming experimental design, caused by the task request, a head movement is inevitable. This is generally reached when it is asked to the patient to perform some motor imagery tasks. Other inevitable movement are caused also by the breath and blood pulsation (figures evolution reported below).

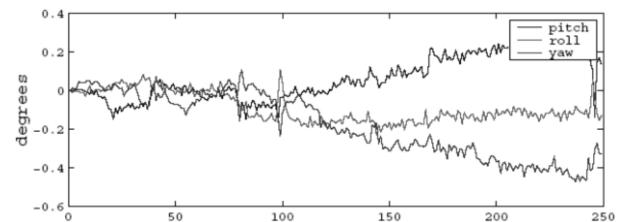
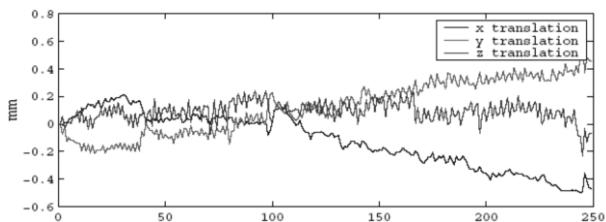


To solve the motion artifact sometimes are used some cage or mask provided with pillow. In the images below reported we can see the possible mechanical solution used to prevent and reduce exponentially the motion artifacts.



Some of these masks are also PET compatible. In this last case the patients could perform a PET and after that an fMRI image acquisition consequently. Usually, these types of thermoplastic mask are not used because the claustrophobic sensation is extremely high. Other methodologies are by the usage of prospective motion correction tools such as optical tracking system. Usually, this last method during the acquisition works quite more efficiently than the previous cages.

After all the possible precautions, we can correct posteriori the head movement by the usage of co – registration algorithms. These methods exploit all the possible movement to correctly realign the volume to the patient chosen reference. In fact, before applying any type of procedure we have to choose which is the reference system for that given patient. The process of move each image and volume to a certain reference is called *registration* or *co – registration*. Generally, all the possible images are registered respect to the reference volume. The admissible movement are mainly six, three rotation and three translation because it is assumed that the brain is modelled with a rigid body. The Euclidean norm calculated with all the six parameters will give us the absolute displacement (it is a weighted Euclidean norm because we are working with element characterised by different unit of measurement). Another measure usable is to calculate the Euclidean norm of the first derivative applied over the six parameters. In the usual case, if the analysis is performed in a time interval where the physical dimension of the head remains fixed, a combination of translations and rotations is enough. In the most of algorithm that are used to perform this process are produced also some files that represent the transformation applied to a given slice or volume to be registered as the reference one. Down below is reported an example of combination of rotation and translation applied over a certain volume.



As could be seen in the images above we have different abscissas values that correspond to indexes of certain slices. From these images we can see that from the 100-th slice both the translation and rotation start assuming value different from an oscillation close to zero. At the end, we can see that the maximum translation applied is of about 0.4 mm (in absolute value) which is a very good result. For this reason, we can consider the measurements performed on the given subject as good ones.

A different way to prevent all the possible geometric distortion, we can compute and acquire the signal and the different slices with increasing factor of parallel imaging by take into account algorithm like GRAPPA and SENSE.

fMRI: Smoothing

Before or generally after having applied a motion correction algorithm it is needed to perform some smoothing processes. These phases are known as *readout distortion correction*. By applying some smoothing process, we increase the signal to noise ratio (*SNR*) of the image acquired. These are largely used in MRI. In this field of application, the possible distortion correction is not just made up of filtering in the space plane but also some other more advance algorithm that recreate the possible magnetic field inhomogeneities and determine the correction factors. Anyway, without entering deeply in the field of readout distortion correction algorithm, the filter mostly used is the Gaussian. When we apply the Gaussian filter, as we know from the bioimaging course, we are reducing the noise effect of the image and increasing thus the SNR. When the Gaussian filter is applied also some blur appears and therefore the *Full Width Half Maximum* (*FWHM*) is modified. This length measures the minimum distance between two distinct objects such that they are still recognizable as two separate ones. In the image reported below we can see the response of the Gaussian filter in the space plane and in frequency.



As could be seen in the images reported above, the Gaussian filter present a specific shape. The mask is generally an approximation of the 3D shape reported before. The most important thing is that the dispersion, or the extension of the region not – filtered is related with the standard deviation σ of the gaussian bidimensional distribution. This present also high correlation with the new *FWHM* of the filtered image. The analytical relation could be written as:

$$FWHM \approx 2.36 \sigma$$

In general, typical values of FWHM are 3 – 10 FWHM. The filtering process could be considered in this case as a weighted average. As cons of the usage of this simple algorithm is the reduced spatial resolution and the increased partial volume effect.

fMRI: Temporal filtering

Another important step to do before the fMRI data analysis is the temporal filtering. This is done by the usage of high pass and low pass filter and should be done because there a lot of artifact that are injected from different sources. Some of this are as instance due to the hardware of the instrumentation, the blood pulsation, the breath and other electronical or physiological sources. In general, the frequency of these components is not too elevated and for this reason it is majorly used a high pass filter instead of the other

typology. This process should be done also to remove eventual baseline drift which could vanish eventual statistical analysis. Usually, the cut off period is equal to 1.5 times the stimulation period. In SPM, the cut off instant of time is equal to 128 sec. This step, as all the other in the chain reported at the beginning of the lesson, could have some potential issues such as remove frequencies related to the hemodynamic response. Thus, we have to be aware of which frequencies are related with physiological artifact and which not.

fMRI: Computation of the SNR

There is a standard way to compute the SNR of a given *Region Of Interest*. We select first this last region outside the cerebral volume without artifact and we compute the mean value μ and the standard deviation σ . Then we select a ROI inside the cerebral volume, and we compute again the mean and the standard deviation respectively indicated with μ_{ROI} and σ_{ROI} . The analytical relations that could be used in our situations are:

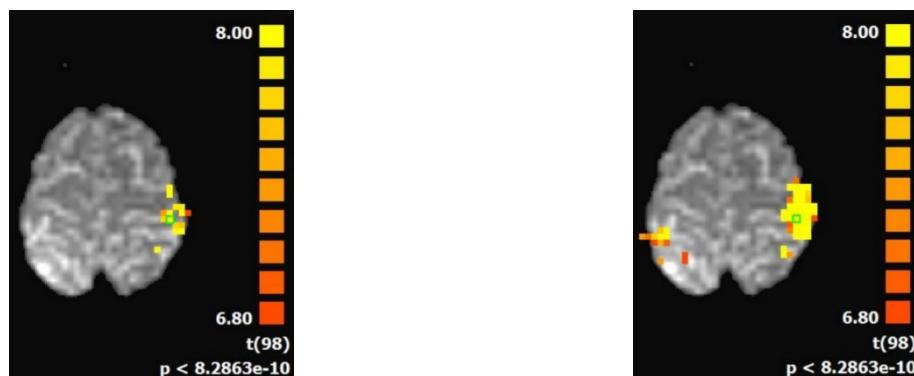
$$SNR = \frac{\mu_{ROI}}{\mu}$$

$$SNR = \frac{\mu_{ROI}}{\sigma}$$

We can quantify the quality of our data by calculating this metric. When we cite and read about the SNR value, we should always state which denominator we have used. Typical values of SNR for the head coil are greater than the ratio 50:1 while if we consider the surface coil, we should reach ratio of 100:1.

fMRI: Key role of pre processing

The role of all the pre processing phases previously reported are very important to know. This because different pre processing steps could lead to different topological brain maps. This is such important because, when we analyse results coming from different studies the notable differences between the studies performed could lead to different results. An example is the results reported below. As could be seen in the two images the left hemisphere (right section of the image, radiological view) is much more activated in the image reported on the right, even if the task requested to be performed by the subject is the same. The main differences are just due to the pre – processing steps applied.



fMRI: Statistical analysis

After having speak and discussed enough about all the pre processing phases to reach the end of the analysis we present how we can compare results and images obtained during different task of the subject.



In the previous chart we can see the presence of some pre processing phase like the one that we have introduced before and at the end of the chain we have the statiscal analysis box. The objective is to find the possible statistical tool to apply on our situation to enhance the activation and the differences between certain brain region during a task phase and the resting state. One possible solution to out problem, to enhance and highlight the activated region, is by simply taking the differences between each voxel during the resting state and the task involved state. This is a very intuitive solution and would be very easy also to implement, but in real cases we obtain a final image which has a very low signal to noise ratio. For this reason, there are a lot of possible technique to apply in a certain situation to achieve better results and improve the differences between the resting state and the active state. These methods are:

- *t – test*
- *Correlation analysis*
- *General Linear Model*
- *Kolmogorov – Smirnov*
- *Fourier analysis*
- *Independent Component Analysis*

fMRI: Statistical analysis – t-test

In the statistical analysis based on the simple concept of t – test we have to define the two different hypotheses. The null one (H_0) and the alternative one (H_1). These are defined as:

- H_0 : The two groups are equal.
- H_1 : The two groups are different.

As usual, the results of the null hypothesis test is a decision. To understand the performance of the test we have to compute the confusion matrix for the t – test. This could be represented as the one reported below.

		DECISION	
		H_0 true Group 1 = Group 2	H_0 rejected Group 1 \neq Group 2
REALITY	H_0 true Group 1 = Group 2	TRUE NEGATIVE	FALSE POSITIVE
	H_0 false Group 1 \neq Group 2	FALSE NEGATIVE	TRUE POSITIVE

Depending on the exams that needed to be done we have to control the false positive or the false negative. This could lead to different situation, in fact, if we are dealing with a cancer removal surgery before starting the intervene, we have to acquire an fMR image and we have to establish which voxel are associated to cancer tissue. In this case, it is better controlling the false negative instead of the false positive. The reason why could be explained in the next few lines. If we have a subject which is affected from brain tumour and he or she may undergo on surgery this is already a bad situation. For this reason, we want to open the scalp and remove completely the tumour from the brain even if the tissue is not damaged. To obtain this results we have to control the false negative, on the other hand we have also to keep under a certain threshold the

value of the false positive. Even if the t – test could be thought as a good solution for our problem, the results of this methods is not always reliable. This because we have a lot of non – deterministic factors that we may consider during our analysis. As instance, the too much noise inside the signal could corrupt it and altered the effective results. Moreover, even the acquisition time instants are not the same for all the slices and we may compare in our analysis also two different sites of the brain, and due to the non – temporal co – registration we may face with wrong results. Thus, before computing all the t – test comparison we need to realign temporally all the possible measure by different algorithm and then we can compute the t – test comparison. This is the only way that we can use to compute more correctly this statistical analysis. These are some of the possible issues by computing the t – test.

In general, the reason why we should use the t – test is because it assumed that the samples have a gaussian distribution. This assumption is meet by fMRI data, as actually during the fMRI acquisition hundreds of time points are measured. On the other side, the pitfall of this technique are the unsolved problems such as head motion correction heavily impact the results, the t – test is not appropriate to answer question about the neuronal activity timing and finally, the t – test evaluate differences on the basis of distribution mean but it is sensitive to changes in distribution variability or in HDR shape.

fMRI: Statistical analysis – General Linear Model

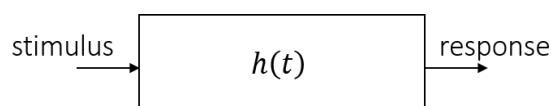
The best way to create the parametric map is by the usage of *General Linear Model (GLM)*. This theory considers the data as a linear combination of several functions summed up with some noise. The basic functions have known shape, but their amplitude need to be estimated. The estimator needed to be used in this case is linear methodology. Our model, in this case, could be described with the analytical expression:

$$Y = X\beta + \epsilon$$

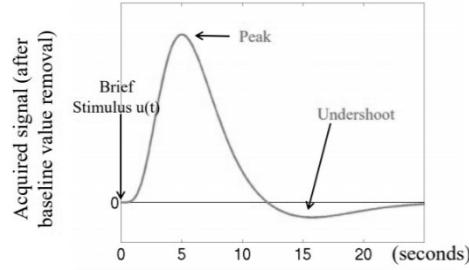
Where each term has the next meaning:

- $Y: Y \in R^{N \times 1}$ is a column vector, as could be seen from the notation, and represent the single voxel acquired data. It represents the output of the model.
- $X: X \in R^{N \times P}$ is a matrix. This represents the haemodynamic response for each voxel.
- $\beta: \beta \in R^{P \times 1}$ is the parameter column vector. This represent the amplitudes of the basic function that needed to be estimated.
- $\epsilon: \epsilon \in R^{N \times 1}$ represent the noise overlapped to each voxel measures.

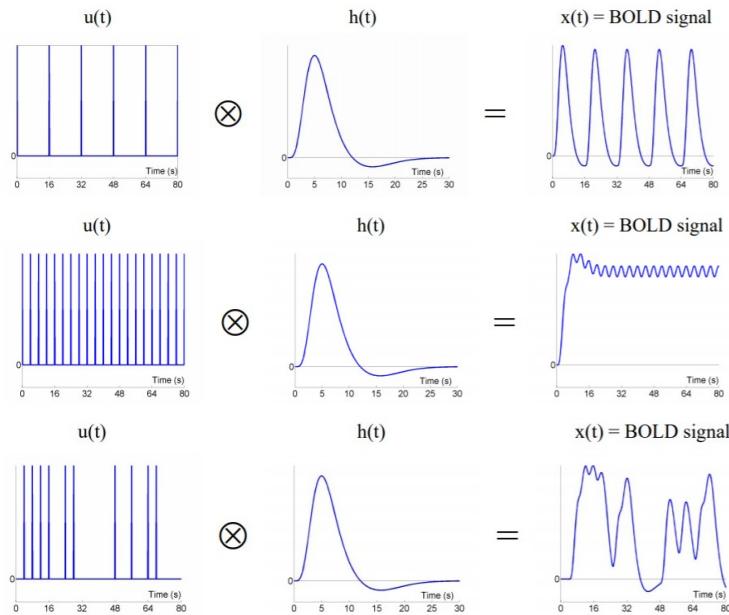
The aim of this analysis is to model the haemodynamic response when a certain stimulus is proposed. In this case to describe correctly and numerically the haemodynamic response model we have to define the impulse response function. This could be represented mathematically with the symbol $h(t)$. From the system point of view, we could write:



The system behaviour when the stimulus is the impulsive stimulus $\delta(t)$ is the next one which perfectly represent the shape of the HDR.



From the image reported above could be clearly see the presence of the usual pattern of HDR introduced in the last lesson. To understand if this model works well or not, we have to describe its behaviour during a multiple impulsive stimulus sequence.



As could be seen from the images reported below, because of the model is a linear one, once find the impulsive response we could determine easily the behaviour during any general stimulus situation. The analytical expression used in this case is the next one:

$$x(t) = u(t) \otimes h(t)$$

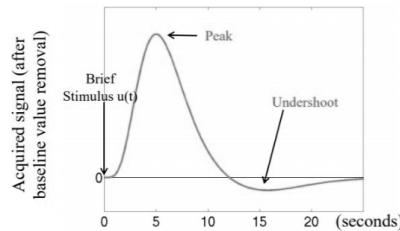
One important thing to determine analytically and mathematically is the impulsive response. In the literature nowadays are available different model to correctly find the behaviour of the HDR. The model that we will use to model the HDR is represented from the next analytical expression proposed by Friston et al. in 1998:

$$h(t) = \left(\frac{t}{d_1} \right)^{a_1} e^{-\frac{(t-d_1)}{b_1}} - c \left(\frac{t}{d_2} \right)^{a_2} e^{-\frac{(t-d_2)}{b_2}} \quad d_i = a_i b_i$$

LESSON 11: fMRI STATISTICAL ANALYSIS AND MULTIPLE COMPARISON CORRECTIONS

fMRI: Statistical analysis – General Linear Model

In this lesson we will finish to speak about the General Linear Model and then we will talk about the different statistical test applicable and the different correction methodologies to get results confirm from the statistical point of view. Just to remember one important thing we report down below the characteristic pattern of the HDR.



As we have already introduced in the last lesson, we want to use a model to generate the different parametric maps. In these last one will be plotted the parameter estimated in the linear model reported below.

$$Y = X \beta + \epsilon$$

Where the meaning of each term are the next ones:

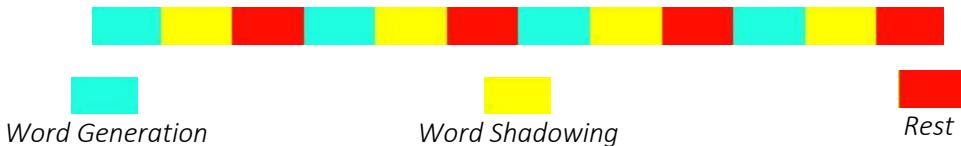
- Y represents the acquired data for each single voxel. It presents the dimension of $N \times 1$, where N indicates the number of samples chosen.
- X Represent the model for the HDR chosen. The dimension of this matrix is $N \times P$, where N is the same number introduced before, while P represents the number of trials decided to be modelled (this concept will be explained better further in the lesson).
- β which represent the coefficient to give to each trial modelled to obtain the best estimation of the signal acquired for that voxel.
- ϵ is defined as an error which is additive and presents a gaussian distribution with samples mutually uncorrelated with standard deviation constant σ . The dimension of the additive error vector is $N \times 1$. In general, we could write: $\epsilon \sim N(0, \sigma^2 I)$.

To introduce and to understand well the analysis that we will see during this lesson, we introduce an example that will be used during all the possible explanation. Let take in consideration a situation where we want to perform an fMRI data acquisition of an experiment made up of 3 different task each of these last 21 seconds. The three task that are used are the next ones:

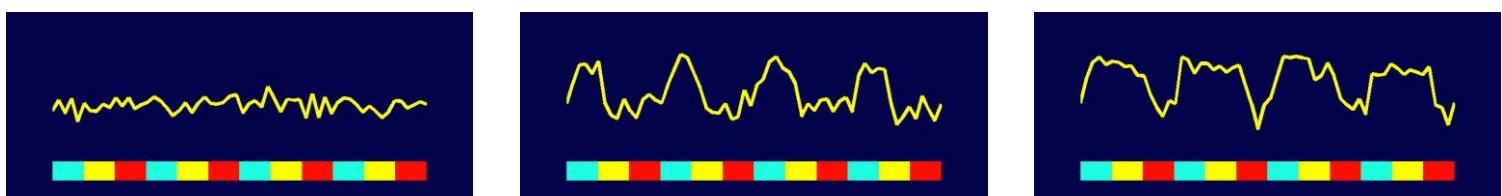
- *First task:* It is called “Word Generation” the objective is to present in a screen a certain word to the subject involved in the analysis and let the subject think voluntary to a certain verb and an action connected to the word presented.
- *Second task:* Its name is “Word Shadowing”. The aim in this case is to present to the subject a certain verb and let the subject think to the verb presented.

- *Third task:* It is named “Rest” and the subject during this phase should do nothing. But with “do nothing” does not mean sleep but try to don’t do and think to anything.

The three tasks are presented with a blocked experimental design as reported below.



From each voxel of the analysis, we can extract a specific time series. By looking on a signal related to a specific voxel (contained inside a specific region of the brain) we could see the presence of specific oscillations in correspondently to specific task. Thus, from an initial visual inspection we could understand which voxel are majorly related to specific task and which ones not.

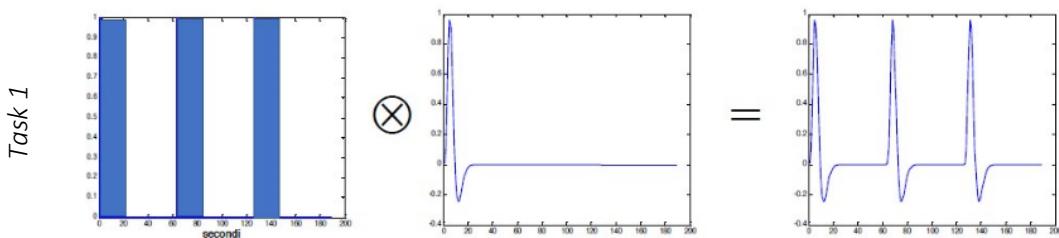


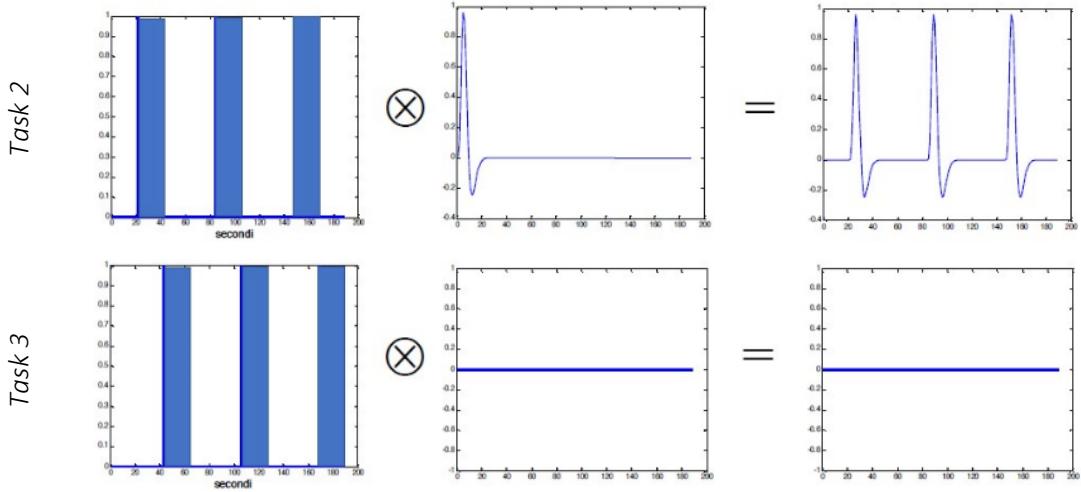
In correspondence of this voxel, we can see that during any activity it is activated an HDR. This means that the anatomical region of the brain where this voxel is located is not demanded to perform the activity requested during the tasks.

In this case, instead of the first one, we can see that an HDR is activated during the “Word Generation” task. From the signal we may also see a little delay in the activation of the haemodynamic response.

Finally, the situation reported above presents the activation of an HDR in correspondence to the first and the second task named “Word Generation” and “Word Shadowing”.

The usage and the introduction of the GLM is to try to fit the data acquired over each voxel. To understand how the fitting process is done we need to understand how the output of the model is created. At the beginning of the lesson, we have introduced again the pattern of the HDR, and we need to use it as the impulse response function of the system. To calculate the output during the activation of a certain task that last 21 second we need to evaluate the convolution between the impulsive response function of the model and a combination of Dirac impulse that define the beginning of the specific task. These convolutions are reported below.



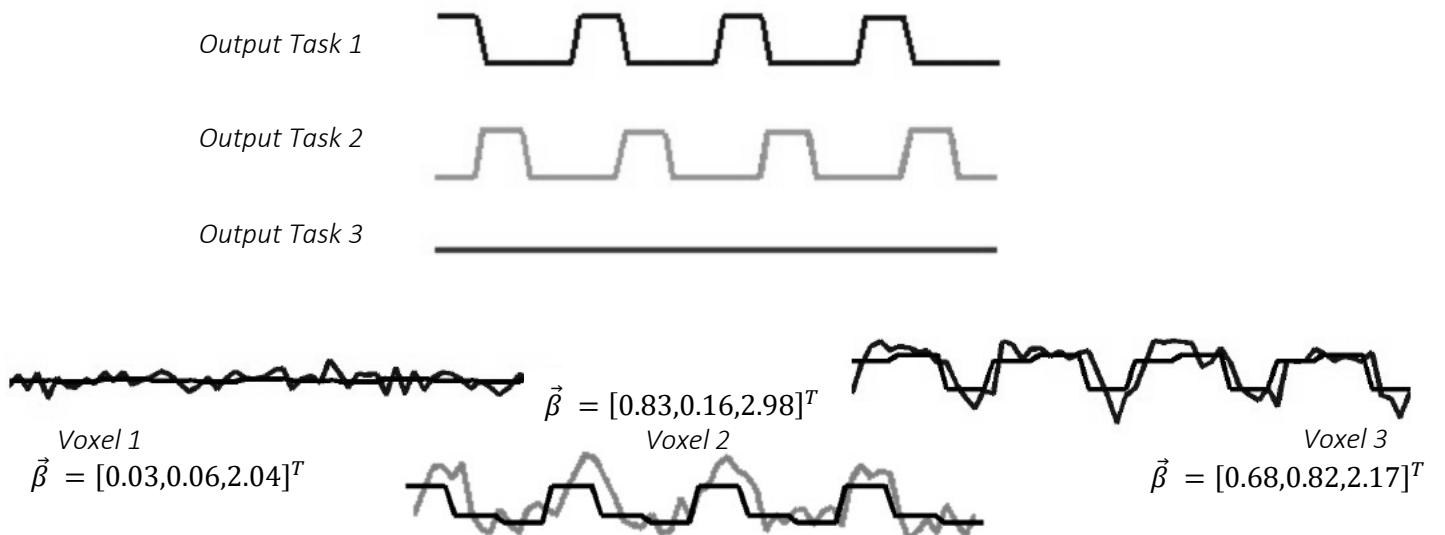


Each output of the model in correspondence to each task are then sum all together and before summing them we scale them for the parameter values contained inside each element of the β . Thus, as instance the output coming from the *task 1* is scaled by the factor β_1 , while the output coming from *task 2* is scaled for the value β_2 and finally *task 3* by β_3 . These coefficients are estimated by fitting the sum of these three signals to the acquired signal for each voxel. Consequently, as result, we obtain a parameter vector of three element for each voxel acquired during the fMRI experiment.

Now it is clearer also the meaning of P , which is the dimension of the parameter vector β . This represents the number of different trials that needed to be consider in our analysis to correctly fit the acquired signal.

Even if this analysis does not show any problem when it is applied, in the real application it is not used as impulsive response the HDR that we have seen but it is used a type of quasi rectangular waves. The reasons of this choice will be explained during the next lessons. Thus, the output signal is not combination of different HDR translated in time but are combination of quasi rectangular waves translated in time and centred respect to the activation time interval of that specific task.

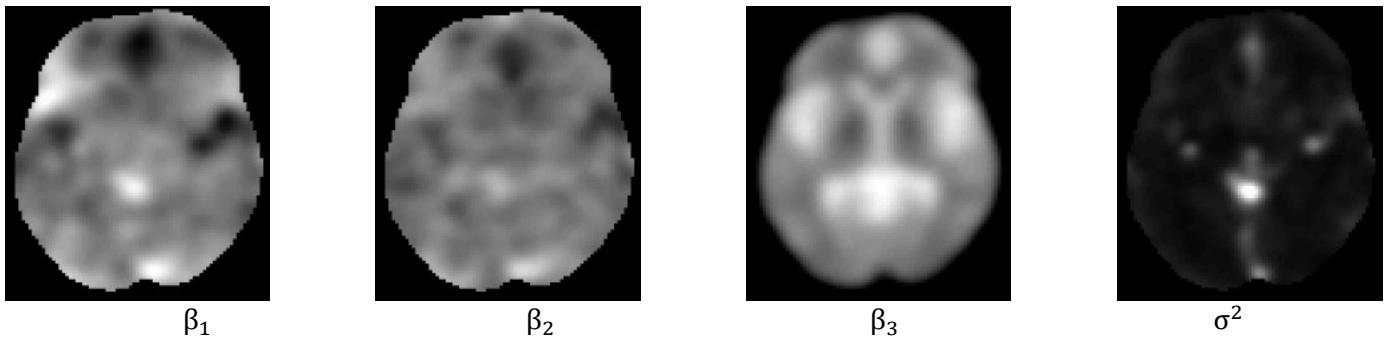
After having applied an estimation process to find the correct values of the parameters vector to give to a specific voxel we could establish which signal between the one modelled present a higher incidence to correctly fit the output of the that specific voxel. By considering the previous example, down below are reported the results of the estimation process.



From the images reported above we can see the different weight that are gives to each output task of the model. Another important parametric map that we can acquire, and compute represents the sum of square errors. This last concept is defined from the analytical expression:

$$\sigma^2 \propto \sum_{i=1}^n e_i^2$$

Down below are reported the three parametric maps that are related to the three components of the parameter vector represented all over the head of the subject.



fMRI: Statistical analysis – General Linear Model: Estimation

The estimation of each component of the parameter's vector is done by starting from the next relation:

$$Y = X\beta + \epsilon$$

With $\epsilon \sim \sigma^2 N(0, I)$. The information about the distribution of the measurement error could be known of unknown. Consequently, the relation to use to correctly perform the estimation process are the next ones:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad \hat{\epsilon} = Y - X\hat{\beta} \quad \hat{\sigma}^2 = \frac{[\epsilon^T \epsilon]}{N - P} \quad \Sigma_\beta = \hat{\sigma}^2 (X^T X)^{-1}$$

Parameter's vector

Residuals

Variance posteriori of
the measurement
error

Variance of the
parameters

The assumption about measurement error id est additive uncorrelated and gaussian with zero mean and variance σ^2 are not completely satisfied by fMRI data. In fMRI data, there is a structural correlation which is related to the low frequency physiological noise linked to hearth beat and breath and to unmodelled neural activity. Using assumptions about error uncorrelation we obtain estimates affected by error bias. Thus, a more adapt model is:

$$Y = X\beta + \epsilon$$

$$\epsilon \sim N(0, \sigma^2 V)$$

Where V is not a diagonal matrix. To obtain unbiased estimates it is necessary to pre-multiply the model by the matrix $W = V^{-1/2}$. By applying the product for this matrix to all the terms of the estimation equation we obtain:

$$WY = WX\beta + W\epsilon = WX\beta + w$$

Where $w \sim N(0, \sigma^2 I)$. Now we can really generate the β s and the methodology to estimate this variable are the same as the one seen before. The reason why could be used a matrix W characterized with this structure is the next one. If we are searching for the matrix W that let the covariance matrix of the measurement error be diagonally dominant we need to calculate the variance of the matrix $W\epsilon$.

$$E[(W\epsilon - E[W\epsilon])^2], E[W\epsilon] = W \times 0 = 0 \Rightarrow E[W^2\epsilon^2] = W^2E[\epsilon^2] = W^2Var[\epsilon] = W^2\sigma^2V = \sigma^2W^2V$$

To let the last term of the equality be a diagonal matrix we need to select the value of W that satisfies the next equality:

$$\sigma^2W^2V = \sigma^2I \Rightarrow W^2V = I \Rightarrow W = V^{-1/2}$$

The main problem to solve now is how we can generate the matrix V .

The way that we may use to determine the matrix V is by the usage of an autoregressive model of first order. To do so, we are assuming that the error temporal variability could be described from a first order autoregressive model. In this last one the output is determined from the previous step and the noise term. Mathematically:

$$\epsilon(t) = \rho\epsilon(t-1) + \xi_1$$

Where the parameters presented in the relation could follow the next description.

$$|\rho| < 1$$

$$\xi_1 \sim N(0, \sigma_1^2)$$

With uncorrelated values. To identify the model, it is necessary to estimate the parameters ρ and σ_1 , that are respectively the model coefficient and the white noise variance. The procedure to build the V matrix requires the usage of the Yule Walker equations. Those equations describe the output process autocorrelation of the AR model. The V matrix is made by the Yule Walker equation coefficients and results to be a Toeplitz matrix id est it is symmetric with diagonal constant elements and positive semidefinite. This has the next analytical and mathematical structure.

$$V = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}$$

ρ is assumed to be constant over all the voxels even if this is not true and is computed as the mean of all values obtained in all voxels. Once V is estimated, we use the Cholesky factorization to obtain:

$$V = (H)^T H$$

Thus:

$$W = (H^T)^{-1}$$

More in details, the W matrix results to be:

$$W = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -\rho R & R & 0 & \dots & 0 \\ 0 & -\rho R & R & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -\rho R & R \end{pmatrix}$$

Where the value of R is equal to:

$$R = \frac{1}{\sqrt{1-\rho^2}}$$

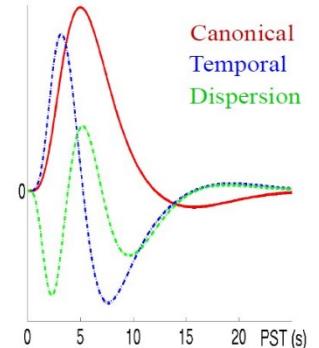
fMRI: Statistical Analysis – General Linear Model: HDR model matrix

The matrix $X(t)$ is build (as previously shown) using the information of the experimental design, $u(t)$ and an appropriate function $h(t)$ that represents the HDR. To describe $h(t)$ several models have been proposed. In the literature we have available different types of HDR model. These are:

- Canonical model
- Temporal model
- Dispersion model
- Fourier Set

These ones present the shape and follows the explanation reported below. The first three models are represented in the figure reported on the left side of the page. From this we can see the different shape that can be used to model the HDR.

- *Canonical model*: This model is the one that represents most HDR. It is largely used in literature and has a specific shape.
- *Temporal model*: This model has been introduced recently and could be used to enhance the detection of the hyperpolarization at the end of the HDR. The temporal model mathematically is expressed by the same formulation for the canonical model but with the addition of the Taylor expansion till the first derivative (in time).
- *Dispersion model*: This was first introduced in the late '90s. In the first 0.5 second of stimulation researchers have observed at that time a first slightly downwards before the peak. Only few research groups were able to replicate this finding. The hypothesized mechanism is that before the flow response has a chance to start, a more rapid increase in oxidative metabolic rate causes the blood to become transiently less oxygenated. This transient effect is then washed away by the large increase in flow that follows. Just a few research papers have spoken about this topic and also nowadays it is not well understood when this deflection could be seen. Anyway, even if the first depolarization is not seen



always, we introduce a dispersion model that it is used to understand if the HDR evoked present this shape.

- *Fourier set:* In this case there is another alternative. This is made up of the Fourier set which consist to describe the BOLD signal detected trough a linear combination of cosine and sine functions with period of T_H , $\frac{T_H}{2}$ and finally $\frac{T_H}{N_S}$.

The main problem with all these types of techniques is that they give us different results which do not present any relations between each other. To find better parameters map we don't have to plot directly the value of a certain parameter estimated for each voxel, but we need to perform some statistical analysis that consist in the application of particular statistical test.

fMRI: Statistical Analysis - Contrast

The contrast is a vector of integer values both positive and negative. By referring to the previous example where in the fMRI acquisition we are asking to the subject to perform three different task we can define the following contrast vector. A contrast in general is represented as a linear combination of different variables. Once define this combination we can express it analytically as:

$$\sum_j \alpha_j \bar{X}_j$$

Where the linear combination presents some coefficient α_j . The contrast vector is made up of this coefficient α_j . This one has to satisfy the mathematical requests that the sum up of all the coefficient α_j need to be equal to zero. It is an important and strong request that needed to be satisfied. Moreover, different contrast vectors could be orthogonal. This is verified if and only if the inner product between the two vector is equal to zero. From the statistical point of view, the contrast is used to compare different population means. In our application, would be used to test differences between different task considered into the acquisition.

fMRI: Statistical Analysis – Statistical test

As said before to achieves better results for the description of the different parametric map we may use statistical test. In fMRI data analysis the t value is called also t contrast. The mathematical expression reported below represents the test applicable in our situation after having perform a general linear model analysis.

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

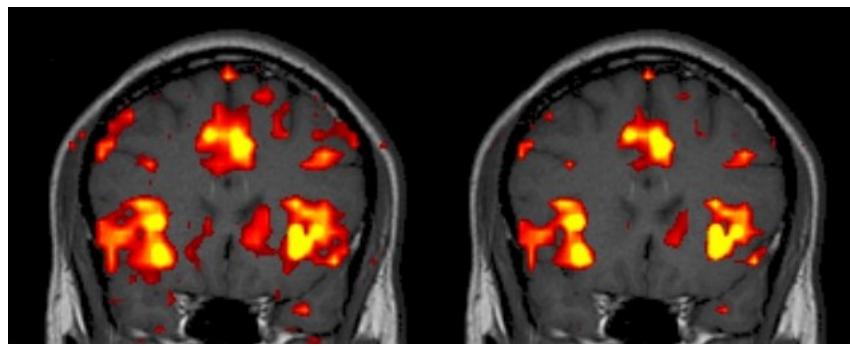
In the numerator we have the parameter weighted for the contrast, while in the denominator there is the uncertainty term. This is made up of residual weighted for the contrast vector. Once the t value is computed we have to remember that we need to build up the statistics. We need to understand which function is generated to establish if the statistical test should be rejected or not. The information available in our situation are *the degree of freedom (df)* and *the significance threshold (α)*. The distribution of the statistical test depends just one the value of the degree of freedom. Instead, the other parameter α is needed to

understand if statistical test computed should be rejected or not. It is possible to show that the degree of freedom in this case of t – Student distribution is equal to $N - p$. The value of N represents the number of acquired volumes (id est the number of samples for each voxel), while p represents the rank of X (id est the maximum number of linearly independent columns of X). One other important information about the analysis and the interpretation of the results of a statistical test is to define the null and the alternative hypothesis. In this case we have:

$$H_0: \text{Different Task}$$

$$H_1: \text{Equivalent Task}$$

To obtain the different results we need to establish in each voxel if the null hypothesis has to be accepted or not. To do so we need to compare the value of the statistic t with the threshold t_α found by the usage of α selected. By the application of diverse values of α we could obtain different results as could be seen in the images down below.



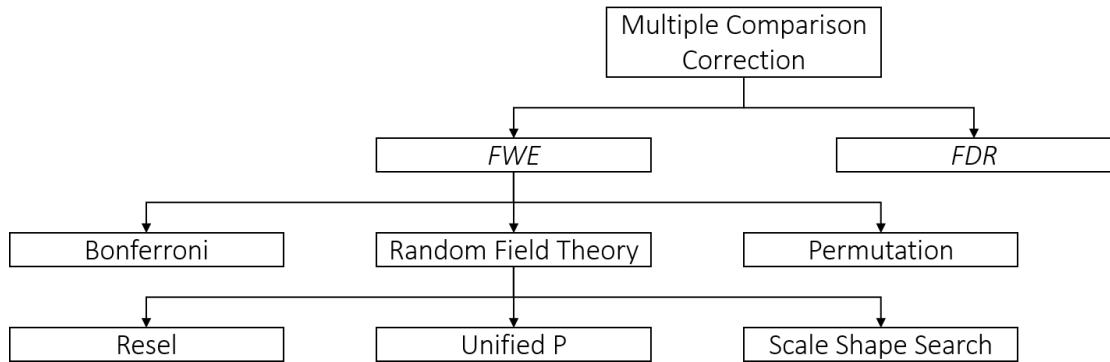
These images represent the results of the elaboration by the usage of two distinct p – values. The image on the left represents the one elaborated with $p < 0.05$ while the one on the right represents the one elaborated with $p < 0.001$. One important problem that invalidate all the results achieved by just the application of a statistical test is the problem of multiple comparison. In the next paragraph we will discuss about the methodologies applicable to solve this problem.

fMRI: Statistical Analysis – Multiple comparison

In those case study where we need to perform a lot of statistical test some problem may arise. This is related with the definition of the significant threshold α used. It represents the percentage of probability to classify a voxel incorrectly. If the value of α is of 0.05 this means that on 100 voxel we will for sure misclassify 5 of them. Obviously, in a fMRI application the number of voxels is not just 100 but are a lot of more and this leads to a higher number of voxels misclassified if still used the significance threshold equal to 0.05. To solve this problem are proposed, in the next paragraphs, some specific techniques. These are the so – called the multiple comparison correction approaches. These are divided mainly in two families:

- **FWE:** It stands for the acronym *Family Wise Error Rate*. This type of techniques controls this quantity which is represented by controlling the chance that a voxel is classified as a false positive.
- **FDR:** It is the acronym for *False Discovery Rate*. With these approaches we could control the fraction of false positive.

By looking to a more complete situation in these two families we could find different methodologies.



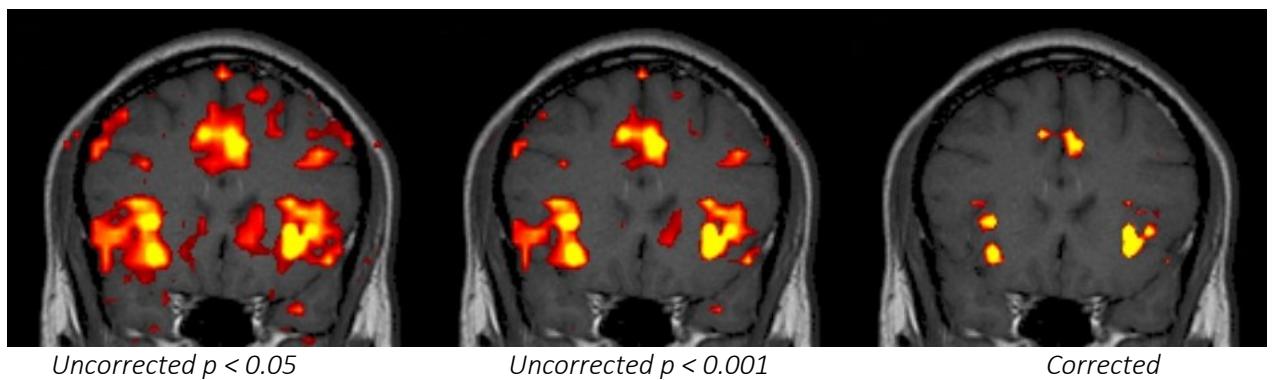
All the techniques that are listed and reported in the scheme above will be explained.

fMRI: Statistical Analysis – Multiple comparison: Bonferroni

In the Bonferroni correction we have to divide the threshold with the number of comparisons that we need to do. In principle, the Bonferroni correction controls the type I errors (classification of a determine test as false positive). First, we need to perform the statistical test and get all the possible values of the statistic and then we need to compare these ones with the corrected value of the significance threshold α . In fact, initially we decide to perform a test with a certain level of significance but before performing all the possible comparison we need to correct this value by dividing it with the number of all the possible test needed to be executed. Therefore, the number the overall probability that a certain voxel in the entire brain is misclassified is equal to:

$$\alpha_{corrected} = \frac{\alpha}{N}$$

There are some pitfalls of this method as instance the possibility of an increase probability to commit an error of type II. This is such a problem sometimes in some field like neurosurgery. Usually, the Bonferroni test is typically overly conservative but due to a lot of spatial autocorrelation we face some problem and only a few voxels survive due to the unsatisfaction of the independence hypothesis.



fMRI: Statistical Analysis – Multiple comparison: Random Field Theory

Bonferroni is based on the independence hypothesis between different variable to compare, but due to the high spatial correlation present between neighbour voxels it cannot be applied. To solve this problem is introduced the *Random Field Theory (RFT)*. In this field it is used the measure of the roughness of a random field. The roughness measures the level of smoothness of the image. First, we need to discretize the image and then we need to compare all with a certain threshold. After having done this is possible to establish the value of the *Euler characteristic* (EC) for that specific value of threshold. The EC represent the number of blobs (separated region in the threshold image). Higher is the threshold lower is the EC. This index is a measure of roughness in *RTF*. This approach leads to control the probability in FWE by evaluating the expectation of the random variable EC. The analytical relation, proposed by Worsley, is the next one:

$$\alpha = E[EC] = R(4 \ln 2)(2\pi)^{-3/2}ue^{-\frac{1}{2}u^2}$$

Where R represent the number of resels. This as definition suggest is a resolution element. The number of resels in an image is similar to the number of independent observations in the image. However, they are not the same as the number of pixels of a given image. A resel is defined as a block of pixels of the same size as the FWHM of the smoothness of the image. In the previous forma exists also the letter u which represent the voxel wise threshold. Even if this methodology corrects some issues of the Bonferroni correction there are some limitations that are the requirement of a sufficient smoothness because if not the estimate is biased. Another stronger assumption that it is needed to be satisfied when perform this type of task is the multivariate normality which is very difficult to check.

fMRI: Statistical Analysis – Multiple comparison: Permutation testing

It is a non – parametric method, and this means that we are not assuming a specific distribution of the data. This is able to give us more reliable result. The permutation test means that we need to compute the statistic t for real data. We need to shuffle labels N times and recompute the statistic and repeat all the process. First, we need to compute the initial statistic value, then we need to create all the possible permutation between the sequences and compute again the statistic for each permutation. After having evaluated all the possible permutation we need to count the number of combinations that present a value of the statistic greater than the first one. After having done this operation, we need to compute the p – value by calculating the ratio between the previous identified value and the total amount of permutation evaluated. This is compared then with the significance threshold and if it is below this last value, we can reject the null hypothesis. This methodology requires only exchangeability and so if we want compares different subjects, we could use this method. Inside the same individual it cannot be used because there are some problems of temporal autocorrelation.

fMRI: Statistical Analysis – Multiple comparison: False Discovery Rate

To control the FDR instead of the FWE we need to follow the next steps. First, we need to compute the p – value for each voxel (without any other correction). Then we need to sort the p – values in an ascending fashion and we need to find the index h such that $p_h < \frac{\alpha h}{N}$. All the statistical test associated to the p – values lower than p_h are then rejected.

```

1 %
2 % Title: Laboratory 4
3 %
4 % Note:
5 % PET and Spectral Analysis
6 %
7 % Author: Matteo Martin
8
9 % THEORY -----
10 %
11 % SPECTRAL ANALYSIS
12 % The Spectral Analysis is a way that we can use to determine the
13 % reversibility or the irreversibility of the compartmental model used.
14 % In particular, with this we can also compute the exact number of the
15 % compartments to use to describe our compartmental model structure.
16 % Remember that the steps that are needed to perform correctly the spectral
17 % analysis are the one listed below:
18 %
19 % - Estimate the grid of the different Betas
20 % - Evaluate the convolution between the plasmatic concnetration and the
21 % exponential function raised to a certain value of betas determined.
22 % - Estimate the different coefficient alpha which need to be strictly
23 % positive to reduce the sparsity of the solution.
24 % - Visualization in a beta/alpha plane of the different spectral lines
25 %
26 % VARIABLES
27 % idif : Represent the plasmatic tracer
28 % concentration corrected for metablites.
29 % [kBq/ml]
30 %
31 % T_idif : Represent the different instant of time
32 % where the samples are taken.
33 % [s]
34 %
35 % t : Represent the duration of the acquisition
36 % of PET data.
37 % [s]
38 %
39 % PET_DYN_rl.nii : Dynamical PET data contained inside a 4D
40 % matrix.
41 %
42 % Hammers_2_PET_GM.nii : Represent the Atlas that we need to use
43 % inside our analysis. The matrix is a 3D
44 % matrix.
45 %
46 % beta : Containes the beta's grid.
47 % [min-1]
48 %
49 % NOTE
50 % An additional note very important to correctly perform the entire
51 % analysis is that the PET data are already coregistered over the Atlas.
52 % Thus, when we will select different anatomical regions by referring to
53 % the Atlas.
54 %

```

```

55 % Another important information is that, when we inspect the concentration
56 % behaviour associated to the different anatomical region we can easily
57 % select which ones have the higher uptake and lower uptake by looking on
58 % the last value of the concentration time series.
59
60 %% CCC
61 % Clear all, close all, clear the command window
62
63 close all; clear all; clc
64
65 %% DATA - LOAD
66 % Load of all the possible data necessary for this Laboratory
67
68 load("T_idif.mat"), load("timePET.mat"),
69
70 %% TIME - GENERATION
71 % Conversion of all the time grids in minutes with 0 as first time points.
72
73 timePET = [0, cumsum(t) - t/2]/60; % Time instant of PET analysis
74 timePCC = T_idif/60; % Time instant of Plasmatic Conc.
75
76 q1 = 0.001; % First time step quantization
77 q2 = 0.1; % Second time step quantization
78
79 timePETG1 = timePET(1):q1:timePET(end); % First time grid
80 timePETG2 = timePET(1):q2:timePET(end); % Second time grid
81
82 clear t; clear T_idif;
83
84 %% DATA - LOAD
85 % Load of the anatomical atlas and the PET data
86
87 ATLAS = load_untouch_nii('Hammers_2_PET_GM.nii');
88 ATLAS = double(ATLAS.img);
89
90 PET = load_untouch_nii('PET_DYN_rl.nii');
91 PET = double(PET.img);
92
93 %% ATLAS - MANIPULATION
94 % Remove from the ATLAS of all the possible not usable anatomical regions.
95 % We need to select just the regions with more than 10 voxels associated to
96 % them or the one that are not cited in the vector RM. The ones that are
97 % listed in the vector RM will be removed from the ATLAS as requested.
98
99 M = max(ATLAS,[],'all');
100 m = min(ATLAS,[],'all');
101
102 TH = 10; % Region threshold size
103 RM = [3,4,17,18,19,44,74,75,45,46,47,48,49]; % Region to remove
104
105 nVOXEL = zeros(1,M);
106
107 for idX = 1:1:M
108     nVOXEL(idX) = sum(ATLAS == idX,'all');

```

```

109     if nVOXEL(idX) < TH, ATLAS(ATLAS == idX) = 0; end
110     if not(isempty(find(RM == idX,1))), ATLAS(ATLAS == idX) = 0; end
111 end
112
113 %% PET - TIME SERIES - EXTRACTION
114 % Time series extraction in correspondence to those specific anatomical
115 % region still presented inside the ATLAS. To do so we need to apply the
116 % next few simple steps.
117 %
118 % - Region's index generator: Generation of a vector which will contain
119 % all the possible index contained inside the
120 % ATLAS. These indexes represent the selected
121 % anatomical brain region at the step before.
122 %
123 % - Time series extraction : The time series extraction is done by first
124 % selecting the index of the region that we
125 % need to extract and then select all the
126 % voxel in the PET image by passing through
127 % the Atlas.
128
129 iREGION = unique(ATLAS);
130 iREGION = iREGION(2:end);
131 nREGION = length(iREGION);
132
133 [nROW,nCOL,nSLICE,nTIME] = size(PET);
134 timeSeries = zeros(nREGION, nTIME);
135
136 for idX = 1:1:nREGION
137     ATLASREGION = find(ATLAS == iREGION(idX));
138     for jdX = 1:1:nTIME
139         PETTIME = squeeze(PET(:,:,:,:,jdX));
140         timeSeries(idX,jdX) = mean(PETTIME(ATLASREGION),'all');
141     end
142 end
143
144 timeSeries = [zeros(nREGION,1), timeSeries];
145
146 % This last line of code is needed to set the initial value of the time
147 % courses to zero. This because at the initial instant of time, which is
148 % zero, the concentration of the radioactive tracer inside our head is
149 % null.
150
151 %% PET - TIME SERIES - VISUALIZATION
152 % Visualization of all the possible time series extracted
153
154 figure, hold on
155     plot(timePET,timeSeries)
156     hold off, grid on, grid minor, box on,
157     title('PET - Time Series'), xlabel('Time [min]'),
158     ylabel('Concentration [kBq/ml]')
159     xlim([0 timePET(end)])
160
161 %% PET - TIME SERIES - MAXIMUM MINIMUM UPTAKE
162 % Evaluation of the PET maximum and minimum uptake by inspecting the last

```

```

163 % value of the concentration time series and finding the correspondence to
164 % a determine anatomical brain region.
165
166 [M, IM] = max(timeSeries(:,end));
167 [m, Im] = min(timeSeries(:,end));
168
169 fprintf('+++++++\n')
170 fprintf('PET - MAXIMUM & MINIMUM UPTAKE \n')
171 fprintf('Region with lower uptake: %i\n', iREGION(Im))
172 fprintf('Region with higher uptake: %i\n', iREGION(IM))
173 fprintf('-----\n')
174
175 %% BETA GRID - LOAD
176 % Load of the beta value's to create the time grid needed to be used to
177 % perform the spectral analysis.
178
179 load('beta.mat'),
180 load('idif.mat'), Cp = idif; clear idif           % Rename
181
182 %% BETA GRID - EXPONENTIAL GRIRD
183 % Exponential grid generation by the usage of the timePETG1 and the values
184 % contained into the variable beta loaded at the previous step
185
186 nBETA      = length(beta);                      % Numbers of beta
187 expBetas = zeros(nBETA, length(timePETG1));       % Matrix initialization
188
189 for idB = 1:1:nBETA, expBetas(idB,:) = exp(-beta(idB)*timePETG1); end
190
191 %% BETA GRID - EXPONENTIAL GRID VISUALIZATION
192 % Visualization of the exponential generated by the stem function all
193 % inside the same chart.
194
195 figure, hold on
196     for idB = 1:1:nBETA, stem(timePETG1, expBetas(idB,:)), end
197         grid on, grid minor, box on,
198         title('EXPONENTIAL GRID GENERATED')
199         xlabel('Time [ms]'), ylabel('Dimensionless []')
200         xlim([timePETG1(1), timePETG1(end)])
201
202 %% BETA GRID - GENERATION
203 % Generation of the grid by the usage of the filter function of the
204 % different value of B(i) to compute then the spectral analysis.
205
206 Cp      = interp1(timePCC, Cp, timePETG1);        % Interpolation on grid 1
207 B       = zeros(nBETA+1,length(timePETG2));        % Grid initialization
208
209 for idB = 1:1:nBETA
210     temp      = q1*conv(Cp,expBetas(idB,:));
211     B(idB,:) = interp1(timePETG1, temp(1:length(timePETG1)), timePETG2);
212 end
213
214 B(nBETA+1,:) = interp1(timePETG1, Cp, timePETG2);
215
216 clear temp

```

```

217
218 % Here we can use either two function:
219 %
220 % - conv: In this the output is represented with a vector made up of
221 % an higher dimension then the two passed as input. Hence, handle
222 % the output we have to consider just the beginning samples.
223 % - filter: In this case the output is of the same dimension of the input
224 % vector. We have just to multiply for the step in time.
225 %
226 % In all the two cases reported above we need always to put in front of the
227 % conv or the filter function the step in time. This is requested
228 % mathematically because the step that is considered in these two function
229 % is homogeneous (thus the time grid need to be homogeneous) and unitary.
230
231 %% BETA GRID - VISUALIZATION
232 % Visualization of the B(i)'s grid by plotting with the stem function the
233 % behaviour of the convolution between the concentration Cp measured and
234 % the exponential of beta(i) times the time grid.
235
236 figure, hold on,
237 plot(timePETG2, B,'-')
238 hold off, grid on, grid minor, box on,
239 title('B(i) GRID GENERATED'),
240 xlabel('Time [ms]'), ylabel('Concentration [kBq/ml]')
241 xlim([timePETG2(1) timePETG2(end)])
242
243 %% BETA GRID - PET TIME INTERPOLATION
244 % Interpolation of the different function contained in the matrix of Beta's
245 % into the timePET acquisition grid. This is done to then compute the
246 % estimation of the different alpha's value and then represent the spectrum
247 % of each identified ROI.
248
249 temp = B; % Temporary variable
250 B = zeros(nBETA+1, length(timePET)); % Initialization
251
252 for idB = 1:1:(nBETA+1)
253     B(idB,:) = interp1(timePETG2, temp(idB,:), timePET);
254 end
255
256 clear temp
257
258 ! TO DO - DOMANDA: Perché passiamo per tre griglie temporali? Ok le ultime
259 ! TO DO - DOMANDA: due ma non capiamo perché la transizione da prima a seconda.
260 % Vengono utilizzate due time grid diverse perché più affinate e meno
261 % affinate si ottengono risultati bene o male probabilmente leggermente
262 % diverse. Non risulta un'errore considerare solamente una griglia ma
263 % quando si valutano integrali e convoluzioni ha senso prendere in esame
264 % sempre delle griglie con diversi livelli di quantizzazione.
265
266 %% ALPHA GRID - COMPUTATION
267 % Computation of the different alpha's grid and values for each ROI
268 % identify inside the previous ATLAS analysis operation.
269
270 A = zeros(nBETA+1, nREGION);

```

```

271
272 for idR = 1:1:nREGION
273
274     CROI      = timeSeries(idR,:)';
275
276     options  = optimset('TolX',1e-5, 'TolFun', 1e-5, 'Display', 'off', ...
277                         'MaxFunEvals',10e5,'MaxIter',1000);
278     A(:,idR) = lsqnonneg(B',CROI, options);
279
280 end
281
282 %% ALPHA GRID - RANDOM VISUALIZATION
283 % Visualization of the alpha values calculated for each ROI that we have
284 % identified during the pre processing phase.
285
286 nROI      = 5;
287 indexROI = ceil(nREGION*rand(nROI));
288
289 figure, for idR = 1:1:5
290         if idR == 5, subplot(3,2,[5 6]), else, subplot(3,2,idR), end
291         stem(beta, A(1:nBETA,indexROI(idR)), 'm*')
292         title(sprintf('SPECTRAL ANALYSIS ROI - n° %i',indexROI(idR)))
293         xlabel('\beta [1/ms]'), ylabel('\alpha []')
294         grid on, grid minor, box on
295     end
296
297 %% ALPHA GRID - ALL VISUALIZATION
298 % Visualization of all the possible alpha values estimated for each ROI
299 % inside the ATLAS manipulated and considered. The visualization is tuned
300 % for 68 region of interest to visualze. In this analysis the successive
301 % information are displayed and computed:
302 %
303 % - Graphical visualization ROI's spectrum
304 % - Evaluation of each ROI's number of spectral lines
305 % - Evaluation of the number of irreversible compartment
306 % - Evaluation of the number of reversible compartment
307 % - Comand window visualization of the ROI's charateristic
308 % - Memorization of irreversible and reversible compartment information
309
310 figure('units','Normalized','Outerposition',[0 0 1 1]),
311 fprintf('+++++++\n')
312 fprintf('----- SPECTRAL ANALYSIS RECAP -----\\n')
313 fprintf('REGION | IRREVERSIBLE/REVERSIBLE | n LINES \\n')
314
315 idS = 1; nSPECTRALLINES = zeros(1,nREGION);
316 nREVERSIBLE = 0; iREVERSIBLE = zeros(1,nREGION);
317 nIRREVERSIBLE = 0; iIRREVERSIBLE = zeros(1,nREGION);
318
319 for idR = 1:1:nREGION
320
321     subplot(2,2,idS)                                % VISUAL
322     stem(beta, A(1:nBETA, idR), '*-')              % VISUAL
323
324     nSPECTRALLINES(idR) = sum(A(1:nBETA,idR)>0);    % ELAB

```

```

325 if A(1,idR) == 0                                % ELAB
326     nIRRVERSIBLE = nIRRVERSIBLE + 1; iIRRVERSIBLE(idR) = 1;    % ELAB
327     TYPE        = '           Irreversible      ';
328 else
329     nREVERSIBLE = nREVERSIBLE + 1; iREVERSIBLE(idR) = 1;    % ELAB
330     TYPE        = '           Reversible       ';
331 end
332
333 title(sprintf('SPECTRAL ANALYSIS ROI - n° %i, LINES = %i ',... % VISUAL
334     iREGION(idR), nSPECTRALLINES(idR)))          % VISUAL
335 xlabel('\beta [1/ms]'), ylabel('\alpha []')         % VISUAL
336 grid on, grid minor, box on                         % VISUAL
337
338 if idS == 4 && not(idR == nREGION)                % VISUAL
339     idS = 1; pause(1);                            % VISUAL
340     figure('units','Normalized','Outerposition',[0 0 1 1]), % VISUAL
341 else
342     idS = idS + 1;                            % VISUAL
343 end
344
345 fprintf('%2i      |%s| %2i\n',iREGION(idR),TYPE,nSPECTRALLINES(idR))
346
347 end
348
349 fprintf('+++++++\n')
350
351 % From the chart reported before we can see the presence of different
352 % spectral charts that shows the presence of a number higher than 3
353 % spectral lines. The maximum number of lines detectable is equal to 7
354 % while the minimum is equal to 3. A lot of ROIs even if they present an
355 % high number of spectral lines usually these are duplication of the same
356 % spectral line. All the ROIs as could be seen from the output
357 % reported in the comand window shows an irreversible behaviour.
358
359 %% CLUSTERING - K MEANS
360 % Application of a K - means clustering technique to reduce the number of
361 % ROIs to analyse. The clustering techniques is performed to find 12 major
362 % centroid that describes our dataset.
363
364 K      = 12;          % Number of clusters
365 [IDX, C] = kmeans(timeSeries,K,'Display','final',...
366                     'Distance','squeuclidean','MaxIter',200,...
367                     'Replicates',10);
368
369 %% CLUSTERING - VISUALIZATION
370 % Visualization of the resutls of the clustering technique applied to the
371 % previous step.
372
373 figure, hold on,
374     for idK = 1:1:K, plot(timePET, C(idK,:),'-'), end
375     hold off, grid on, grid minor, box on,
376     title('CLUSTERING K - MEANS - 12 Cluster')
377     xlabel('Time [ms]'), ylabel('Concentration [kBq/ml]')
378     xlim([timePET(1) timePET(end)])

```

```

379
380 %% LOGAN - NUMBER OF TIME INSTANT
381 % Evaluation of each the logan plot to determine the necessary time intasnt
382 % where we have to fit the regression line.
383
384 Cp      = interp1(timePETG1, Cp, timePET);
385 ICp     = cumtrapz(timePET,Cp);
386
387 figure('units','Normalized','Outerposition',[0 0 1 1]),
388 idX = 1;
389
390 for idK = 1:1:K
391     Ck = C(idK,:);
392     IC = cumtrapz(timePET, Ck);
393
394     X = ICp./Ck;
395     Y = IC./Ck;
396
397     subplot(3,4,idX), hold on, plot(X,Y,'b-*','LineWidth',2),
398                 plot(X,X,'k--'), hold off,
399                 grid on, grid minor, box on,
400                 title(sprintf('LOGAN PLOT %i',idK)),
401                 xlabel('X [ICp/C]'), ylabel('Y [IC/C]')
402     idX = idX + 1;
403 end
404
405 TSTART = 18;
406
407 %% LOGAN - REGRESSION
408 % Estimation of the Volume Distribution for each ROIs. The Logan plot fits
409 % the data contained inside the last 11 samples of the time courses.
410
411 Vd = zeros(nREGION,1);
412
413 for idK = 1:1:K
414
415     SELECTED = find(IDX == idK);
416     nSELECTED = length(SELECTED);
417
418     for idR = 1:1:nSELECTED
419         Ck = timeSeries(SELECTED(idR),:);
420         IC = cumtrapz(timePET, Ck);
421
422         X = ICp./Ck; X = X(TSTART:end)';
423         Y = IC./Ck; Y = Y(TSTART:end)';
424
425         P = ((X'*X)^(-1))*X'*Y;
426         Vd(SELECTED(idR)) = P(1);
427     end
428
429 end
430
431 %% LOGAN - PLACEMENT
432 % Placement of the volume distribution value in the correct location.

```

```
433
434 TEMPET = zeros(size(PET));
435 TEMPATL = repmat(ATLAS, [1,1,1,nTIME]);
436
437 for idR = 1:1:nREGION
438
439     SELECTED = find(TEMPATL == iREGION(idR));
440     TEMPET(SELECTED) = Vd(idR);
441
442 end
443
444 ! TO DO: DOMANDA - La parametric map deve essere stimata complessivamente
445 ! TO DO: DOMANDA - valutando il singolo VOXEL oppure valutando la ROI?
446
```

LESSON 12: FMRI AND DEFAULT MODE NETWORK

Integration and Segregation: Introduction

To introduce the argument that speak about the connectivity analysis of the brain we need to introduce first two important definition that are recurrent in this field. These are:

- *Integration*: The functional integration is the study of how brain regions work together to process information and effect responses. Though functional integration frequently relies on anatomic knowledge of the connection between brain areas, the emphasis is on how large cluster of neurons – numbering in the thousands of millions – fire together under various stimuli.
- *Segregation*: The functional segregation instead, refers to the ability of the brain to work as a group of clusters and keeps up different processing phase simultaneously in a parallel way.

Default Mode Network: Introduction

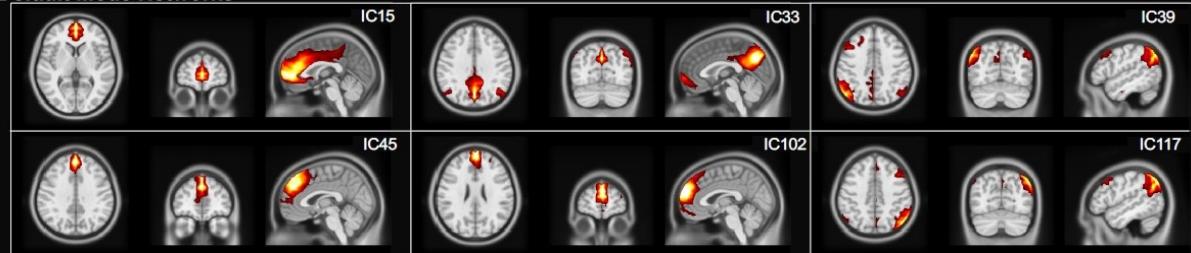
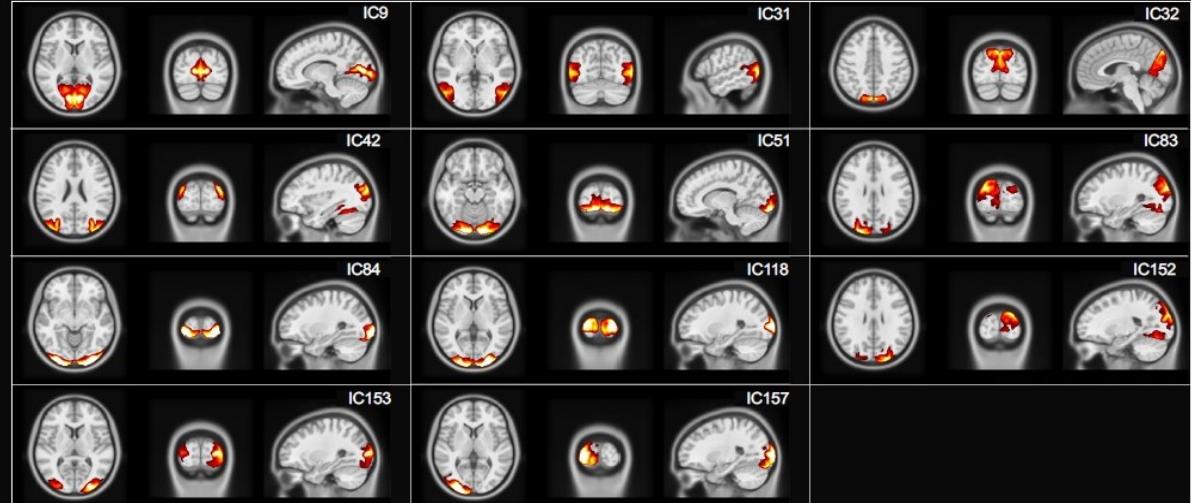
To study these two concepts on functional maps we need to use specific data that comes from our analysis. This last one has to be performed in situation where the patient is in a relaxed state to get good information about the resting state network characteristic of that subject.

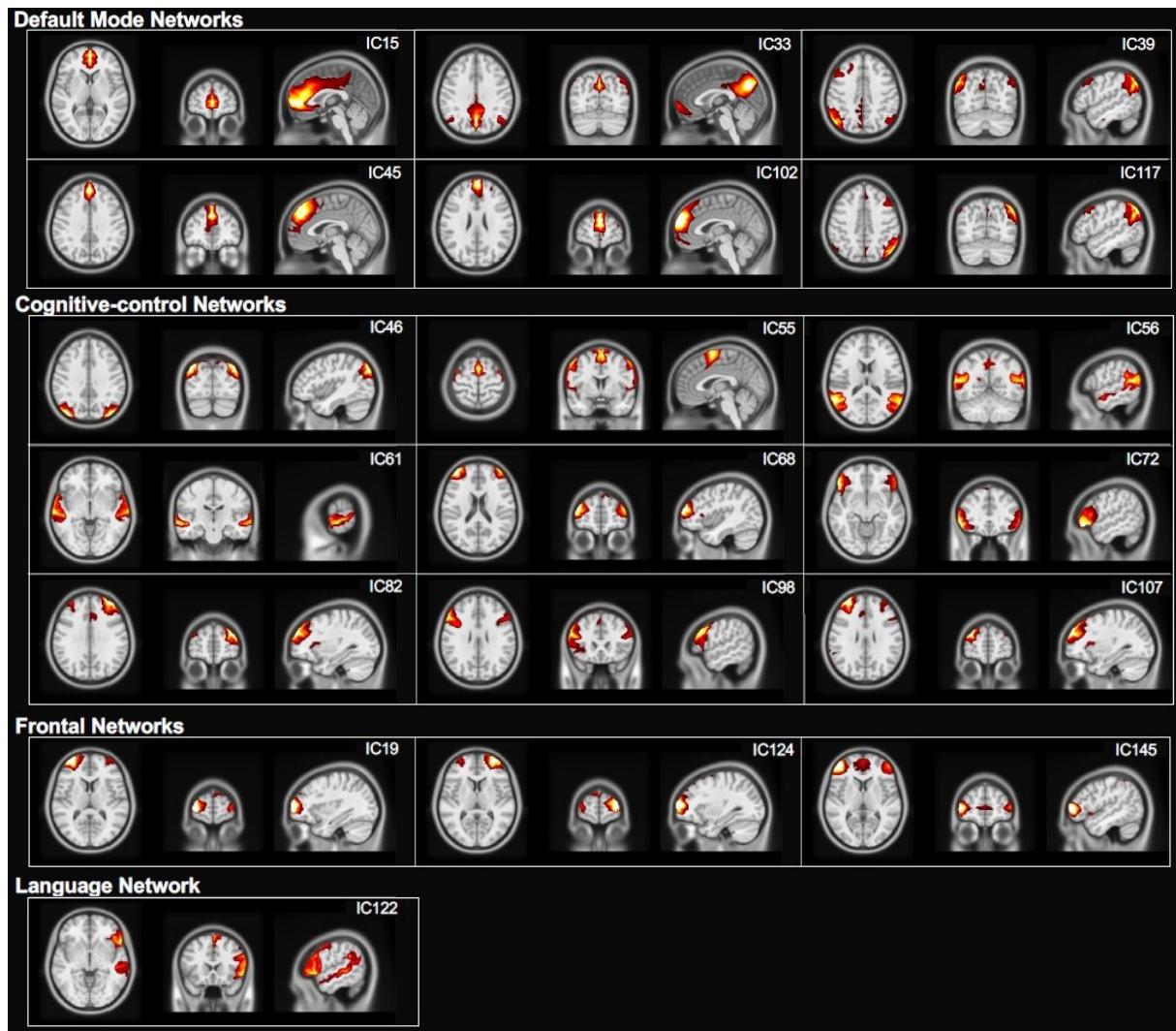
A common observation in brain imaging research is that a specific set of brain regions referred to as default network is engaged when individuals are left to think to themselves undisturbed. In this state the subject is performing a so-called *do-nothing* task and even if this last one is apparently doing nothing the brain is still working and processing information. From data retrieved during this state we can acquire a lot of information about on how the subject's brain is interconnected. These types of investigation started with Louis Sokoloff and colleagues in 1955. They were requested to study the cerebral metabolism changes globally when one subject goes from a quiet rest state to performing a challenging arithmetic problem or in general a task that demands focused cognitive effort. To their surprise, metabolism remained constant. All the results were resumed in the published paper that read:

Having thus resolved the question of site, we have attempted to determine whether the mental processes involved in the performance of mental arithmetic are of a type in which energy utilization or transformation occurs. A positive result that through processes require the utilization of additional quantities of oxygen by the brain would have led to such a conclusion. In view of our negative results, however, the conclusions are fat less obvious, and we must consider several possible interpretations. [...]

The oxygen consumption of this fraction of the brain may be so small even under conditions of maximal activity, that we are unable to detect changes in it by means of a relatively crude method for measuring total cerebral oxygen consumption.

Thus, this first results means that there is always an activity during also the *do-nothing* condition and the study of this state could be more relevant than though at that time. Nowadays, the study of this condition is such important and let all the researchers define from data acquire during this type of task the so-called *Default Mode Network (DMN)*. This concept in history was first introduced from Raichle, Gusnard and colleagues. In the following pages are reported some examples of nowadays studies on the DMN and the objective currently reached.

Default Mode Networks**Visual Networks****Sensory-Motor Networks****Auditory Network****Cingulo-opercular Networks****Dorsal-attention Networks****Fronto-parietal Networks**



Default Mode Network: Instrumentation

The instrumentation that are used to investigate the brain networks could be different and usually we can recognize mainly, PET which was used at the beginning of these analysis, fMRI that is currently largely used and it is the best way to get information about the brain state. Finally, another alternative to investigate the connectivity between different regions is the EEG. This last one, even if it is a less expensive exam gives results that are not reliable as the one obtained with fMRI.

Default Mode Network: Experimental design

The experimental design that it is used to get some information about the brain's state is a large and long resting state condition. In fact, during the fMRI data analysis that we need to perform to investigate the possible connection between brain's region we need to ask to the patient to do a *do-nothing* activity. This could be very difficult even for the most trained patient. Due to this condition, it could not be used the GLM method to inspect the data that we get from this analysis. This is caused by a leak of Haemodynamic responses generated from the brain. For this reason, the analysis is based on other methodology:

- *Seed based Correlation*
- *ICA*
- *Hierarchical Clustering*
- *Self – Organization Map*

In the session where we need to investigate the resting state behaviour of the network, we have to ask to the subject to lie still. This is the condition where the resting state fMRI data come from. The thing that could be misleading is that even if it is called resting state the data are not of rest but reveal the so-called Default Mode Network. This is a little bit controversial such that some people of the field called this state as task free condition. The resting state fMRI enables assessment of intrinsic function activity. To get these results we need some graphical or numerical output that quantifies and let be interpretable the fMRI data acquired. To propose these results, we need to process the data by the usage of specific techniques. The methodology chosen is very important and rely on different characteristic such as the model type, the domain, and the hypothesis. Down below is reported a table that subdivide the principal techniques used to investigate the resting state fMRI data.

<i>Domain</i>	<i>Hypothesis – Driven</i>	<i>Model Type</i>
		<i>Data - Driven</i>
<i>Temporal</i>	Seed – based Correlation	Regional Homogeneity Principal Component Analysis Independent Component Analysis
<i>Frequency</i>	Coherence	Amplitude of low frequency fluctuation (ALFF) Fractional ALFF (fALFF)
<i>Time – Frequency</i>		Four – dimensional consistency of local neural activity (FOCA)

In the table above we can see the presence of different methodology and the choice depends also on if we want to perform a global or a local analysis. If we are interested in this last one, we should choose the:

- *Regional Homogeneity (ReHo)* is one analysis that could be performed in the temporal domain and the underlying assumption of using this is that the neighbour voxels are homogeneous to the one analysed and we should look to the concordance by calculating the Kendall's coefficient of concordance and we look to the interconnectivity within that region.
- *Amplitude of low frequency fluctuation (ALFF)*, this analysis could be performed in the frequency domain and look to the strength of the interconnectivity between two voxels in the same region. The amplitude of low frequency fluctuation methods looks to the low frequency's amplitude fluctuations recordable between two voxels of the same local region. The analogies that usually is done with this technique is to look at a map and the traffic amount along specific streets.
- *Fractional ALFF (fALFF)*, this analysis is the same as before but in this case the amplitude is normalized with the total amount of power in the frequency range. Even for this technique there is an analogy which is to look at the traffic density along a specific street normalizing the number of cars for people that live in that city.
- *Four – dimensional consistency of local neural activity (FOCA)*. This is an extension of the *ReHo* method in the frequency domain. In this case in fact, we can look even to the temporal and frequency interconnectivity between neighbour voxels.

If we are interested in performing a wider analysis instead, we can use the next methods that are subdivided into the two classes, the hypothesis driven methods and the data driven methods. In the first category we could find the next methods:

- *Seed based correlation*: It is the most basic and rudimentary way to inspect the data that we have obtained and measured. It is based on some initial hypothesis to take the right seeds where to perform the analysis. The correlation found is just temporal and the found results depends on just the initial seeds selected.
- *Coherence*: It is the analogous as the seed – based methods but inspected in the frequency domain.

For the data driven approaches instead we could find the next set of techniques:

- *Principal component Analysis (PCA)*: This one is a technique of data reduction methods by considering only the component that explain most of the variance of our data. Usually, the number of components considered is proportional to the amount of variance that we want to explain in our dataset. The drawback of this technique is that the component identified are not independent one each other.
- *Independent Component Analysis (ICA)*: This is the methodology that is used to get independent component between the different analysed. Nowadays researchers use both the ICA and PCA. This because the ICA need some data reduction analysis before performing the algorithm.

In the usual application of task involved analysis one of the best methods to use is the *General Linear Model (GLM)* that involves the next structures to get some final information to compute the different parametric maps.

$$X = G \beta + \epsilon$$

Where X represent the voxel time courses, G represent the transfer matrix containing the different haemodynamic curves, β the parameters vector and finally ϵ which represents the noise contribute. By visualizing and representing the β estimated all over the maps we can create different *statistical parametric maps (SPMs)*. The problem of resting state fMRI is that we no longer have the transfer matrix because there is no task activated in our data and thus, also the HDR should not be excepted.

In this case we could face with a so – called *cocktail party* problem where we have different distinct sources that are measured with the fMRI measurement system and we want to estimate the single individual activity. One way to solve this problem is by the usage of the *Independent Component Analysis*. This could be formalized with the mathematical problem:

$$\vec{x}(t) = A\vec{s}(t)$$

Where x represent the vector of measurement for each instant of time, s represent the vector of the behaviour of the sources for each instant of time and finally we have A which represent the weighting matrix. This problem could be solved by searching for the matrix W which maximizes the independence between different time courses in the next relation:

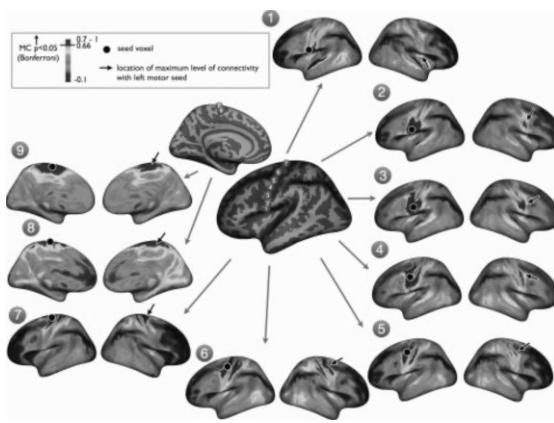
$$\vec{y}(t) = W\vec{x}(t)$$

This is very similar to the problem introduced with the GLM. The only difference is that in the ICA we do not have any transfer matrix G but we have the matrix A which represent the component course.

Default Mode Network: Investigation techniques - Seed based Correlation approach

The *Seed based Correlation approach* is one of the methodology most used nowadays to investigate and interpret the fMRI data coming from a resting state task. It is a very simple technique because after having identified the possible seeds where to perform the computation (usually these points are set up with the help of a doctor) we need to investigate the statistical correlation between the seeds chosen and all the voxel inside our scans. In this way we produce some maps representative of the statistical correlation between the seeds and the different ROIs. This approach is not at the voxel and even at the ROI level because the seeds is defined as a small portion of a specific ROI made up of 6 or 7 or 9 voxels.

The statistical correlation is an index that could be both positive and negative but generally the negative values are discarded because of not interest. The same analysis of correlation investigation could be performed from different initial seeds and this leads to the definition of multiple important maps.



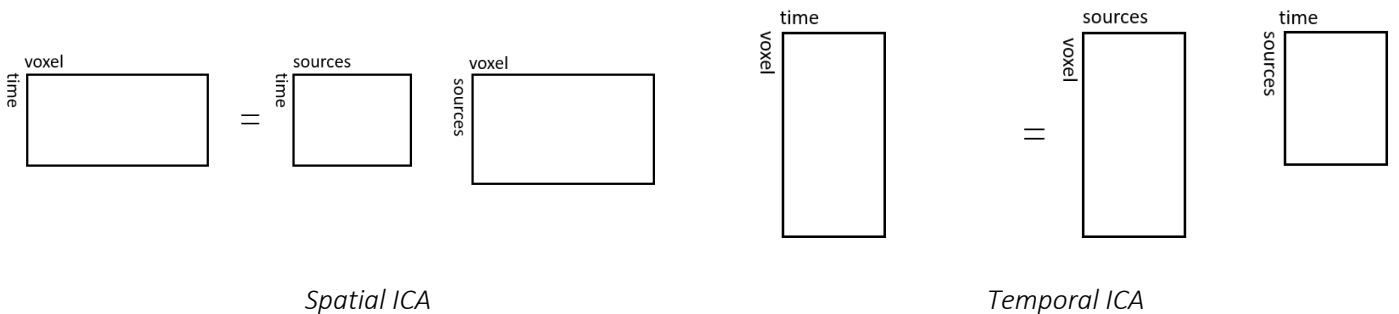
The seeds – based correlation analysis always allows to ask straightforward question and get an easily interpretable answer. Obviously, this analysis talks only the seeds that we have asked about. To solve this problem, we do not perform the analysis with seeds but are used the anatomical parcellation contained in specific atlas to determine the regions between which we must compute the statistical correlation.

Default Mode Network: Investigation techniques - ICA

Another way to investigate the Default Mode Network is by the usage of the ICA. This last one is the acronym for *Independent Component Analysis*. This last one is a method for recovering underlying signals from linear mixtures of those signals, draws upon higher – order signal statistics to determine a set of components that are maximally independent of each other. This method is used also to take care about the inter individual variability of the Default Mode Network that could be finally estimated. This method is created to achieve the maximal independence in space or time yields two varieties of ICA meaningful for functional MRI (fMRI) application:

- *Spatial ICA (SICA)*
- *Temporal ICA (TICA)*

These two techniques rely on two distinct situations represented in the scheme reported in the next page. We can see from this one the main differences between the two types of analysis. More in particular the spatial ICA is the transposons form of the temporal one.



In the two previous representation we can see the different ingredients that are necessary to perform correctly the Spatial and Temporal ICA. In the first case what we want to finally obtain is a matrix made up in the column of the different voxel and in the rows of the different time instant. Usually, thanks to the different measures this last matrix is known and could be found expressed with a different form that consist in change the voxels with specific parcels. The problem that we need to solve through the ICA algorithm is to find the most appropriate matrix that once reverse the expression let the combination of the different measures acquired more independent as possible. This is achieved with specific implemented algorithm as negentropy one. The Temporal ICA is based on the same concept, but the different composition of the problem are just transposed between column and rows. One important information that we need to highlight for the different measurements is that we need a lot of data. In general, are acquired just 6 to 7 minutes of task free state but this could change between research centres. In Washington centre in fact the acquisition time interval could also reach 30 minutes.

There are some important aspects because using the ICA. This last one is a process that create independent component spatially and in time and this leads to the possibility to identify eventual scanner, physiological artefacts and resting state networks. As for the other possible technique usable in fMRI data analysis we have some pro and some cons. These last one is that some components can be hard to interpret, and you may not get component that clearly relates to the brain bit you cared about. From the pros point of view instead the entire dataset is decomposed into “all” the different networks present. Another thing that could be seen as a pro of the technique is that the fMRI data does not need any pre – processing steps.

Default Mode Network: Investigation techniques – ICA: Open issues

In the ICA techniques there are some still open issues. These could be listed and summarized in the three down below categories.

- *Number of sources k*
- *Components of interest*
- *Spatial or temporal ICA*

The open issues are related most with the choice of the number of sources to use. This is not trivial because with fMRI data we are investigating the activity of thousands of millions of neurons. We can use some addition knowledge to determine the number of sources to estimate as instance through the PCA which is based on the quantity of variability to reconstruct. In general, we could incur in two distinct situations. Given the right number of components represented with \hat{n} we could decide to use a number n that could be:

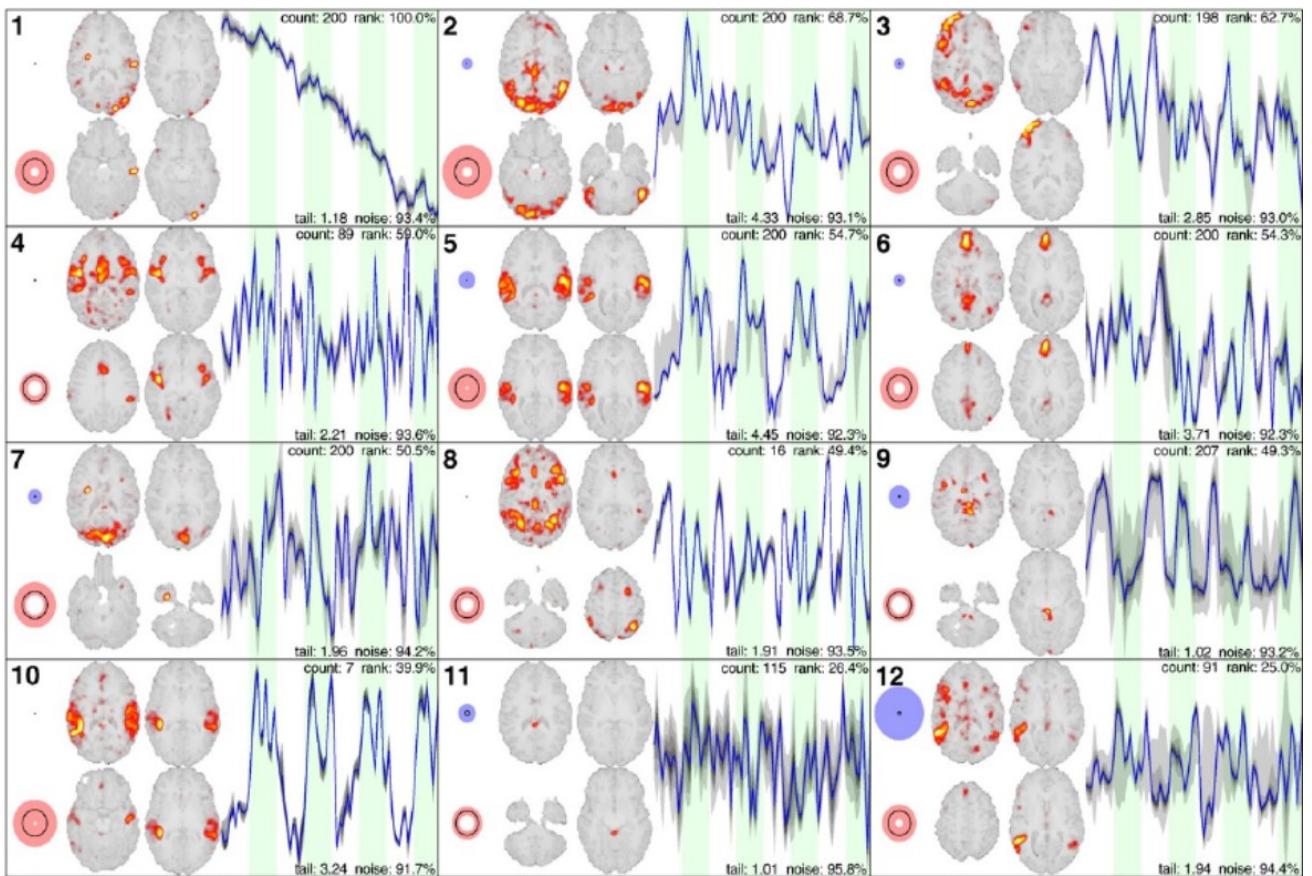
- $n > \hat{n}$: In this first case we are over splitting the components.

- $n < \hat{n}$: In this second case we are over clumping the components.

To solve this problem, we can perform as previously reported at first a PCA analysis and then we perform the ICA. An alternative is by selecting a high number of components to induce oversplitting and perform some clustering analysis post ICA to cluster together time courses that are not related with different brain's areas, but they are referring to the same.

Default Mode Network: Investigation techniques – ICA: Analysis example

In the image reported below we can see an example of fMRI data analysed with the ICA technique. The results obtained are reported already in some brain's maps. The acquisition protocol involves 10 different subjects with some auditory trials lasting 30 seconds and 30 seconds of rest. The ICA components extracted were 15 and, in the image, down below are represented only 12 of these 15 components.



From this image we can see that the number 5 and the number 10 are strongly related together. Maybe, due to the high number of component that are splitted into the different time courses is too high and we have induced some oversplitting phenomenon. This could be solved by just merging together those regions that seems spatially connected. Moreover, in the image 5 and 10 we can see how the signals is much more related with the task than the rest state. This gives us important results because we can see that the region of the brain activated is mainly localized over the auditory cortex.

Some components that are related with the blood flow are the number 11 and the number 4 (maybe also the number 12). In this case we can see that the behaviour of the fMRI signal is such stochastic and does not show a specific relation with a particular task.

Default Mode Network: Investigation techniques – ICA: Results interpretation

After having performed the ICA analysis, a bioengineer or with the supports of neurologist should interpret the results that comes out and establish if they are relevant or not for our purposes. To correctly perform the analysis, we need to inspect the different spatial maps and their relative time courses.

In literature there are several methods that are developed to order the output components. These could be summarized in three different relevant categories:

- *FIX (FMRIB's ICA – based X – noiseifier)*: FMRIB Software Library (FSL) uses a machine learning method to classify each ICA components as good or bad and regress bad ICA time courses.
- *Empirically*: As summarized recently by Allen and his research team, the empirical criteria to select RFNs from ICs are based on the expectations that RFNs should exhibit peak activations in gray matter, low spatial overlap with known vascular, ventricular, motion, and susceptibility artifacts, and dominated by low frequency fluctuations below 0.08 – 0.1 Hz.
- *Expert Neurologist*: This is widely the most important and reliable methods.

Brain connectivity

All these analyses are performed to prepare the data to be computed in to obtain some brain connectivity representation. This type of analysis investigates the principles of organization of the brain. More specifically how it segregates the different brain's area and how it integrates the different information. These two different concepts relays on the initially mentioned integration and segregation measures. The brain connectivity analysis starts when researchers found that specific area of the brain are involved in specific task. The first proofs of functional segregation were retrieved while inspecting the brain area activation during motor and visual task. In the history were developed two distinct functional concepts. These are nowadays recognized as:

- *Functional segregation*
- *Functional integration*

Brain Connectivity: Functional segregation

This concept relays on the theory that each part of the brain is specialised for different functions. This separation could be mainly seen in the cortical areas. In this case, the studies began with the observations that specific functions are localised in anatomic cortical regions. This leads to the born of the *Localizationism* theory which is nowadays well developed and known as the Functional Segregation. This type of analysis that can be conduct on the DMN try to search and identify the regions specialized for task. Usually, it is just a univariate analysis.

Brain Connectivity: Functional integration

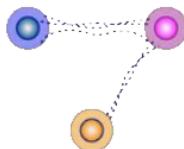
This concept relays on the possibility that exists networks of interactions among specialised areas. This point of view born with the *Globalism* theory which assumes that the brain works as a whole, extent of brain damage is more important than its location. This theory is based also know on the connectionism which represent the concept of interconnected networks. This analysis tries to explain how different regions of the brain in a neuronal system interact. This analysis could be conduct with univariate and multivariate analysis which are the functional and effective connectivity.

LESSON 13: BRAIN CONNECTIVITY

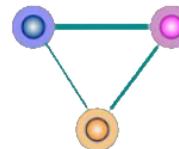
Brain connectivity: Introduction

In the last lesson we have talk about the different methodologies to analyse the *Default Mode Network* and find some relevant results to describe the interconnectivity of the brain. During the last part of the lecture, we have also begun to talk about brain connectivity and the two main branches of this sort of analysis. We would now introduce, in this lesson, some metrics and methods to find some information to study the brain connectivity map. In general, we could use methods based on the correlation and methods based on the causality. The main difference between these two is that the first just look on the possible statistical interaction between signal with a certain predetermine delay, while the second focus its attention on the causality's relation between different signals. Thus, with the first technique we could not determine the directionality of the interaction while with the second of course.

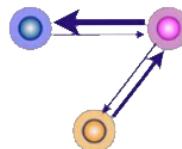
Another important information to know before going into deep of all the possible application about these types of studies is that we could recognize different classes of analysis based on the information that could be found. In general, the brain connectivity could be subdivided into three major categories:



Structural connectivity



Functional connectivity



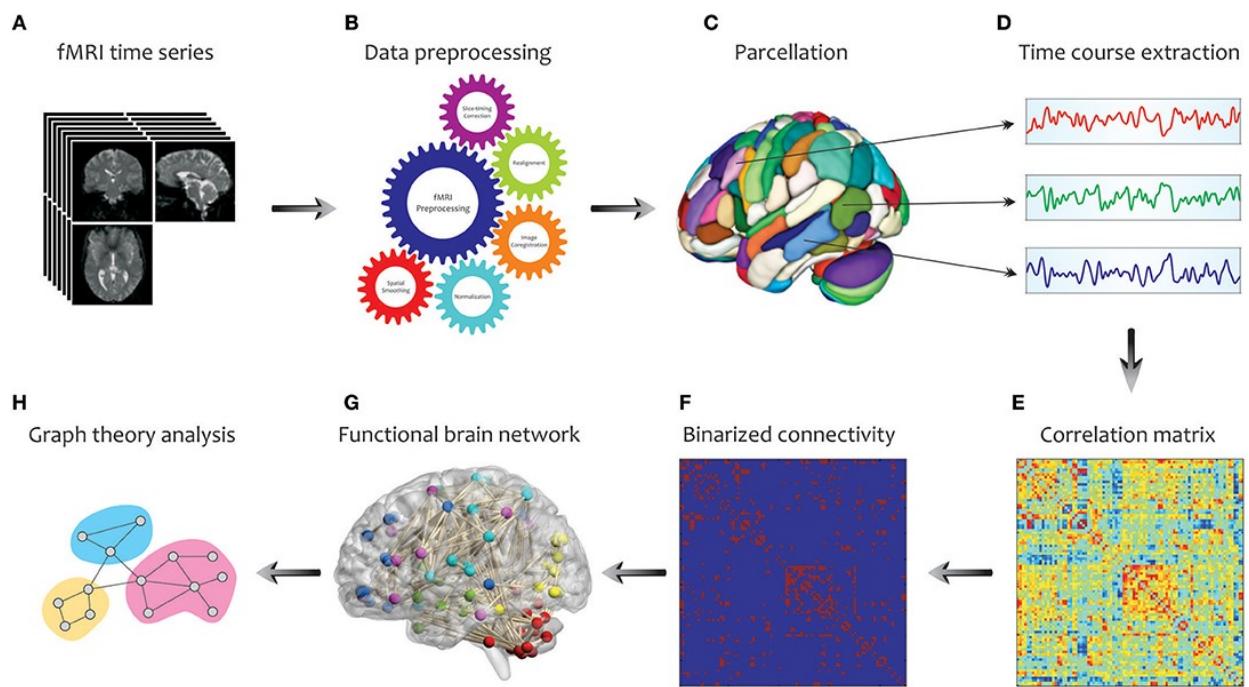
Effective connectivity

The differences between these three analyses involves not just the type of information that could be found but also the instrumentation and the algorithm that needed to be used to correctly process all the information. In fact, with the *structural connectivity* we highlight the presence of anatomical connections id est axons connecting different brain areas. This type of image could be obtained with the diffusion MRI. Instead, the *functional connectivity* map represents the statistical dependencies between the activity in different brain areas. It does not provide directionality or causality as previously said. Finally, the *effective connectivity* represents the causal directed influence that one neuronal system exerts over another. The instrumentation used in this last two techniques are respectively the fMRI, the EEG and maybe even the MEG.

Brain connectivity: Functional connectivity - Processing steps

The final aim of the brain connectivity analysis is to extract all the possible information form the signals measured and thanks to the usage of different mathematical instrumentation try to analyse the brain as a graph. To do so, we need to move from the signal measured to a graph. To reach this result we need to perform some elaboration on the data analysed, in particular we need to follow the following steps. First, we measured all the possible time series. In the different discussion that we will report in the following of the lesson we will refer to the data measured as fMRI data. This because this last measurement technique is well defined and widely used to perform connectivity analysis nowadays. After all the measurements processes, we need to apply some data pre – processing that could involve as instance some filtering, motion

corrections and registration operations. After having cleaned all the possible data, we need to get some information about the anatomy of the brain and subdivide the different brain regions into anatomical parcels that will be used to understand from which voxels needed to be extracted the time courses. After also the extraction of this last elements we could compute, by the usage of different techniques, the *functional connectivity matrix* also known as *adjacency matrix*. Now, the researcher, could choose if apply a binarization of the connectivity matrix or not and we could distinguish these two cases in the definition of a *weighted* or a *non-weighted undirected adjacency matrix*. Once this thing is computed we could obtain a representation of the brain as a network and apply some graph theory analysis to extract some information and indexes that could be helpful to enhance the differences between the healthy and non-healthy subjects involved in the study.



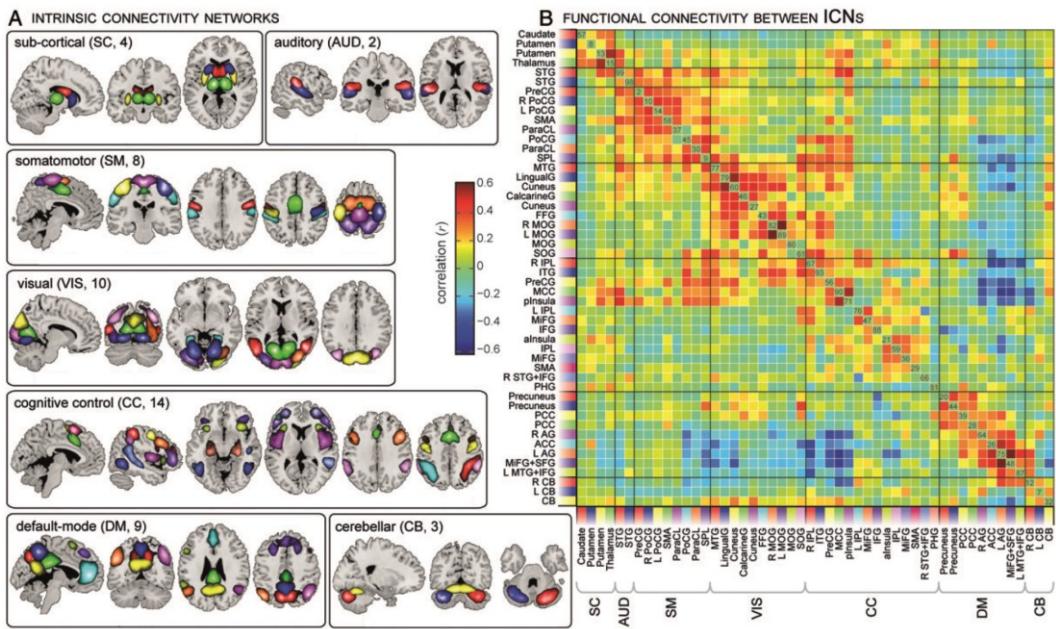
Before moving on, we should understand why could be useful or not calculate the binarization of the functional connectivity matrix. To understand this, we need to start from the definition of how the functional connectivity matrix is build up. The methods that could be used as instance are the ICA, the in-time correlation and delayed version of the correlation. Thus, the entries of the functional connectivity matrix represent the weight of the correlation between two given elements. This is such important because represent the strength of the statistical connection between signals. Thus, when we apply the binarization of the connectivity matrix we lost this information and sometimes instead, could be very relevant to be maintained.

Brain connectivity: Functional connectivity – ICA analysis

One possible type of analysis that could be performed on the fMRI data that we have measured is the functional connectivity analysis through the usage of the ICA. In this case the scheme of analysis is a little bit more different form the previous proposed. In fact, in this case we do not use any atlas to parcel the brain anatomically because the first step to apply is an independent component analysis (ICA). Thanks to this technique the signals are decomposed into *Intrinsic Connectivity Networks (ICN)*. This last one incorporate

two or more segregated submodules within and between hemispheres. The functional connectivity matrix can be calculated within the submodules of each ICN after having extracted from the ICN the signals relative to each brain region identified. This process is very useful because from the ICN we could build up the DMN and we can compare two different brain connectivity maps generated during a task related and a do – nothing condition by looking just only on the connection between the DMN. This is very useful to understand the changes on the brain functional connectivity between different regions of the brain.

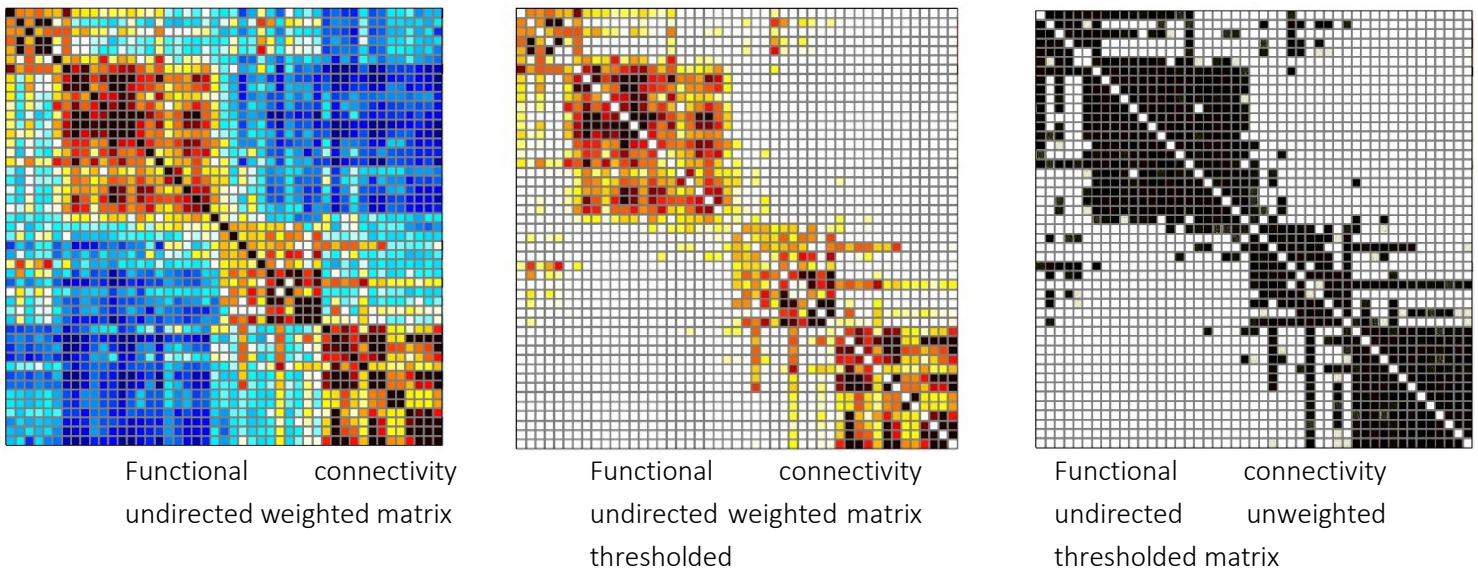
When all the calculation are done, we obtain as result the *functional connectivity matrix*. In the image reported below it can be seen the different correlation strength between ICNs. In the image is also possible to see the different ICN and how they are organized anatomically and spatially inside the brain.



In this image we could appreciate a lot of entries of the matrix that assume a high value and they are mainly localized over the main diagonal. The matrix in this case is symmetric because we are performing a correlation analysis which induce the loss of the directionality of the link. Moreover, to enhance other properties of the network previously identified is that there are a lot of non – zero element outside the square build on the main diagonal. These are sign of interconnections between segregated brain regions and index of functional integration.

Brain connectivity: Graph theory analysis

To extract some possible indexes relevant for the analysis we could keep just an amount of overall element, or we can set a threshold and directly binarize the matrix. In both cases, all the negative correlation entries of the matrix are discarded and set them to zero. Then, we could choose between maintain a certain amount of correlation inside the functional connectivity matrix or by maintaining just the entries above a certain threshold. In general, weak, and non – significant edges may represent spurious connections that tend to obscure the topology of strong and significant connections and as a result are often discarded by the usage of the previous techniques. The choice made at this point is crucial because will change the result that we obtain from the graph theory analysis. At this point, we could choose to maintain the strength information or not and thus to apply or not a binarization processes but remember that this step is not mandatory.



Brain connectivity: Graph theory - Nodes

Before moving on to presents some metrics that could be used to describe a graph, we need to introduce some basic concept. The first element of the graph analysis that made up a graph is the *node*. This is the basic concept of a graph and represents a certain brain area. In general, the node extraction and selection needed to be done carefully and could rely on different methodologies. In fact, we could extract nodes from atlas (after the registration over our fMRI data), activation studies, ICA and other techniques. Once identified all the nodes of our net we need to extract the time activity courses related with these. The time series could be processed over all the voxels related with the anatomical regions associated with that node to obtain just one representative signal of that region.

Brain connectivity: Graph theory - Edges

In the brain connectivity analysis, the connection between different nodes is called an *edge*. This is usually inside the functional connectivity matrix and is represented with a binary value or not. If it is not binary is also associated to the edge a weight representative of the connection strength between two different nodes. The measures that could be used to identify the weight inside a functional connectivity analysis is the *cross correlation* which represent in time the statistical dependencies between two given signals. Assumed that the two signals are represented with y_1 and y_2 we can represent the cross correlation between these two signals with the mathematical expression:

$$R_{12} = \frac{Cov(y_1, y_2)}{\sqrt{Var(y_1)Var(y_2)}}$$

An alternative to compute the edges weight is the *lagged cross correlation*, also known as *delayed cross correlation*. This evaluates the cross correlation between two signals at a certain degree of delay. The last measure that could be used to compute, in the functional analysis, the weight of edge is the *coherence*. This element is based on the frequency domain and could be expressed with the analytical relation reported below:

$$C_{12} = \frac{|G(y_1, y_2)|^2}{G(y_1)G(y_2)}$$

Where $G(y_1, y_2)$ represents the cross spectral density, while the $G(y_1)$ represents the spectral density.

Brain connectivity: Graph theory - Matrices

Once define the nodes and the edges, basic concept to build up a net, we need to introduce the two types of matrix that could be analysed.

The first one is known as the *incidence matrix* represented analytically with the symbol $I(G)$ of the graph G . It is a matrix of dimension $n \times m$ where the entries m_{ij} is 1 if the vertex v_i is adjacent to the edge e_j and 0 otherwise. Thus, this matrix represents the structure of the network and we could get information about the edges directly connected with specific nodes.

The other matrix is instead called the *adjacency matrix*. It is represented with the symbol $A(G)$ of the graph G and has dimension $n \times n$ whose entries a_{ij} represents the weight of the edge between the two vertices v_i and v_j . The adjacency matrix can have even binary entries:

$$a_{ij} = \begin{cases} 1 & \text{if the edge } j \rightarrow i \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

In other words, the adjacency matrix defined the topology of the graph by representing nodes as matrix rows and columns and representing edges as binary or weighted matrix entries. For simple graph, all the diagonal elements, representing the loop edges, are zeroes. Furthermore, as previously said, the adjacency matrix is symmetric for undirected graph.

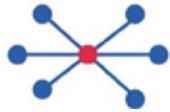
Brain connectivity: Graph theory – Measures

There are a lot of measure that could be found and useful to describe our identified graph. These are listed and highlighted down below.

- *Degree*. This index represents the number of the node connected to the one analysed with just one edges between them. In the image reported on the right the node in red has a degree of three and by considering the subnet we can say that the red node is the one with the highest degree.
- *Shortest path*. This measure represents the shortest path distance between two distinct nodes of the net considered. Usually, this metric is computed with the Dijkstra algorithm. In the image on the right is possible to see just the shortest path between the two nodes.
- *Clustering*. The clustering index represents how much a subnet is characterized by a segregation characteristic from the rest of the net. It is evaluated as the number of triangles around the node considered.



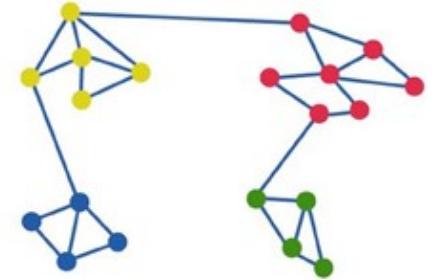
- *Betweenness centrality*. This measure can highlight nodes that are central inside a specific net. This is important because it identifies the zones where the information is majorly integrated inside a specific subnet or inside the overall net.



- *Hub*. This is majorly a properties of certain node of the node of the net. In this case the hub's node represent specific nodes that if extracted from the net creates two specific well segregated and not integrated subnet. This could be seen in the image reported on the right side.



- *Modularity*. Another important metric is the modularity one. In this case the net is recognize and subdivide into different clusters that works independently from the other but have at a higher level all these separated clusters integrate the elaborated information.



In general, these measures could be subdivided into three major categories which are respectively:

Measures of centrality

These measures represent and highlight the centrality of a given node respect to the entire net or a given subnet. The most representative measures are the degree of a node, the betweenness centrality of that node and finally the participation coefficient.

Measures of segregation

These measures represent the level of segregation between different subnets. This is very important because can let us understand the performances in parallel processing of our brain. One typical measure of segregation is the clustering coefficient of the node.

Measures of integration

On the other hand, the integration measures represent the ability of integrate between results coming from different subnets. In general, the most representative integration measure is the characteristic path length between two nodes and the global efficiency. This last one is represented as the average inverse shortest path length.

Nowadays in the literature there are some packages software that integrate all the possible measures of graph analysis. One of the most used is the *Brain Connectivity Toolbox* developed by Sporns, Rubinov and their research teammate.

Brain connectivity: Graph theory – Direct and undirect connections

A very crucial thing to always keep in mind is that there always exist approaches that produces only direct connections and other approaches that estimate during the different calculations undirect connections.

Sometimes could be useful try to find both but in the most of cases we are just interested in the direct ones. This because, given three nodes that we will call A, B and C, assume that with the effective connectivity we found that A is connected with B, B with C but there is no relation and connection between A and C. By applying some direct analysis, we could find the connection between A and B and B and C and this reflects the real situation. Instead, if we perform some undirect analysis we could even estimate the relation between A and C even if there is not. Hence, we understand that by using undirected methods we are estimating even undirected connections. This is sometimes undesired and if we look on the results that give to us these methods we could not recognize as the previous situation the undirect connections because we need some additional information that discriminate between the two situations.

In general, the functional connectivity analysis relies on the correlation or the coherence and these set of methods take part in the group of undirect analysis. Thus, with the functional connectivity we can estimate even the undirect connections. A way to solve this problem is to use the partial correlation instead of using the correlation itself. This method gives us quite more better results and measures the association between two variables after controlling for or adjusting for, the effects of one or more additional variables. The analytical definition is reported below:

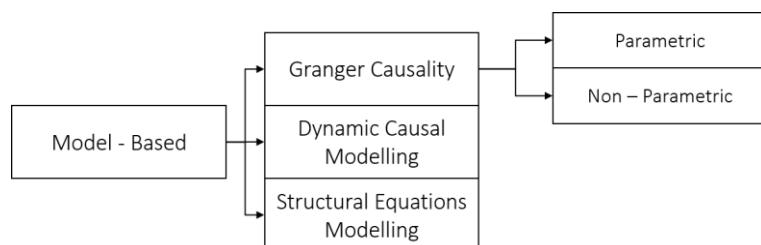
$$r_{xy \cdot z} = \frac{r_{xy} - (r_{xy})(r_{yz})}{\sqrt{1 - r_{xz}^2} \sqrt{1 - r_{yz}^2}}$$

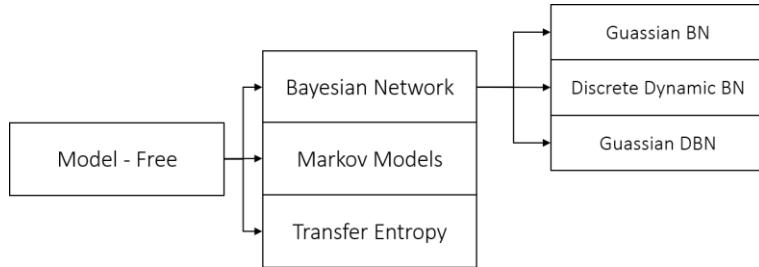
In this case we need to solve the linear regression problem for the signals x and y , we compute the residuals and then we calculate the results of the fraction. There are a lot of differences between the final correlation matrix obtained with the partial correlation instead of using the usual simple correlation. This induces to lower the sparsity of the matrix and solve the problem about the undirect correlation between regions. Nowadays aware of this problem, researcher still use correlation because it is important to induce sparsity in the data.

Brain connectivity: Effective connectivity

Another type of analysis that can be done on the data is to also extract information about the direction of the connections. In this case, the analysis gives not more a functional connectivity matrix but generate an *effective connectivity matrix*. In this last case, is measured the influence that one neuronal system exerts over another. It could be based on regression analysis, could be model dependent and hypothesis driven. A crucial property is that we can do causal analysis. This means that we can inspect the causalities relation between two given signals, and this is what makes the analysis creating a directed matrix. Thus, for this last motivation the matrix that represent the graph is not more symmetric.

In literature there a lot of methodology that could be used to correctly compute an effective connectivity matrix. In the scheme reported below are listed all the possible techniques.





Brain Connectivity: Effective connectivity - Granger causality

One of the methodologies that we will see during this course and also in the laboratory section is the *Granger Causality*. This method was formalized from Granger in 1969 with the concept of causality. During these years the PET was at the beginning. Granger formalized the concept of causality introduced by Wiener giving a mathematical formulation and providing a powerful tool analysis of the relationships between stationary time series. Weiner have defined the stationary causality between two simultaneously measured signals as:

If the prediction of the first signal improves using the past information from the second signal, unless the usage of the past information of the second signal, thus we could say that the second signal causes the first.

Starting from this idea of causalities and from the fact that the causes precede the effects, Granger created the stability concept in mathematical terms on an autoregressive model of temporal series. This concept is applied the most for the EEG signals to understand which electrical trace causes another. Even for fMRI data we could apply this technique but it is less use because in the research filed is prefer the method proposed by Karl Friston that we will see in the next lecture.

The calculation of the Granger Causalities is done by using a multivariate autoregressive model. This model is a discrete time, linear and time invariant model and it is described by difference equations. The MVAR model with N variable is expressed as:

$$y(n) = - \sum_{k=1}^p A(k)y(n-k) + e(n)$$

Where $y(n) = [y_1(n), y_2(n), \dots, y_N(n)]^T$ is the data vector of dimension N containing the $n - samples$ of the N time series, p is the model order, $A(k)$ with $k = 1, 2, \dots, p$ are the $N \times N$ matrices containing model coefficients, $e(n) = [e_1(n), e_2(n), \dots, e_N(n)]^T$ is the vector containing the $n - samples$ of the prediction errors. Id est it is a multivariate white noise process with diagonal covariance matrix $\Sigma_e = diag[\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2]$.

As an example, by considering three distinct regions. We want to measure the causalities from y_2 to y_1 given y_3 the complete MVAR model is the following.

$$\begin{aligned} y_1(n) &= - \sum_{k=1}^p a_{1,1}(k)y_1(n-k) - \sum_{k=1}^p a_{1,2}y_2(n-k) - \sum_{k=1}^p a_{1,3}y_3(n-k) + e_1(n) \\ y_2(n) &= - \sum_{k=1}^p a_{2,1}(k)y_1(n-k) - \sum_{k=1}^p a_{2,2}y_2(n-k) - \sum_{k=1}^p a_{2,3}(k)y_3(n-k) + e_2(n) \end{aligned}$$

$$y_3(n) = - \sum_{k=1}^p a_{3,1}(k)y_1(n-k) - \sum_{k=1}^p a_{3,2}(k)y_2(n-k) - \sum_{k=1}^p a_{3,3}y_3(n-k) + e_3(n)$$

These are the differential equations that describes the behaviour of the model. Model coefficients $a_{i,j}$ are estimated using the multichannel Yule Walker method. The model order p is estimated according to some criterion as instance *AIC* or other. As result of the model identification, we obtain an estimate of the error prediction:

$$\sigma_1^2 = \text{var}(e_1) \quad \sigma_2^2 = \text{var}(e_2) \quad \sigma_3^2 = \text{var}(e_3)$$

The variance represents the remaining variance that it not explained from the model. After having identified the multivariate autoregressive model, we can just discard the contribution of the second signal, and we estimate the bidimensional multivariate model. From this we could derive the residual variance related and associated to the model. Now, we have estimated the variance inside the complete and partial autoregressive model. From a theoretical point of view we could say that the signal eliminated from the system is causing one of the signals inspected if the total variance in the eliminated version is higher than the one inside the complete version of the autoregressive model.

To define that the granger causality that the signal y_2 causes the signal y_1 ($y_2 \rightarrow y_1$) we could use:

$$GC_{(y_2 \rightarrow y_1 | y_3)} = \ln \frac{\rho_1^2}{\sigma_1^2}$$

If the value of GC is greater than one, thus the variance at the denominator (variance of the complete autoregressive model) is lower of the variance at the numerator (variance of the eliminated version of autoregressive model). In this case we could say that $y_2 \rightarrow y_1$. If this value is instead zero, this means that the variance ratio is equal to one and thus the signal y_2 does not cause y_1 . To test if a given value is positive significantly, we can use the *F-test* of the variance. Hence:

$$F_{y_i \rightarrow y_j | y} = \frac{\frac{RSS_{\text{restricted}} - RSS_{\text{unrestricted}}}{p}}{\frac{RSS_{\text{unrestricted}}}{(L - 2p - 1)}}$$

Where L represents the number of measured data points, while RSS represents the residuals sum of squares of restricted and unrestricted models and finally p represents the model order.

Brain Connectivity: Effective connectivity - Granger causality: Pro and Cons

The pros in the usage of Granger Causality is that can be applied directly to any given timeseries. The model selection (significant connections) is completely data – driven and do not need hypothesis on model structure. As cons we are not considering the haemodynamic response variability in different brain regions. In general, for fMRI data we should not use the Granger Causality.

```

1 %
2 % Title: Laboratory 5
3 %
4 % Note:
5 % fMRI Analysis and generation of the functional connectivity map
6 %
7 % Author: Matteo Martin
8
9 % THEORY -----
10 %
11 % fMRI
12 % The goal of this laboratory is to perform the pre processing of resting
13 % state functional MRI data and arrive to a functional connectivity matrix.
14 % Once all the timeSeries are build up we can create the matrix of
15 % functional connectivity through the usage of the correlation function.
16 % Higher is the value in a specific entry higher is correlation between two
17 % signals. One more important thing is that after the computation of the
18 % correlation index we can compute the Fisher z transform by applying the
19 % atanh function to each element of the correlation matrix created.
20 %
21 % PRE PROCESSING STEPS
22 % The different steps that are needed to correctly perform the entire
23 % analysis are the next ones:
24 %
25 % - MASK: Generation of the different masks that are needed to activate
26 %           specific brain regions. The mask to be produced are:
27 %
28 %           FMask : Which hold information about the voxels where it is
29 %           detected a certain fMRI activity, the voxels refer to a
30 %           gray matter element and there is the association of that
31 %           voxel with an anatomical region contained in the Atlas.
32 %           WMMask : It represents the mask for the detection of which voxels
33 %           are associated with white matter.
34 %           CSFMask: It represent the mask used for the detection of the
35 %           voxels associated with the cerebral spinal fluid.
36 %
37 %           After this mask generation we extract the timeseries curves
38 %           associated to each non zero element inside the different masks.
39 %
40 % - FILTER: Filtering of the data by the usage of some parameters that are
41 %           associated with the movement correction. Application of a
42 %           denoising phase associated with non neural noise and temporal
43 %           filtering with an apposite given funciton. After the filtering
44 %           in terms of noise we censor the slices acquired in certain
45 %           instant of time in which the displacement get bigger than a
46 %           certain given threshold (maximum value of displacement
47 %           accessible).
48 %
49 % - ROI: Extraction of the different curves associated with specific ROIs
50 %           that should not be neglected in our analysis.
51 %
52 % - FUNCTIONAL CONNECTIVITY: Computation of the functional connectivity
53 %           matrix by the evaluation of the pearson
54 %           correlation coefficient by normalizing it with

```

```

55 %                               the Fisher z - score. By the usage of the corr
56 %                               function we extract also the different p -
57 %                               values assocaited with each element of
58 %                               correlation computed.
59
60 %% CCC
61 % Clear all, close all, clear the comand window
62
63 close all; clear all; clc
64
65 %% LOAD - DATA
66 % Load of all the possible data inside the workspace of Matlab with the
67 % apposite function of Nifty toolbox.
68
69 ATLAS = load_untouch_nii('Hammers_2_T1_2mm_int.nii');
70 ATLAS = double(ATLAS.img);
71
72 FMRI = load_untouch_nii('FMRI_2_T1_2mm.nii');
73 FMRI = double(FMRI.img);
74
75 MASKF = load_untouch_nii('mask_sumepi.nii');
76 MASKF = double(MASKF.img);
77
78 CSF = load_untouch_nii('csf_mask_0.8.nii');
79 CSF = double(CSF.img);
80
81 GM = load_untouch_nii('gm_mask_0.4.nii');
82 GM = double(GM.img);
83
84 WM = load_untouch_nii('wm_mask_0.8.nii');
85 WM = double(WM.img);
86
87 [nROW, nCOL, nSLICE, nTIME] = size(FMRI);
88
89 %% MASK - GENERATION
90 % Generation of three diffent mask.
91 %
92 % - FinalMask: Mask the Hammers atlas with two masks: the GrayMatter and
93 %                 the sumEPI mask. The first is used to get just the
94 %                 information from the anatomical point in the atlas that are
95 %                 belonging to the gray matter and the second beacuse only a
96 %                 region of the fMRI image is active.
97 %
98 % - WMMask : Mask of the White Matter mask with the sumEPI to active
99 %               only the portion of the white matter that contain
100 %               informative information for the fMRI
101 %
102 % - CSFMask : Mask of the CSF with the sumEPI to active only the portion
103 %               of the CSF that it is active in the fMRI analysis.
104 %
105 % After have done all these possible pre processing phase we need to mask
106 % the fMRI with the mask created.
107
108 GMMask = GM.*MASKF;

```

```

109 FMask      = (GMMask.*ATLAS);
110 WMMask     = WM.*MASKF;
111 CSFMask    = CSF.*MASKF;
112
113 F2DMean   = sum((sum(FMRI,4)).*(FMask>0),3);
114 GM2DMean   = sum((sum(FMRI,4)).*GMMask,3);
115 WM2DMean   = sum((sum(FMRI,4)).*WMMask,3);
116 CSF2DMean = sum((sum(FMRI,4)).*CSFMask,3);
117
118 F2DActivity = zeros(nTIME, length(find(FMask)));
119 GM2DActivity = zeros(nTIME, length(find(GMMask)));
120 WM2DActivity = zeros(nTIME, length(find(WMMask)));
121 CSF2DActivity = zeros(nTIME, length(find(CSFMask)));
122
123 for idT = 1:1:nTIME
124     temp           = FMRI (:,:,:,:,idT);
125     F2DActivity(idT,:) = temp(logical(FMask));
126     GM2DActivity(idT,:) = temp(logical(GMMask));
127     WM2DActivity(idT,:) = temp(logical(WMMask));
128     CSF2DActivity(idT,:)= temp(logical(CSFMask));
129 end
130
131 clear temp
132
133 %% MASK - VISUALIZATION
134 % Visualization of the mask created at the step before
135
136 M = max([max(F2DMean,[],'all'), ...
137           max(WM2DMean,[],'all'), ...
138           max(CSF2DMean, [], 'all')]);
139
140 m = min([min(F2DMean,[],'all'), ...
141           min(WM2DMean,[],'all'), ...
142           min(CSF2DMean, [], 'all')]);
143
144 figure, h(1) = subplot(1,3,1); imagesc(F2DMean),
145         title('FMRI - Gary Matter Activity'), colormap 'jet';
146 h(2) = subplot(1,3,2); imagesc(WM2DMean'),
147         title('FMRI - White Matter Activity'), colormap 'jet';
148 h(3) = subplot(1,3,3); imagesc(CSF2DMean'),
149         title('FMRI - CSF Matter Activity'), colormap 'jet';
150 set(h,'Clim',[m M]), set(h, 'YDir', 'normal')
151
152 clear M m;
153
154 %% MASK - VISUALIZATION
155 % Visualization of the mask created by the usgae of the implay function
156
157 implay(WMMask)
158 implay(CSFMask)
159 implay(FMMask)
160
161 %% DYNAMICS - VISUALIZATON
162 % Visualization of the dynamics extracted in the steps before to get some

```

```

163 % information about the behaviour in each voxel of the FMRI data acquired.
164
165
166 TR = 2.6;                      % Repetition time
167 t = TR : TR : nTIME*TR;        % Time vector
168
169 figure, subplot(1,3,1), plot(t, mean(F2DActivity,2)),
170     title('GRAY MATTER - Dynamics'), grid on, grid minor, box on,
171     xlabel('Time [s]'), ylabel('Amplitude []')
172 subplot(1,3,2), plot(t, mean(WM2DActivity,2)),
173     title('WHITE MATTER - Dynamics'), grid on, grid minor, box on,
174     xlabel('Time [s]'), ylabel('Amplitude []')
175 subplot(1,3,3), plot(t, mean(CSF2DActivity,2)),
176     title('CSF - Dynamics'), grid on, grid minor, box on,
177     xlabel('Time [s]'), ylabel('Amplitude []')
178
179 %% LOAD - DATA
180 % Loading of the data inside the workspace. The data to be loaded is the
181 % MOCOparams.mat which represent the behaviour of the motion correction
182 % parameter used and created.
183
184 load('MOCOparams.mat'), MOCOP = newMOCOparams; clear newMOCOparams;
185
186 %% REGRESSOR - MATRIX
187 % Generation of the motion matrix that will be used in the next step to
188 % perform the noise regression from the identified signals. Visualization
189 % of the z score regressor matrix.
190
191 MOCOP(:,7) = mean(WM2DActivity,2);    % White matter mean activity
192 MOCOP(:,8) = mean(CSF2DActivity,2);    % Cerebral Spinal Fluid mean activity
193 ZMOCOP      = zscore(MOCOP);          % Z - Score
194
195 figure, imagesc(ZMOCOP), title('REGRESSOR MATRIX'), colormap gray; colorbar
196 figure, subplot(2,1,1), plot(MOCOP(:,1:3)),
197     title('TRANSLATION [mm]'), grid on, grid minor, box on,
198     xlabel('Volumes []'), ylabel('Translations [mm]'), xlim([0, nTIME])
199 subplot(2,1,2), plot(MOCOP(:,4:6)),
200     title('ROTARION [°]'), grid on, grid minor, box on,
201     xlabel('Volumes []'), ylabel('Rotation [°]'), xlim([0, nTIME])
202
203 %% NOISE - REGRESSION
204 % Regression of the noise using the regression parameter estimated during
205 % the motion correction phase.
206
207 PARAM      = (ZMOCOP'*ZMOCOP) \ ZMOCOP'*F2DActivity;    % Regression
208 CF2DActivity = F2DActivity - ZMOCOP * PARAM;            % Cleaning NonNeuro
209 CF2DActivity = hp_filter(CF2DActivity, nTIME, TR, 1/128); % Cleaning Slow
210
211 % In this way we have clean all the entire activity with the motion
212 % parameter and the white matter and cerebral spinal fluid influences. In
213 % this way we have subtract from the signal the non neural component.
214 % After that we have cleaned the signal from the slow components which
215 % could be due to different noise's sources such as breath and cardiac
216 % oscillations.

```

```

217
218 %% DYNAMICS - VISUALIZATION
219 % Visualization and comparison of the dynamics extracted with the initial
220 % method and the one elaborated and cleaned from different types of noise.
221
222 figure, hold on, plot(t, mean(F2DActivity,2)),
223             plot(t, mean(CF2DActivity,2)), hold off,
224             title('GRAY MATTER - Dynamics Comparison'),
225             xlabel('Time [s]'), ylabel('Amplitude []'),
226             grid on, grid minor, box on, xlim([0 nTIME])
227             legend('Non-Filter','Filter')
228
229 %% LOAD - DATA
230 % Loading of the data inside the workspace of Matlab to handle it and use
231 % them. We rename the data loaded to get a better uniformity of notations.
232
233 load('FDparams.mat'), TH = 0.35;
234
235 %% DATA - VISUALIZATION
236 % Visualization of the behaviour of the FD vector loaded inside the
237 % workspace of Matlab used to check the maximal displacement.
238
239 figure, plot(FD), hold on,
240             yline(0,'k--','0','LineWidth',2)
241             yline(TH,'r--','TH','LineWidth',2)
242             plot(find(FD>TH),FD(FD>TH), 'or','LineWidth',2), hold off
243             title('DISPLACEMENT [mm]'),
244             xlabel('Volumes []'), ylabel('Displacement [mm]')
245             xlim([0 nTIME]), grid on, grid minor, box on,
246
247 %% VOLUME - CENSORING
248 % Application of the volume censoring algorithm. In this way we need to
249 % inspect the behaviour of the FD vector loaded at the previous step. In
250 % this way we can determine, if a certain element present a value greater
251 % than 0.35 mm displacement, the volxes that should not be considered.
252
253 idT = 1; EXTRACT = [];
254
255 while idT <= nTIME
256
257     if FD(idT) < TH
258         EXTRACT = [EXTRACT,; idT];
259         idT = idT + 1;
260     else
261         if idT == nTIME
262             EXTRACT(find(EXTRACT == (idT-1),1)) = [];
263         elseif not(idT == 1)
264             EXTRACT(find(EXTRACT == (idT-1),1)) = [];
265         end
266         idT = idT + 3;
267     end
268
269 end
270

```

```

271 VCF2DActivity = CF2DActivity(EXTRACT,:);
272 tExtract      = t(EXTRACT);
273
274 % In this situation we extract only the volumes that shows a FD value
275 % associated with a maximum displacement of 0.35 mm. In this situation,
276 % when a certain volume present this artifact also the previous and the two
277 % after are cancelled out from the indexes vector.
278
279 %% ROI - ACTIVITY
280 % Extraction of the ROI time activity for each atlas brain region masked.
281 % Except for the brain regions inside the set: amygdala, cerebellum,
282 % brainstem, corpus callosum, substantia nigra and ventricles.
283
284 dROI = [0 3 4 17 18 19 44 45 46 47 48 49 74 75];
285 aROI = unique(ATLAS);
286 iROI = setdiff(aROI, dROI);
287
288 nROI      = length(iROI);
289 aINDEX     = find(FMask);
290 ROIActivity = zeros(size(F2DActivity,1),nROI);
291 CROIActivity = zeros(size(VCF2DActivity,1),nROI);
292
293 for idR = 1:nROI
294     iINDEX = find(FMask == iROI(idR));
295     sINDEX = zeros(1,length(aINDEX));
296
297     for idI = 1:length(iINDEX)
298         sINDEX(aINDEX == iINDEX(idI)) = 1;
299     end
300
301     ROIActivity(:,idR) = mean(F2DActivity(:,logical(sINDEX)),2);
302     CROIActivity(:,idR) = mean(VCF2DActivity(:,logical(sINDEX)),2);
303 end
304
305 %% HIPPOCAMPUS - VISUALIZATION
306 % Visualization of the activity of Hippocampus regions before and after the
307 % denoising step. This to check if the analysis are performed correctly or
308 % not.
309
310 figure, hold on,
311
312     plot(t, ROIActivity(:,iROI == 1) - ...
313             mean(ROIActivity(:,iROI == 1)))
314     plot(tExtract, CROIActivity(:,iROI == 1) - ...
315             mean(CROIActivity(:,iROI == 1))),
316
317     hold off, grid on, grid minor, box on,
318     title('HIPPOCAMPUS - Dynamics Comparsion'),
319     xlabel('Time [s]'), ylabel('Amplitude []'),
320     xlim([0 t(end)])
321
322 %% FUNCTIONAL CONNECTIVITY
323 % Computation of the funcitonal connectivity matrix by the usage of the
324 % Pearson's correlation and visualization of the different identified

```

```
325 % matrix visually.
326
327 [R, P] = corr(CROIActivity, CROIActivity);
328 Z      = atanh(R);
329 M      = max(triu(Z,1),[],'all');
330 m      = min(triu(Z,1),[],'all');
331
332 figure, subplot(1,2,1), imagesc(Z), colormap 'jet'; colorbar
333             title('FUNCTIONAL CONNECTIVITY'), caxis([m M])
334 subplot(1,2,2), imagesc(P), colormap 'jet'; colorbar
335             title('P - VALUES')
336
337
```

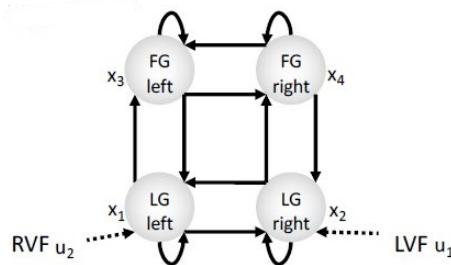
LESSON 14: DYNAMIC CAUSAL MODELLING

In the today's lecture we will going to see the methodology based on the techniques proposed by Karl Friston in 2004 and then revised and extended by Stephan and its research group. In general for the effective connectivity analysis we could cite a lot of different methodologies that could be used to get these types of information. Some examples are the *regression models* proposed by Friston in 1997, the *structural equation models (SEM)* proposed by McIntosh in 1991 and the *Volterra Kernels* first introduced in 2000.

Brain connectivity: Effective Connectivity - Dynamic Causal Modelling

These methods as previously said was first introduced from Friston and its research group in 2003. This instead of the granger causality method is physiologically driven instead of model driven. This is the reason why the *Dynamic Causal Modelling (DCM)* it is widely used nowadays to perform elaborations. The first publication of the previous mentioned technique comes from the 2003 and the reference paper was cited more than 300 times inside other different literature. In particular, in SPM the software used inside the labs is embedded some functions that could be useful to implement this type of analysis.

To introduce these methods we consider the next example of linear model of interacting visual regions. Down below is reported the full scheme and we could see the presence of four different compartment where are analysed also some specific autocorrelation of effect inside the same compartment. If we need to describe this model with the GLM technique we have a strong simplification of what it is happening in the system due to the strong simplification considered for the HDR. But if we consider the situation with the DCM approach thanks to the mixture between the model used and some physiological knowledge we could reach better and more physiologically relevant results. This is the reason why, as in the previous paragraph said, is much more used the DCM instead of the GLM.



Let try to apply and explain where is based the *Dynamical Causal Modelling* by considering the previous chart. In this case we have two inputs represented from the dashed arrows connected with the lower compartments. In this case, the input has name respectively *LVF* and *RVF* which are the acronym of visual input in the *Left* and *Right Visual Field*. Then we have four distinct compartments called *LG* and *FG* which means respectively *Lingual Gyrus* and *Fusiform Gyrus*. Thanks to the physiology and the knowledge in this field we could express and justify the different connections between all the compartments. Moreover, thanks to the physiological knowledge all the autocorrelation on the same compartment could be hold and are physiologically relevant. From this representation we could understand how the two hemispheres interact between each other when a certain stimulus is proposed to the patient. To analyse this model we need to perform and enhance two specific assumptions. These are:

- The first assumption is that all the region of interest that we could consider could be represented with a compartment where the overall state is represented with the vector x .
- The second assumption instead is much stronger than the previous and in this we are assuming that everything is linear.

Thanks to the previous last assumption we could translate the graphical visualization of the system inside a mathematical system of equation. In this case, by considering that the model is represented from a compartmental model all the equations are differential ones. Thus the previous model could be summarized as:

$$\begin{aligned}\dot{x}_1 &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + c_{12}u_2 \\ \dot{x}_2 &= a_{21}x_1 + a_{22}x_2 + a_{24}x_4 + c_{21}u_1 \\ \dot{x}_3 &= a_{31}x_1 + a_{33}x_3 + a_{34}x_4 \\ \dot{x}_4 &= a_{42}x_2 + a_{43}x_3 + a_{44}x_4\end{aligned}$$

Where a_{ij} represent the strength of the correlation. From the previous mathematical description there is a little bit of difference inside the compartment 3 and the compartment 4. In these two last ones in fact, there is any direct input inside these two compartments. Moreover, the a_{ij} coefficient are assumed to enter inside the compartment, and they are not split in the two components where one enters in the compartment j and the other in the compartment i . Thus, in the effective connectivity analysis in this case there is not a symmetrical matrix. The model could be represented also with a matrix vector notation where could be seen the non – symmetrical matrix:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & 0 & a_{24} \\ a_{31} & 0 & a_{33} & a_{34} \\ 0 & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 & c_{12} \\ c_{21} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Where the first column vector is also known as the state changes of the model. The matrix is known as the effective connectivity matrix between the four compartments analysed, the second vector represents the system state, the second matrix the input parameters and finally the vectors that represents the external inputs of the system. With a more compressed form we could represented the previous relation as:

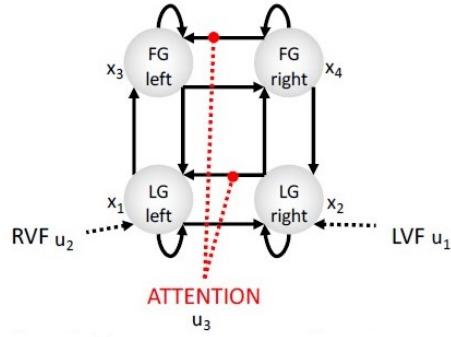
$$\dot{x} = Ax + Cu$$

Where the things that needed to be estimated are the matrix A and the input parameters C . Thus the complete parameters vector is codified as the next set of elements.

$$\theta = \{A, C\}$$

All the terms reported inside the previous set of elements θ are unknown and needed to be estimated expect for the value of the input parameters. In this case we know that the interaction between the right and the left hemisphere could be modulated some other input. To improve the physiology and the explanations of the interconnections between different brain region in the model is inserted the *attention* input. This is physiological explainable and meaningful because the interaction between the left and the right visual cortex

interact higher only when the attention is involved in the elaboration. Thus the interaction is called has a control from an external source. In this case is modified the scheme of the bilinear model considered.



In this case, because we need to consider even the input u_3 we have a different representation of the model in the vector matrix form. This will follow the next analytical expression:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \left(\begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & 0 & a_{24} \\ a_{31} & 0 & a_{33} & a_{34} \\ 0 & a_{42} & a_{43} & a_{44} \end{bmatrix} + u_3 \begin{bmatrix} 0 & b_{12}^{(3)} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_{34}^{(3)} \\ 0 & 0 & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 & c_{12} \\ c_{21} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

The system in a close and general form could be so described by the next expression:

$$\dot{x} = \left(A + \sum_{j=1}^m u_j B^{(j)} \right) x + Cu$$

We are modelling the previous system with some linear relation when we know that there is no such differences. Consequently, it could be used this type of simplification by justifying it with the Taylor series expansion around a stable state $x = 0$ and input $u = 0$ until the second order of a non – linear dynamics. From an analytical point of view we could write:

$$\frac{dx}{dt} = f(x, u) \approx f(x_0, 0) + \frac{\partial f}{\partial x} x + \frac{\partial f}{\partial u} u + \frac{\partial^2 f}{\partial x \partial u} ux + \dots$$

All the derivative expressed in the previous differential equation are calculated in the coordinate of the equilibrium point which are represented in the next few equations:

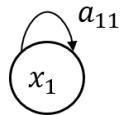
$$A = \frac{\partial f}{\partial x} \Big|_{u=0} \quad C = \frac{\partial f}{\partial u} \Big|_{x=0} \quad B = \frac{\partial^2 f}{\partial x \partial u} \Big|_{(x,u)=(0,0)}$$

From these equations we could derive the Bilinear state equation of the system which is a generalization of what previously seen in the example:

$$\frac{dx}{dt} = \left(A + \sum_{i=1}^m u_i B^{(i)} \right) x + Cu$$

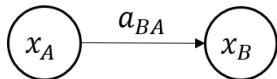
Brain Connectivity: Effective Connectivity – DCM: Parameters

The ingredients to solve the system inside the approach of dynamic causal modelling needed to be contextualized every time. In the situation reported above as instance there is just one compartment where it could be seen the presence of an auto correlation. Hence, we could express our system with the next mathematical formulation:



$$\frac{dx_1}{dt} = a_{11}x_1 \Rightarrow x_1(t) = x_1(0)e^{a_{11}t}$$

By solving the differential equation we could see that the auto coefficient represents the decay time. This has unit of measurements equal to the one used for the time expression but taken the inverse. Thus, if the value is greater than zero the state x_1 of the compartment will increase its content, instead, if the value is lower than zero the state x_1 will decrease in time. Another important and basic situation that could be analysed with this technique is when we consider the model made up of just two compartments. This is such a basic and simple solved situation. The differential equation and the model graphical representation is reported below:

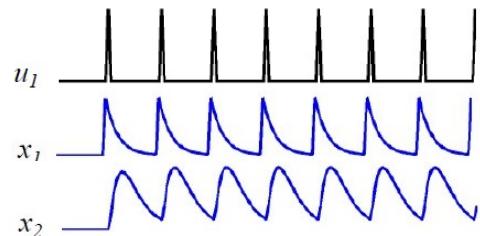
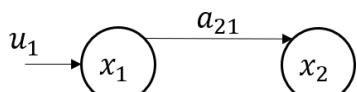


$$\frac{dx_B}{dt} = a_{BA}x_A \Rightarrow x_B = f(x_A, t, a_{BA})$$

In this case the analytical solution of the system is equivalent as the one reported and found for the system made up of just one compartment. In this case the coefficient a_{BA} let us be able to identify the fraction of content of the compartment A that leaves this last one and moves towards B .

Brain Connectivity: Effective Connectivity – DCM: Context dependent enhancement

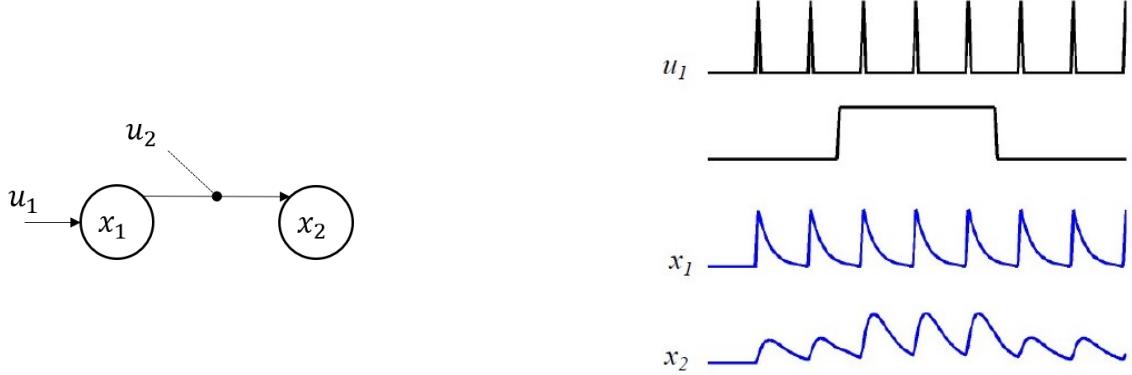
With the DCM technique we could investigate if some stimulus proposed to the subject could changes the neuronal connection between the different neuronal populations identified inside the net. Especially, if we proposed to the previous situation an external generical stimulus u_1 we could induce some changes in the net.



In this case we are introducing some perturbances generated with the input u_1 and we could measure the signal intensity coming from the neuronal populations inside the compartment 1 and the compartment 2. This situation presents also a basic mathematical description summarized in the next vector matrix form:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} c_{11} \\ 0 \end{bmatrix} u_1$$

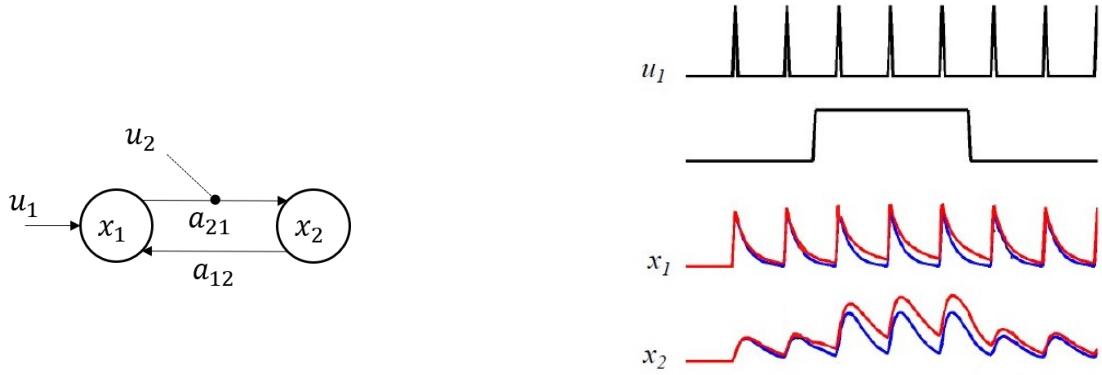
Another possible situation that we could face with is the next one. In this case, the changing of the state x_2 is modified and dependent on the value of the context input u_2 . Thus, this means that, when certain contexts are activated the second compartment could face some modification on the state variation.



In this case it could be seen that when the value of the context is in a particular situation the state of the compartment number 2 changes and its state too. Respect to the normal situation there is an enhancement of the state in the compartment number 2 dependent on if the context is in a particular condition or not. The mathematical description of this situation is reported below:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + u_2 \begin{bmatrix} 0 & 0 \\ b_{21}^{(2)} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} c_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

In some situation it can be that two considered compartments have reciprocal connection during specific context but in other not. This could be seen with the next model structure and behaviour inside the different analysed term.



In this case we could see that when the context u_2 occurs the activity inside the second compartment changes. Due to the feedback effect also the one inside the first compartment changes a little bit the pattern. From the mathematical point of view the situation could be summarized with the next expression:

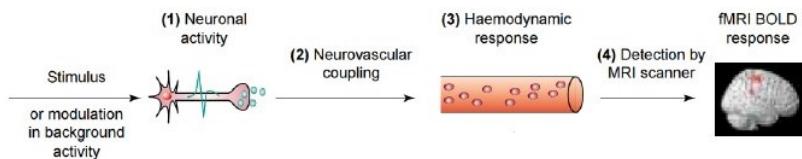
$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + u_2 \begin{bmatrix} 0 & 0 \\ b_{21}^{(2)} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} c_{11} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

Brain Connectivity: Effective Connectivity – DCM: Haemodynamic model

In the previous structure all the compartment represents the activity of the singular populations of neurons. From the state equations point of view the mode could be described as:

$$\dot{x} = f(x, y, \theta^C) = \left(A + \sum_{j=1}^m u_j B^{(j)} \right) x + Cu$$

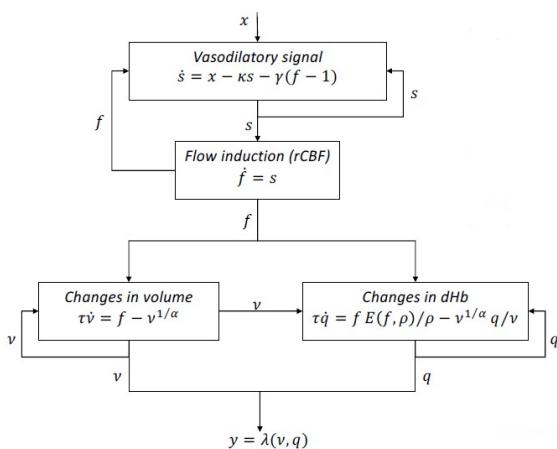
Where the θ^C vector is defined as $\theta^C = [A, B, C]$ which represents the connectivity parameters. These one are the objective of our studies. As we have seen before, these parameters and configurations are dependent on the state condition analysed and inspected. Moreover, during the measurements we could not analyse the neuronal states because they are hidden. Consequently we need to define some haemodynamic model to link the neuronal state to the BOLD signal measured. In this case the cascade of events that will reach and be developed as the neuronal state is summarized in the next scheme:



From a mathematical point of view we could link the neuronal state to the BOLD signal measured through the following relation:

$$y = g(x, \theta^h) + \epsilon$$

Where θ^h represents the haemodynamic parameters vector and ϵ the measurement error. The model is based on the CBF and the CBV and it is based on the changes of these two factors to have a proper description of the BOLD signal measured.

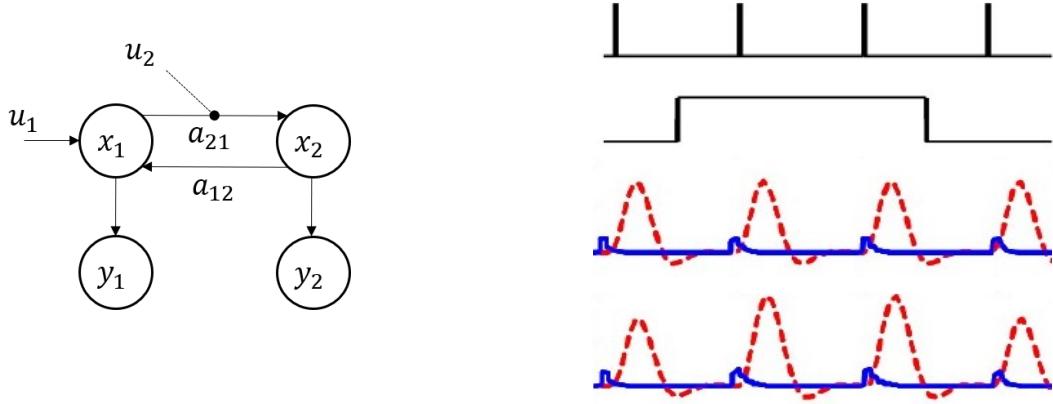


Parameter	Description	Prior mean η_θ	Prior variance C_θ
κ	Rate of signal decay	0.65 per second	0.015
γ	Rate of flow-dependent elimination	0.41 per second	0.002
τ	Haemodynamic transit time	0.98 second	0.0568
α	Grubb's exponent	0.32	0.0015
ρ	Resting oxygen extraction fraction	0.34	0.0024

This schematic shows the architecture of the haemodynamic model for a single region. Neuronal activity induces a vasodilatory and activity-dependent signal s that increases the flow f . Flow causes changes in volume and deoxyhaemoglobin (v and q). These two haemodynamic states enter an output non-linearity to

give the observed BOLD response y . The parameters that needed to be estimated for each region are reported in the table above.

The final problem is made up of estimating the parameters for the haemodynamic model and the neuronal activity model. Down below is represented graphically the problem, the final structure comprehensive also of the haemodynamic model and the neuronal activity model. Here it has been used the cascade of events previously reported that describes the activation of the haemodynamic response starting from the neuronal activity model. In this case it is also possible to see the presence of a bold and the neuronal activity signals overlapped.



Where in this case the first couple signals reflect the activity of the two neuronal populations without the context activated. While the second couple signals reflect the activity of the neuronal populations when the context intervenes.

Brain Connectivity: Effective Connectivity – DCM: Parameter's estimation

In this case we need to estimate a lot of parameters. In general we could refer to our models by summarizing them with the next two following relations:

$$\dot{x} = f(x, u, \theta^C)$$

$$y = g(x, \theta^h) + \epsilon$$

Where all the terms are well and already defined in the previous paragraph. Thus, we could not use a linear estimation paradigm to find the whole information. For this reason we need to solve the DCM with some Bayesian estimation approach. This last one is characterized from a structural point of view as:

$$P(\theta|y, M) = \frac{P(y|\theta, M)P(\theta|M)}{P(y|M)}$$

Where the terms that made up the previous relation has specific meaning. Specifically:

- $P(\theta|y, M)$ is called the *posterior*.
- $P(y|\theta, M)$ is called the *likelihood*.
- $P(\theta|M)$ is called the *prior*.
- $P(y|M)$ is called the *model evidence*.

In all the previous representation are used the θ and the M terms. The first one represents the parameters that needed to be estimated. In fact, in this case we have: $\theta = [\theta^h, \theta^c]$. While M represents the model. In fact, the parameters distribution is conditional to the model tested. In the next paragraph we will analyse each term of the Bayes relation applied in this condition.

The first thing that we are going to analyse is the *model*. This is a key aspect of DCM in fact the parameters estimated depends on it. Models can differ, as previously seen, for the structure as instance the matrix A , where the stimulus enter information contained inside the matrix C or which connection the stimulus is modulating represented as the matrix B . The self – connections could not vary the parameters estimation because they are always present inside for theoretical and physiological reason. In general is not possible to test all the possible combinations of connections and stimuli (grow exponentially with the number of regions). We need some knowledge about the system to correctly analyse it.

The *likelihood* represents another important concept and is determined by the model error. In fact, if the haemodynamic model used presents some additive noise with Gaussian distribution described by zero mean and covariance matrix C_e the prior has a Gaussian distribution too. Hence:

$$\begin{aligned}\dot{x} &= f(x, u, \theta^c) \\ y &= g(x, \theta^h) + \epsilon \Rightarrow P(y|\theta, M) \sim N(g(x, \theta), C_e) \\ \epsilon &\sim N(0, C_e) \\ \Rightarrow P(y|\theta, M) &= \frac{1}{\sqrt{(2\pi)^k \det(C_e)}} e^{-\frac{1}{2}(y-g(x, \theta))^T C_e^{-1} (y-g(x, \theta))}\end{aligned}$$

Generally as all the other estimation processes (in the most of cases) the covariance matrix C_e is assumed to be known up to a proportionally constant $C_e = \sigma^2 B$. B is known while σ^2 is estimated together with the other parameters in fact there is a prior for σ^2 too.

The *priors* represent all the a priori information available to solve the system and find the parameters value to give to each term in the parameter vector. The prior used in the DCM analysis are of different types and they are assumed to have a Gaussian distribution. They could be categorized inside the next following categories:

- *Principled priors*: The system must be stable, all the self – connections must be negative and thus they can be called as self – inhibition connections.
- *Physiological priors*: In this case instead the haemodynamic priors are set from the knowledge available from the literature.
- *Shrinkage priors*: For all the connections other than self – connections, the prior mean is set to 0. The idea is that we need high evidence from the data to prove that the consider connection exists. The variance of the connections is chosen to allow for neural transient with a half life in the range of 300 ms to 2 second.

The *model evidence* instead is given by integrating out the dependence on the model parameters. Mathematically this could be expressed with the relation reported in the following page.

$$P(y|M) = \int P(y, \theta|M) d\theta = \int P(y|\theta, M)P(\theta|M) d\theta$$

Because we have generated the marginal distribution over θ the evidence is also known as the marginal likelihood. For linear Gaussian models there is an analytic expression for the model evidence. DCM are highly non – linear and thus the integral does not have any analytical solution and numerical integration is too expensive. The posterior distribution $P(\theta|y, M)$ cannot be easily evaluated. It is just approximated.

Brain Connectivity: Effective Connectivity – DCM: Bayesian Model Selection

When we perform some DCM analysis, and we estimate all the parameters with the Bayesian approach also the model selection could be performed with a Bayesian approach too. Thus the analytical relation used to describe that with a certain parameters vector is necessary a certain model M is defined in the next way:

$$P(M|y) = \frac{P(y|M)P(M)}{P(y)}$$

Where the terms inside this expression could be considered as:

- $P(M, y)$ represents the *posterior*.
- $P(y|M)$ represents the *likelihood*.
- $P(M)$ represents the *prior* information.
- $P(y)$ represents the probability that the given y is assumed.

The value of the parameters vector is computed for a few amounts of combinations and models. We need to identify one which is the basic one where we could compute the possible ratio. Then we could calculate all the *Bayes factor* by evaluating the ratio between the model evidence of the two considered model M_1 and M_2 . This last expression follows the next analytical representation:

$$B_{12} = \frac{P(M_1|y)}{P(M_2|y)} = \frac{P(y|M_1)P(M_1)}{P(y|M_2)P(M_2)} = \frac{P(y|M_1)}{P(y|M_2)}$$

The value of B_{12} usually could be computed also with natural logarithm applied to each term of the relation. In this case we could have the following definition:

$$B_{12} = \log P(y|M_1) - \log P(y|M_2) \approx F_1 - F_2$$

In this way we could select which is the best model to apply in our situation. Now that we have seen the overall general analysis and category of application of the DCM we continue the lesson with some relevant examples.

Brain Connectivity: Effective Connectivity – DCM: Example of Synesthesia

The first example that we consider is the synesthesia phenomenon. This pattern was and it is still widely studied and the most relevant results were found by analysing the model with the DCM approach. In this case before introducing the studies that were performed, we need to consider and understand the definition

of the synesthesia phenomenon. In this case the specific sensory stimuli lead to unusual additional experiences. In particular the experiments that were performed is to undergo the subjects on the auditory stimulus of presenting a certain word and let them associate all the possible colours to the different letters. Generally in subject that are characterized from the Grapheme – colour synesthesia they will activate an involuntary automatic association of each letter to a specific colour. The potential investigated causes are due to the aberrant cross activation between brain areas. Specifically the *grapheme encoding area*, the *colour area V4* and the *superior parietal lobule SPL*. In two studies directed from Ramachandran and Grossenbacher shows that the effective connectivity reflects the conscious experience. In fact, with the two approaches used it could be possible identify the two-following models:



Brain Connectivity: Effective Connectivity – DCM: Example of Attention to Visual Motion

This experiment was used to assess the site of attention modulation during visual motion processing of fMRI paradigm reported by Buchel and Friston. In this situation we need to analyse three distinct brain areas and we need to investigate which is the model that connects the next three brain region define with the abbreviation *SPC*, *V1* and finally *V5*. By using different experiments we should investigate the connection between them. Each of the three – brain region involves the photic, the motion and finally the attention properties of the subject. To reach the results and the conclusion the research group need to think a lot on what happens inside the different model structure and finally through a model selection process has reached the final configuration of the model.

Brain Connectivity: Effective Connectivity – DCM: Pro and Cons

As all the techniques analysed during the course also the DCM presents some pro and some cons. In particular as pro it incorporates a physiologically plausible haemodynamic model that consider region differences, it is a robust framework from model selection, the results are evaluated via electrophysiological studies and it is fully Bayesian. On the other side, the con of this technique is that it is nonlinear and thus is very computationally expensive. In addition requires some hypothesis about the models to be tested.

LESSON 15: DIFFUSION MRI BASIC PRINCIPLES

With the last lecture is end up the chapter on fMRI. To conclude the analysis we need to say that the fMRI nowadays is used the most inside the research field. Sometimes this could be used also in clinical practice especially when we need to perform some neurosurgery. In this case is important to have some pre scan of where the different brain areas and regions are localized.

Before introducing the concept of the diffusion MRI we need to remember that each MRI image is based on the signals coming from the different protons. These signals could be modified as we have seen in the fMRI paradigm from the presence of haemoglobin or deoxyhaemoglobin. The modification of the signals could be also due to the different gradient applied during the acquisition or the different paradigm used to acquire the image. Some of these examples are the T1w and the T2w images.

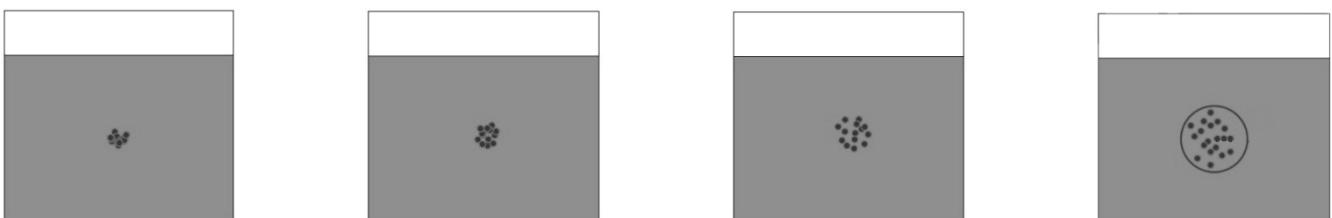
Diffusion MRI

In the Diffusion Magnetic Resonance Imaging is combined the two previous effect to change and modify the magnetic signal produced to enhance certain components. Especially in Diffusion MRI is combined the protons movement and the different gradient applied to obtain different results that enhanced different brain structures. This technique is more used in clinical practice than the fMRI because reveal some physiological and anatomical information inside the brain which are very useful to enhance specific brain pathologies or injuries.

The Diffusion MRI are subdivided into mainly two classes the *DWI* and the *DTI*. The first one is the acronym which stand for *Diffusion Weighted Imaging* while the second stand for *Diffusion Tensor Imaging*. As previously said the techniques are largely used to detect the protons movement. But because the major quantity of protons is contained inside the water molecules. Hence, in these techniques, is like studying the movement known as diffusion of the water molecule inside the brain. In physics, diffusion is the results of the thermal translation movement of the molecules. Thus the movement is called *Brownian*. This last motion has been first described by Einstein in 1905. The movement seems to be random at the first sight, but it is not.

Diffusion MRI: Diffusion – Isotropic situation

To understand this motion we need to consider the dumb example reported below. In the following image are reported a little group of molecules concentrated in the middle of the box fill with water. Our objective is to describe the movement of the different molecules analytically.



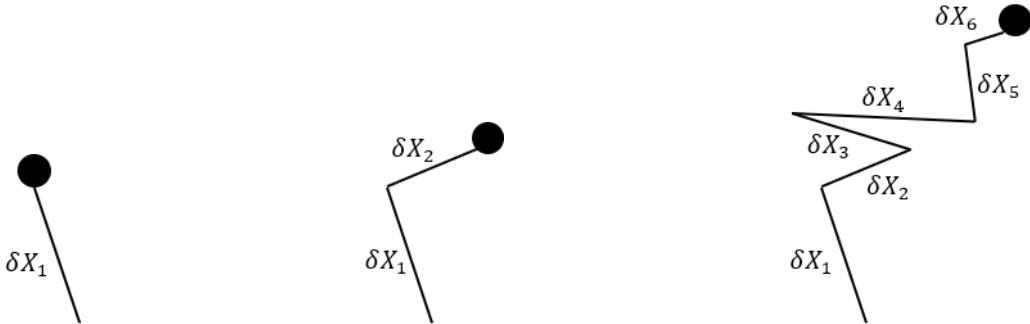
We can see that from an initial instant of time where all the different particles are constraint close an initial point and are let free to move inside the liquid. In this situation after a certain interval of time the particles

changes their initial position to a more spread configuration. This is due to the Brownian movement. Now, let suppose that after a time T_d (last figure reported on the left), all the particles could be considered as contained inside a circumference of radius r . In this case by applying the Einstein law for the Brownian motion we could get:

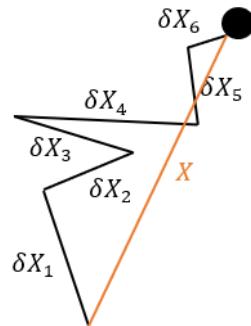
$$\langle r^2 \rangle = 2DT_d$$

Where in this case the value of r and T_d are already defined, while the value of D represents the value of the diffusion coefficient. In this situation the diffusion has been analysed without a preferential direction. For this reason, the Brownian motion in this case is called an isotropic diffusion. Thus no preferential direction could be found. In particular, because of we will see that the diffusion coefficient D could be also a matrix, in this case of bidimensional movement is made up of two terms, the term D_x , D_y and D_z . In this case, the term D_z does not exists because the situation considered is just planar and not three dimensional. However, inside a situation where could be studied an isotropic diffusion movement all the diffusion coefficient needs to be equal.

From the mathematical point of view the Brownian path can be split into the sum of N small random steps, performed at each small – time interval, with mean equal to 0 and fixed values of variance. Let us visualize down below some of the steps of the random walk. Again, it is important to specify that each step has zero mean and fixed variance.



The first situation reported on the left represent a situation where the random walk has just performed one single step. In the second two and in the third six. During the random walk after a certain number of steps we are not interested in defining each step in a deterministic way. What is important and also what we could actually do is understanding how the random movement X is statistically distributed in the space.



Because of the value of X is represented as the sum of all the possible infinitesimal steps δX_i for all the values of i inside the set of elements $\{1, \dots, N\}$ where N represents the number of steps executed from the

random walker. In this case, because of the single infinitesimal steps does not have a statistical description we could assumed them as general random variable. But, because of the variable X is sum of unknown random variable, for a high number of steps, is a random variable too and thanks to the central limit theorem the distribution of the random variable is Gaussian. The probability distribution of X is known also as the *Diffusion propagator*. The variance of the diffusion propagator is described by the Einstein's law. Hence:

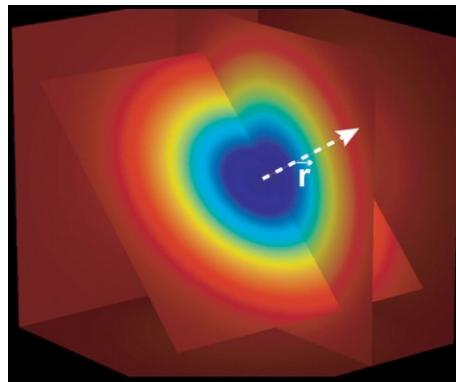
$$\langle X^2 \rangle = 2DT_d$$

This relation and these results are valid only when the diffusion is isotropic. From the molecular level the diffusion coefficient depends on the *temperature*, the nature of the medium known as *viscosity*, and the *molecule dimension*. The diffusion coefficient of the free water at 37° C is of about $3.5 \times 10^{-9} m^2 s^{-1}$. In tissues the water coefficient is lower because of obstacles such as membrane, cells, macromolecules, fibers and other stuff. Thus, the first important and practical consideration is that the diffusion coefficient of the water inside our brain is not equal to the diffusion coefficient of the free water. This because there are a lot of tissues, molecules, and cells.

If we assume that water molecules can diffuse, in a completely free way, then diffusion propagator has a Gaussian distribution with zero mean. Thus analytically this could be expressed with the next probability density function:

$$P(x, t) = \frac{1}{\sqrt{4\pi D t}} e^{-\frac{x^2}{4Dt}}$$

As we could see, if we fix the value assumed from D , bigger is t bigger is the width of the Gaussian distribution. Instead, if we fixed the instant of time t , bigger D bigger is the width of the distribution. If we take a slice of the region where the diffusion shows an isotropic trend, we could expect the next spatial distribution.



We could see from the image that the diffusion happens along the three major direction. At 37°C with a diffusion time of $T_d = 50 \text{ ms}$ the 32% of the molecules in a environment that let them a free diffusion will move of $17 \mu\text{m}$, while only the 5% of the molecules will move of $34 \mu\text{m}$. Passing from the plane x to a three-dimensional space, the Gaussian distribution assume the analytical expression reported below:

$$P(r, t) = \frac{1}{\sqrt{(4\pi Dt)^3}} e^{-\frac{r^T r}{4Dt}}$$

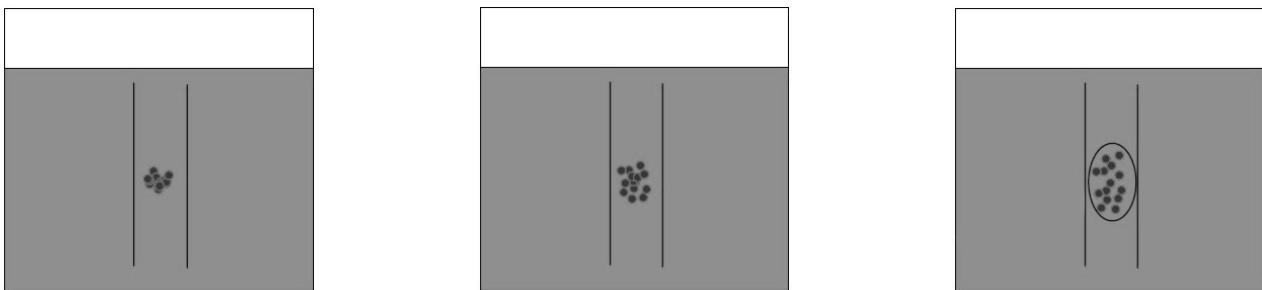
This surface is called diffusion sphere when the next equality is satisfied.

$$x^2 + y^2 + z^2 = (\sqrt{2Dt})^2$$

The radius of the sphere is equal to the root of the Einstein Law. In this case because the value of D is constant among the main directions we talk about *Diffusion Weighted Imaging* or *DWI*.

Diffusion MRI: Diffusion – Anisotropic situation

Let's consider the same glass of water of the before analysis. Let consider now that we have also some constraints inside the glass to let the particles just spread over a preferential direction. As the previous case, down below are reported some images that described the evolution in time of different particles distribution.



As could be seen from the different cases reported before the initial constraint of the particles close to the initial point is as the time pass not truer. In fact, as the time pass, due to the presence of two walls along the x direction the molecules just spread over the y direction. Thus, the diffusion coefficient in this case is not more equal for all the direction. This condition is known as *anisotropy diffusion movement*. In this case, the diffusion among the x direction is not more equal to the one towards z and y. This mathematically could be expressed with the next inequality:

$$D_x \neq D_y \neq D_z$$

One important difference between this condition and the previous situation where the diffusion was isotropic is that, if we create the figure with principal axes the diffusion dimension, we could get in this situation an ellipsoid while in the isotropic situation a perfect sphere.

Human tissues can survive only in a narrow interval of temperatures in which most tissue components are liquid. Therefore, water's diffusion plays a fundamental role in enzymes transport of metabolic substrates and of metabolites. Moreover, tissues show a very inhomogeneous macroscopical structure. Cellular membranes and cell's organ hindered the free motion of water and other molecules. Hence, if we think to the brain this is an organ made up of different tissues and depending on what surrounds the water this last substrate will behaves differently. This is way at the beginning of the lesson we have talked and introduced

that measuring the water motion can be a valid tool to describe tissues microscale structure, beyond the resolution of usual imaging techniques.

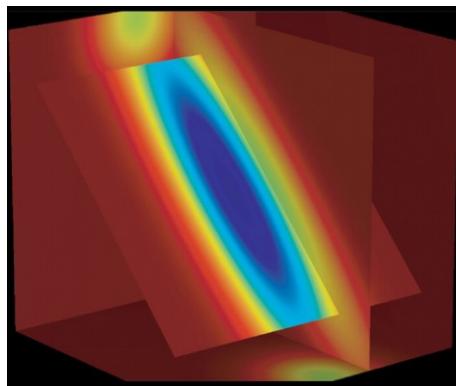
To describe the anisotropic diffusion with the probabilities concept we could use the next analytical expression:

$$P(r, t) = \frac{1}{\sqrt{(4\pi|D|t)^3}} e^{-\frac{r^T D^{-1} r}{4t}}$$

In this case, the value of D could be summarized with a tensor that account for the different value of the diffusion coefficient along the principal directionalities.

$$D = \begin{bmatrix} D_{xx} & D_{xy} & D_{yx} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}$$

Where in this matrix we account for all the direction and an important consideration which will be used later on during the analysis of the tensor estimation is that this matrix could be seen as symmetric. The surface that is generated in this case is called *diffusion ellipsoid* and shows the next trend obviously with a specific direction that could differ from this example.



In general when we would like to investigate the directionality of the diffusion coefficient, we are performing an anisotropic diffusion analysis and we could talk about the *Diffusion Tensor Imaging* or *DTI*.

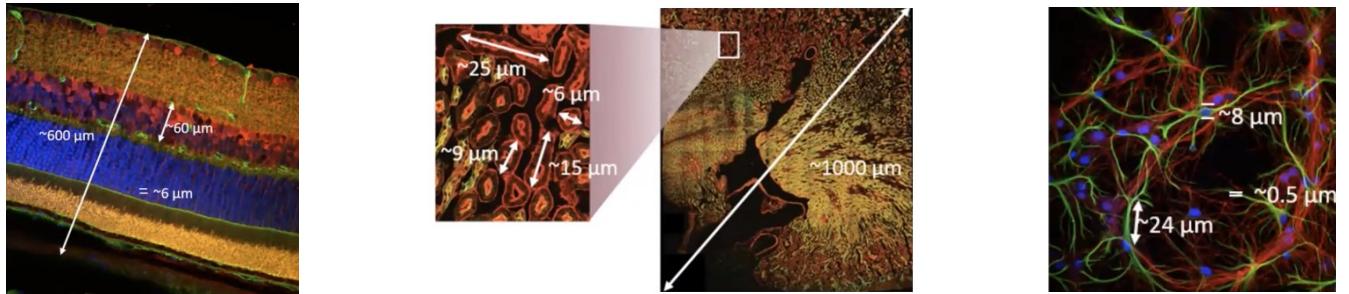
Diffusion MRI: Diffusion – General practical concepts

With *DWI* or *DTI* it is possible to obtain images in which the signal intensity is related to the random water molecules motion, thanks to the sequence of intense magnetic field gradient pulses, applied before and after a radiofrequency pulse of 180 degree. In this case the static magnetic field do not need to reach high magnetic field intensity because the number of magnetic fields applied is greater than the initial T1w or T2w spin echo paradigm.

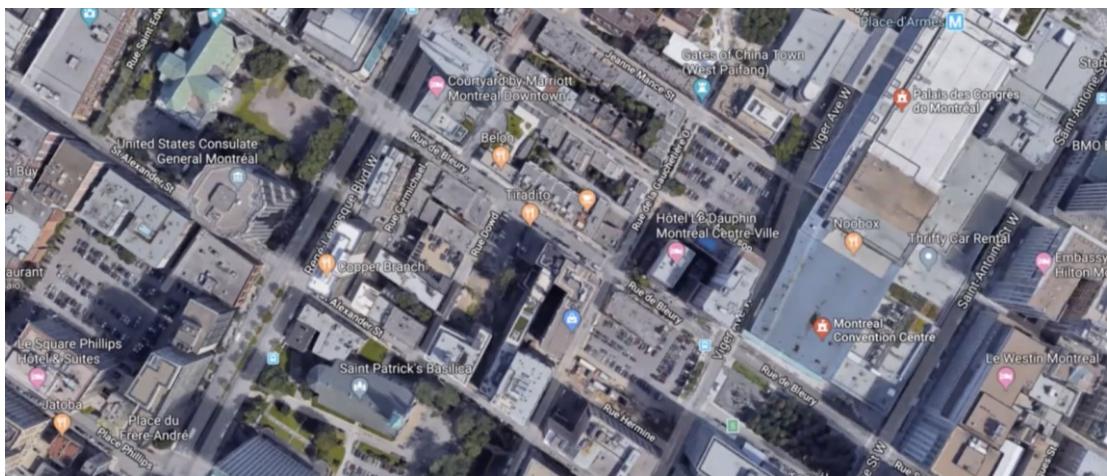
In practice, interaction between the intra and extra cellular space influences the water molecules diffusion, causing a perturbation of its diffusive flux. Diffusion measures, therefore, allow to investigate cellular functionality and integrity, both in normal and pathological conditions.

Diffusion MRI: Diffusion – Diffusion overview

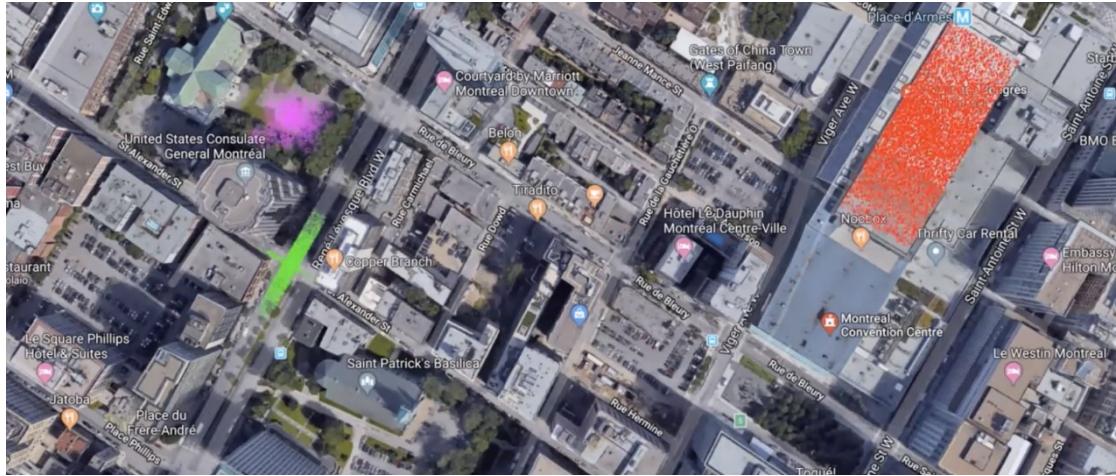
Our aim of these analysis is to understand the tissue macrostructure by using the diffusion MRI. In any biological tissue and not only in the brain presents a tissue microstructure heterogeneous and complex. In fact, as could be seen below in the figures the tissue structure is made up of different component size. Down below are reported some images that enhance the wide variability available inside our body in terms of length scale.



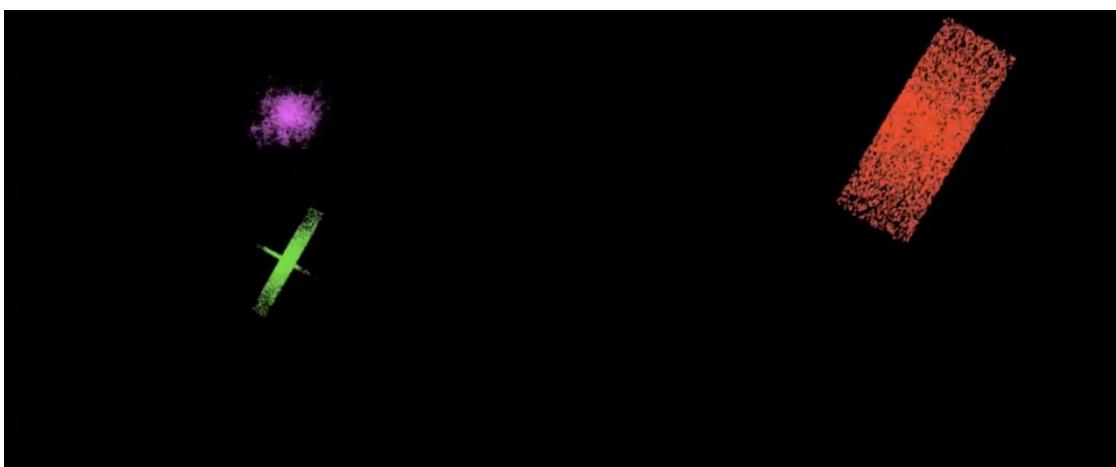
We would like to have an image that could quantify this complexity in terms of tissues structures and also something that could quantify the different length scale. With the usage of usual MRI we could get just information with 1 mm of accuracies. Thus the T1w and the T2w are not suitable for this application. Instead, by the usage of the histological images we could have a better resolution until $2 \mu\text{m}$ but presents a cons because it is a very hard invasive technique. Even if it is invasive, it is the best way to get some information on the different time scales in our situation. The challenge is to fill the gap between the MRI resolution and the microscopy. If we look at the typical distance probe by diffusing endogenous water, we could reach at least 1 to $100 \mu\text{m}$. Thus the diffusion of water could give us information about the tissue microstructure and at least the previous length scale mentioned. Now let see a very simple example of how the diffusion could give us important information about the tissue micro and macro structure. This example is not strictly related with the physiology, medicine or biology but consider a google map street view.



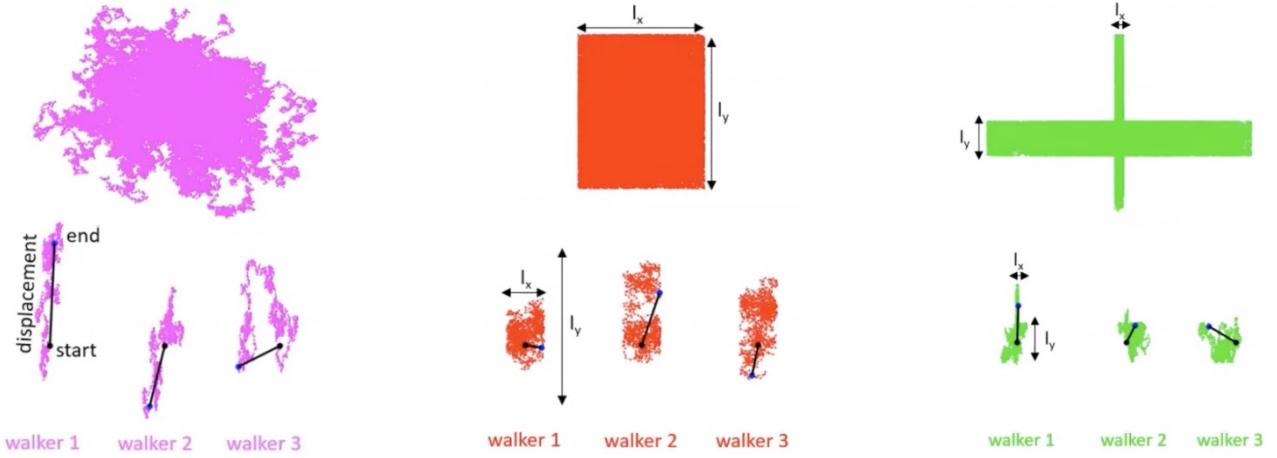
As we could see the image represent a map where we would like to explain what the diffusion is. In particular if we put a marker for each person inside the *Palais des Congres de Montreal*, the people inside the corner between *St Alexander Street* and the main road, and finally inside the people is the park in front of the *United States Consulate General Montreal* we could see after a certain interval of time that the google map reported before start to be more coloured.



We could see that the people inside the congress centre will move just inside the congress building, the people at the corner between the two streets will follows the available streets and finally the people inside the park diffuse themselves trying to avoid the different trees. Thus we could understand that we if we remove the google maps from the bottom of the maps where are enhanced the different diffusion movement of all the different marked people, we still have some important information that could let us understand what is happening to those people marked. In fact, in the next image it could be seen that the people on the congress moves inside a rectangle and knowing that these information are retrieved from a geographical map we could easily understand that this shape is characteristic of a building. The green diffusion pattern is characteristic of two streets that crosses and finally we could not say nothing for the purple distribution but in general it could be said that the movement is almost free.



That's the power of diffusion. If we transpose this concept on the water molecules inside the tissue, we could see and look to the displacement and we can infer the structure where the molecules are moving. In general the displacement is the distance between the initial and the end point of the random walk performed from a molecule. Sometimes, the random walk could be constraint from some characteristic length of the diffusion condition. To conclude the previous example we report down below the three different configuration that could be analysed.



The patterns of diffusion displacement is sensitive to underlying structure. Now let see how this concept could be applied to biology in general or to a specific tissue. The main concept that we have already mentioned and defined before is that the diffusion could be considered as a random walk described with a random variable with mean zero and a given σ^2 variance. Thus these two last considerations means that we will not stay fix in a determine position, but we will move far apart.

After a certain amount of time T_d a given random walker is moved away from an initial position and the overall displacement id given, as previously said, as the sum of all the random steps δX_i . The random variable because of the central limit theorem follows a gaussian distribution. To describe completely the Gaussian distribution we need to determine the mean value and its variance. The mean is zero because of the mean of the single variable, instead the variance is:

$$\langle X^2 \rangle \sim N\sigma^2 = T_d \frac{\sigma^2}{\delta t}$$

We could calculate also the mean square displacement MSD and we could account for the variance of the different steps defining the diffusion constant D :

$$\langle X^2 \rangle \sim DT_d \quad D \sim \frac{\sigma^2}{\delta t}$$

In the case where the central limit theorem holds the relationship between the mean square displacement and the diffusion time T_d is linear and thus the diffusion coefficient is constant.

The random walk approximation of the diffusion holds just for a timescale greater than the inter molecular interaction. This means that the diffusion time needed to be much greater than the value of 10^{-9} s. In fact if we look to the dynamics of water molecule inside small times, they interact a lot between each other. In

fact at that time scale we could not describe the diffusion as a random walk. Typical T_d in diffusion MRI experiments is of about $10^{-3} - 0.5$ s. This time scales are a perfect to study the Brownian motion. The typical values of D in biological tissues are $D \sim 0.5 - 3\mu\text{m}^2/\text{ms}$.

In complex system or diffusion restricted within structured the Central Limit Theorem is violated and the diffusion propagator is not Gaussian. The diffusion propagator can be shown characterised from the next properties:

$$\frac{\partial P(X, t)}{\partial t} = D \frac{\partial^2 P(X, t)}{\partial X^2}$$

One-dimensional case

$$\frac{\partial P(X, t)}{\partial t} = D \nabla^2 P(X, t)$$

Three-dimensional case

The solution for dimension d , in the case of free diffusion is still a Gaussian and the variance is proportional to the mean diffusivity time, the coefficient of diffusion and changes according to the dimension considered.

$$P(X, t) = \frac{1}{(4d\pi Dt)^{d/2}} e^{-\frac{X^2}{4dDt}}$$

Diffusion propagator

$$\langle X^2 \rangle = 2dDt$$

Variance of diffusion propagator

Einstein relate the diffusion coefficient with the *temperature*, medium *viscosity* and *hydrodynamic radius*. This is very important because in this way we are differentiating the behaviour and the pattern of the molecules in the space.

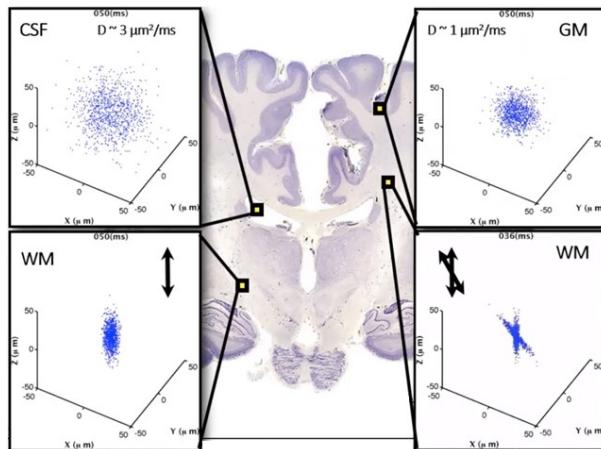
$$D \sim \frac{T}{R\eta}$$

This is very important because in those pathologies where the viscosity increases in specific brain area it could be detected by estimating first the diffusion coefficient. As previously reported the variance of the diffusion propagator is dependent with the analysed time T_d , thus also the probability distribution in time changes in particular it could become flatter because the diffusion spreads all the different molecules. The measured diffusion weighted MRI signal is proportional to the mean diffusion propagator. This last one is the convolution between the structure and the diffusion movement of the particles. Thus, the mean diffusion propagator represents a situation where the molecules are already distributed inside a specific region and they keep staying inside that specific region. This does not mean that the diffusion propagator is not informative, but it is informative the basic of the structure. But if we would like to relate the diffusion propagator with the MRI signal, we need to connect its meaning with the mean diffusion propagator. We cannot really follow the small movement of molecules, but we can use the mean value to derive the path. The mean is still very informative.

When we take the inverse of the square root of the mean squared displacement is informative for the cell size and thus diffusion MRI probes tissue macrostructure.

In the brain as we know we have three main components the white and grey matter and the cerebral spinal fluid. In the list reported below we could see the different patterns for each tissue of the brain analysable.

- *Grey matter*: In the Grey matter the diffusivity is lower, this because there are a lot of cells that represents some physical constraints of movement for the water molecules. In this case the diffusion coefficient is of about $1\mu\text{m}^2/\text{ms}$.
- *White matter*: In the White matter the situation is very relevant. In fact, in this case we could observe something very different. The propagation is not more isotropic, and this is due to the high amount of fibers that are present inside the brain. In particular along the fibers directionality the distribution of the different molecules could be modeled as a Gaussian, but this would not be Gaussian in the perpendicular direction of the fibers. This last distribution in fact clearly shows the hallmark of the restricted compartment. In the parallel to fibers direction the diffusion coefficient is about $2\mu\text{m}^2/\text{ms}$ while in the perpendicular direction is of about $0\mu\text{m}^2/\text{ms}$.
- *Cerebral Spinal Fluid*: In this case the diffusion coefficient is of about $3\mu\text{m}^2/\text{ms}$. The diffusion is isotropic and thus there is not a preferential direction that the different particles will follow. The previous value is the maximum value that we could find when we look inside our brain.



Diffusion MRI: Signal

To quantify the diffusion coefficient or the diffusion tensor we need to get some measurements where to perform some fitting process. The signals that are acquired that could let us be able to identify these quantities are coming from MR scanners where are applied specific paradigm. Before introducing the main paradigm used, we need to have a look on where we can extract these types of information from the MR signal.

First, we need to consider the application of magnetic field gradient along an axis, for example x, this cause a dephasing ϕ of the spin magnetic moment μ , that is function of its position x_i along the x axis and is equal to:

$$\phi_i = \int_0^\delta \gamma G x_i dt$$

Where G represents the intensity of the gradient applied along the x axis, δ its duration and γ the gyromagnetic ratio. If the spin is stationary the application of two consecutive pulses of bipolar gradient (same direction but opposite amplitude to re – phase the spins makes the dephasing null).

If the spin is subjected to a diffusive process, in its motion along gradient direction, it undergoes a net dephasing equal to:

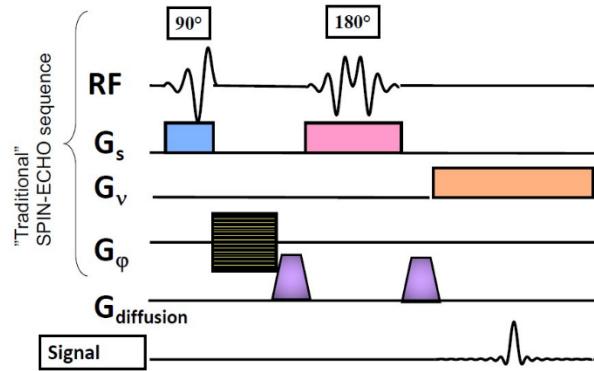
$$\Delta\phi = \phi_1 - \phi_2 = \gamma\delta G(x_1 - x_2)$$

If the particles do not move the value of the dephasing is zero. Nuclear spins subjected to a magnetic field gradient pulses with amplitude G , duration d , applied after an interval Δt , will be subjected to a dephasing that will be translated to a net reduction of the total magnetization.

Using a specific MRI sequence called *Stejeskal – Tanner*, MRI signal results is related to G , δ , γ , Δt and D . This leads to acquire the next magnetic signal:

$$S = S_0 e^{[-(\gamma\delta G)^2 \Delta t D]}$$

Where S_0 is the MRI signal without the application of the gradient G . The sequence used in the scanner to get the previous described signal is made up of the next paradigm.



The real problem is to move from the dephasing equation reported at the beginning of this page to the last equation reported previously. This passage is possible because we can introduce in the last equation a term that account for a part of the structure of the dephasing equation. This term that is introduced is the b value.

$$b = (\gamma\delta G)^2 \Delta t \Rightarrow S = S_0 e^{-b\Delta}$$

Magnetization reduction is proportional to the intensity of the gradient G and to the application time of the gradient δ , that should be short, otherwise the diffusion happening during the gradient application will be no more negligible.

The b value could be changes by modifying the interval of time where the gradient is applied, or the gradient intensity as previously said. Finally, we could change the signal S acquired from the scanner by modulating the value of b , the value of D and the value of S_0 . Thus, because of we could measure the intensity of the signal S_0 but we could not modify the diffusion tensor D the only thing that we could menage is the value of b that is adjusted by modifying the interval of time for which the gradient is applied. This because the intensity of the gradient G is chosen to be not modified.

```
1 %
2 % Title: Laboratory 6
3 %
4 % Note:
5 % fMRI Analysis and generation of the functional connectivity map
6 %
7 % Author: Matteo Martin
8
9 % THEORY -----
10 %
11 % GRANGER CAUSALITY
12 % The Granger Causality it is a recent technique developped in the '90 that
13 % help to discover and compute effective connectivity maps. This last type
14 % of map find the real connection between sources of signals. It is model
15 % based method and its crucial steps relay on idenfity the MVAR model
16 % between the signal considered. Specifically with this approach we need to
17 % identify an MVAR model and an AR model. This analysis is based on this
18 % concept. Given a signal  $x_1(t)$  and a signal  $x_2(t)$ , if we use just the
19 % past of the signal  $x_1(t)$  (known as target in the next analysis) to
20 % predict the instant of time  $t$  of the same signal and we obtain a variance
21 % of the output error equal to  $s_1$ , and then we consider also the past of
22 % the signal  $x_2(t)$  to predict the signal  $x_1(t)$  and the variance  $s_2$  of the
23 % output error is lower than the one before ( $s_1$ ) hence there is causality
24 % between the signal  $x_2(t) \rightarrow x_1(t)$ . To confirm this results also from a
25 % statistical point of view is requested to be performed an F - test.
26 %
27 % PERFORMANCES
28 % To evaluate the performances of the previous technique and its
29 % reliability we need to use compute different metrics such as the one
30 % reported below:
31 %
32 % - ACCURACY: The accuracy represents the amount of signals for which
33 % the eventual causalities is presented or not which are
34 % correctly identified.
35 % - TPR : TRP is the acronym for True Positive Rate. This index let
36 % us be able to identify the fraction on how many positive
37 % are really determine as it.
38 % - FPR : FPR stands for False Positive Rate. This index let us be
39 % able to identify the fraction of missclassification.
40 %
41 % Obviously to compute these metrics we must know the real classification
42 % and relationship between the different signals that are generated. For
43 % this reason we use specific function that let us be able to generate
44 % couple of signals that are connected or not and with a specific
45 % directionality.
46 %
47 % FUNCTIONS -----
48 %
49 % For this laboratory we need to introduce some specific functions that are
50 % necessary to be used to get the results requested. In particular the
51 % funtcions that are not known but are needed inside this program are:
52 %
53 % simuldata_uniform(nTIME, P, G, INT): This fucntion could be used to
54 % generate data with specific
```

```

55 %
56 %
57 %
58 % reshape_yar(STARGET, P)                                characteristic passed as argument
59 %                                                               inside nTIME, P, G and finally INT
60 %
61 %
62 % SCRIPTS -----
63 %
64 % The scripts that are written are different and in this program are
65 % contained two of these. In particular the first part of this code it is
66 % used to generate the matrix of signals and the real link between those.
67 % Next we identify the AR and the MVAR model and we perform an F test to
68 % evaluate if the two variance are statistically different or not.
69 % Finally, we evalute all the previous cited performances. Always in this
70 % code it is reported a section that could be used to identify the best
71 % MVAR model order to predit the possible causality between two distinct
72 % signals by looking on some parsimony criterion such as the Akaike and the
73 % Bayes Information Criterion.
74
75 %% CCC
76 % Clear all, close all, clear the comand window
77
78 close all; clear all; clc
79
80 %% LOAD - DATA
81 % Load of all the possible data inside the workspace of Matlab to compute
82 % correctly the functional connectvitiy matrix
83
84 load('rois_fmri.mat'), rTAC = ROI_fmriTAC; clear ROI_fmriTAC;
85
86 %% FUNCTIONAL CONNECTIVITY
87 % Computation of the funcitonal connectivity matrix by the usage of the
88 % Pearson's correlation and visualization of the different identified
89 % matrix visually.
90
91 [R, P] = corr(rTAC, rTAC);
92 Z      = atanh(R);
93 M      = max(triu(Z,1),[],'all');
94 m      = min(triu(Z,1),[],'all');
95
96 figure, subplot(1,2,1), imagesc(Z), colormap 'jet'; colorbar
97         title('FUNCTIONAL CONNECTIVITY'), caxis([-1 2])
98         subplot(1,2,2), imagesc(P), colormap 'jet'; colorbar
99         title('P - VALUES'), caxis([0 1])
100
101 % The matrix shows the functional connectivity matrix. The Fisher Z
102 % transformation have just an effect of the two tail of the distribution.
103 % There are a lot of entries of the matrxi that are positive but also some
104 % that are positive. Due to the Fisher z - fisher modify the interval range
105 % and let the maximum value be 2 and the minimum equal to -1. The p -
106 % values matrix is important too beacuse help us to understand where we can
107 % accept the correlation found or not. We can erase matrix's entries with
108 % the application of different threhsolding algorithm.

```



```

163                               xlabel('Time [ms]'), ylabel('Amplitude []')
164                               xlim([0 nTIME])
165 subplot(nVISUAL, 1, 3), hold on,
166                               plot(data(index(3),:,1),'b')
167                               plot(data(index(3),:,2),'r')
168                               hold off, grid on, grid minor, box on,
169                               title('SIM 3 - x_1(t) x_2(t)')
170                               xlabel('Time [ms]'), ylabel('Amplitude []')
171                               xlim([0 nTIME])
172 subplot(nVISUAL, 1, 4), hold on,
173                               plot(data(index(4),:,1),'b')
174                               plot(data(index(4),:,2),'r')
175                               hold off, grid on, grid minor, box on,
176                               title('SIM 4 - x_1(t) x_2(t)')
177                               xlabel('Time [ms]'), ylabel('Amplitude []')
178                               xlim([0 nTIME])
179
180 %% GRANGER CAUSALITY
181 % Computation of the Granger Causality by the usage of the provided
182 % functions. In this case, we take a target signal, we compute the
183 % parameters of the regression only by using a part of its samples. We
184 % compute the residual to identify the predictive power of the signal
185 % over itself. Then, we take the full system by considering the other
186 % signals along with the target using the full matrix. We compute again the
187 % predictive power of the autoregressive model using the information of
188 % also the other signals. Finally we compare the variances of the two
189 % residuals and by computing an F Test too we determine which signal is
190 % caused by another. In the next lines of code we use the next notation:
191 %
192 % Causality relation : sCOMPARISON -> sTARGET
193 % Vector LINK (if causality): [sTARGET, sCOMPARISON] = [0, 1]
194
195 nTIMETRAIN = 1:200;
196 nTIMETEST = 201:nTIME;
197 alpha = .05;
198 LINK = zeros(nTRIAL, M);
199
200 for idT = 1:1:nTRIAL
201     for idM = 1:1:M
202
203         if idM == 1, jdM = 2; end
204         if idM == 2, jdM = 1; end
205
206         % SINGLE PREDICTION
207         sTARGET = squeeze(data(idT,:,:idM))';
208         [Y, X] = reshape_yar(sTARGET(nTIMETRAIN), P);
209         B = (X' * X) \ X' * Y;
210         [Y, X] = reshape_yar(sTARGET(nTIMETEST), P);
211         SR = Y - X*B;
212
213         % MULTIPLE PREDICTION
214         sCOMPARISON = squeeze(data(idT,:,:jdM))';
215         sFULL = [sTARGET, sCOMPARISON];
216         [Yf, Xf] = reshape_yar(sFULL(nTIMETRAIN,:), P);

```

```

217      B          = (Xf' * Xf) \ Xf' * Yf;
218      [Yf, Xf]   = reshape_yar(sFULL(nTIMETEST,:), P);
219      mR        = Yf - Xf*B;
220
221      % F - TEST
222      sRSS = sum(sR.^2);
223      mRSS = sum(mR(:,1).^2);
224      F    = ((sRSS - mRSS)/P) / (mRSS/(nTIME - M*P));
225      FTH = finv(1-alpha, P, (nTIME - M*P));
226      if F > FTH, LINK(idT,jdM) = 1; end
227
228  end
229 end
230
231 %% PERFORMANCES - COMPUTATION
232 % Evaluation of the performances by the usage of the initial information
233 % contained inside the INT vector and inside the LINK vector. We compare
234 % the content inside the LINK vector with the content inside the INT one.
235
236 % LINK           INT           RESULT
237 % 0  0           -1           FN = FN + 1; TN = TN + 1
238 % 0  0           1            FN = FN + 1; TN = TN + 1
239 % 0  1           -1           FN = FN + 1; FP = FP + 1
240 % 0  1           1            TN = TN + 1; TP = TP + 1
241 % 1  0           -1           TN = TN + 1; TP = TP + 1
242 % 1  0           1            FN = FN + 1; FP = FP + 1
243 % 1  1           -1           FP = FP + 1; TP = TP + 1
244 % 1  1           1            FP = FP + 1; TP = TP + 1
245
246 TN = 0; FN = 0; TP = 0; FP = 0;
247
248 for idT = 1:nTRIAL
249
250     if (LINK(idT,1)==0 && LINK(idT,2)==0)
251         FN = FN + 1; TN = TN + 1;
252     end
253
254     if (LINK(idT,1)==1 && LINK(idT,2)==1)
255         FP = FP + 1; TP = TP + 1;
256     end
257
258     if (LINK(idT,1)==0 && LINK(idT,2)==1 && INT(idT)>0) || ...
259         (LINK(idT,1)==1 && LINK(idT,2)==0 && INT(idT)<0)
260         FN = FN + 1; FP = FP + 1;
261     end
262
263     if (LINK(idT,1)==0 && LINK(idT,2)==1 && INT(idT)<0) || ...
264         (LINK(idT,1)==1 && LINK(idT,2)==0 && INT(idT)>0)
265         TN = TN + 1; TP = TP + 1;
266     end
267
268 end
269
270 ACCURACY = (TP + TN) / (TP + TN + FP + FN);    % Accuracy

```

```

271 TPR      = TP / (TP + FN);                      % True Positive Rate
272 FPR      = FP / (FP + TN);                      % False Positive Rate
273
274 %% PERFORMANCES - VISUALIZATION
275 % Visualization of all the performances computed at the step before inside
276 % the comand window of Matlab.
277
278 fprintf('+++++++++++++++++++++++++++++++++++++\n')
279 fprintf('\tPERFORMANCES\n\n')
280 fprintf('CONFUSION MATRIX\n')
281 fprintf('      REAL\n')
282 fprintf('      0    1\n')
283 fprintf('PRED   0  %3i %3i\n',TN,FN)
284 fprintf('      1  %3i %3i\n\n',FP,TP)
285 fprintf('STATISTICS\n')
286 fprintf('ACCURACY:\t %1.3f\n',ACCURACY)
287 fprintf('TPR      :\t %1.3f\n',TPR)
288 fprintf('FPR      :\t %1.3f\n',FPR)
289 fprintf('-----\n')
290
291 %% MVAR - ORDER SELECTION
292 % Selection of the model order most suitable for the couple of signals that
293 % are considered during the iteration. The indexes calculated are the
294 % Akaike Information Criterion and the Bayes Information Criterion.
295
296 N      = length(nTIMETEST);
297 alpha  = .05;
298 LINK   = zeros(nTRIAL, M);
299 POBIC  = zeros(1,nTRIAL);
300 POAIC  = zeros(1,nTRIAL);
301
302 for idT = 1:1:nTRIAL
303
304     AIC = zeros(1, 9);
305     BIC = zeros(1, 9);
306
307     for P    = 2:1:10
308
309         mRT = zeros(M, N - P);
310
311         for idM = 1:1:M
312
313             if idM == 1, jdM = 2; end
314             if idM == 2, jdM = 1; end
315
316             sTARGET      = squeeze(data(idT,:,idM))';
317             sCOMPARISON  = squeeze(data(idT,:,jdM))';
318             sFULL        = [sTARGET, sCOMPARISON];
319             [Yf, Xf]      = reshape_yar(sFULL(nTIMETRAIN,:), P);
320             B            = (Xf' * Xf) \ Xf' * Yf;
321             [Yf, Xf]      = reshape_yar(sFULL(nTIMETEST,:), P);
322             mR           = Yf - Xf*B;
323             mRT(idM, :) = mR(:,1)';
324

```

```
325     end
326
327     Sp      = (1/(N - P))*( mRT * mRT' );
328     AIC(P-1) = log(det(Sp)) + 2*(P * M^2)/N;
329     BIC(P-1) = log(det(Sp)) + log(N)*(P * M^2)/N;
330
331 end
332
333 [ ~, temp] = min(BIC);
334 POBIC(idT) = temp + 1;
335 [ ~, temp] = min(AIC);
336 POAIC(idT) = temp + 1;
337
338 end
339
340 %% MVAR - ORDER SELECTION VISUALIZATION
341 % Visualization of the best model order by plotting the histogram.
342
343 figure, hold on,
344     histogram(POBIC),
345     histogram(POAIC),
346     hold off,
347     title('MODEL ORDER DISTRIBUTION'),
348     xlabel('ORDER []'), ylabel('A.F. []')
349     grid on, grid minor, box on
350     legend('BIC','AIC')
```

LESSON 16: DWI AND DTI

In the last lesson we have talk about the diffusion MRI and the different possible diffusion movement inside the grey and the white matter of the water molecules. Specifically the water molecules inside the grey matter could diffuse without any problem among the three main direction without difference inside the diffusion coefficient. Instead, in the white matter the water movement is allow in just one specific direction.

Diffusion MRI: Signal

The magnetic signal measured in the diffusion MRI is retrieved with addition of two gradient to spin echo sequence. The gradients that are applied are regulated from the *b – value* introduced the last time. The analytical expression of this component is the following one:

$$b = (\gamma\delta G)^2 \Delta t$$

From this expression we could understand that there some parameters on which we could act and other that we could not. Specifically, because γ represents the gyromagnetic factor we could not intervene on it, while on the value of δ (duration), G (gradient amplitude), Δt (time interval before the onset of the gradient). Thus, because there are some hardware limitations that block the maximum gradient G applicable the only thing that we could manipulate is the duration δ .

The electromagnetic radiation detected is strictly correlated with the *b – value*. Especially, with the movement of the different nuclei. If they are still in the same position during all the acquisition and stimulation phase the signal acquired present a dephasing content equal to zero. Instead if the nuclei are in motion the dephasing content of the signal is different from zero. These two conditions could be obtained from the relation between the signal dephasing and the position of the nuclei.

$$\Delta\phi = \phi_1 - \phi_2 = \gamma\delta G(x_1 - x_2)$$

Still $x_1 = x_2$
 $\Delta\phi = 0$

Movement $x_1 \neq x_2$
 $\Delta\phi \neq 0$

In general the acquisition sequence uses an interval of time stimulation equal to Δ which is substituted over the Δt quantity. Thus the relation becomes:

$$\Delta = \Delta t - \delta/3 \Rightarrow b = (\gamma\delta G)^2 \Delta$$

In fact, inside the real application what we change is the value of Δ and not the value of δ . This because if the subject undergoes for a long time of a gradient magnetic stimulation it could incur in some consequence that damage its health. Thus, the interstimulus time is much easier to control and for this reason this is the real variable finally controlled. The diffusion signal relay also on the next sequence of equivalence:

$$-bD = -(\gamma\delta G)^2 \left(\Delta t - \frac{\delta}{3} \right) D = \ln \left(\frac{S}{S_0} \right)$$

In this case we could see that if we increase the value of b by changing the Δ term we increase also the weight between the signal S and the signal S_0 . Notice that this sequence of equation is obtained by manipulating the signal coming from the scanner that connect mathematically the b – *value* and the diffusion coefficient D . In fact, S_0 represents the reference signal that is measured inside the magnetic resonance instrumentation without the application of particular b – *value*. Moreover if we increase the value of b we decrease the *SNR*. This because in the relation between b and the ratio S/S_0 there is a minus sign. Another effect that we could observe if the b – *value* increases is that the diffusion time increases too and thus we could obtain information more related with slow diffusion movement.

Diffusion MRI: Apparent Diffusion Coefficient

Let consider now an isotropic tissue as instance the grey matter inside the brain. In this case we have seen in the last lesson that the diffusion coefficient D is equal over the main three direction x, y, z . In this case with the diffusion MRI we could estimate the *Apparent Diffusion Coefficient* known as *ADC*. In the following paragraph we will see on what this concept relay on.

From a mathematical point of view we could describe the relation between the diffusion coefficient and the two signals S and S_0 with the next analytical expression.

$$-bD = \ln \frac{S}{S_0}$$

Notice that the relation between the logarithm of the ratio define as the signal S over the reference signal S_0 is proportional apart from the minus term of the b – *value* choose. Thus, if we plot the previous logarithm over the value of b we need to obtain, by varying this last quantity a straight line with negative slope.

$$\ln \frac{S}{S_0} = -bD \Rightarrow y = -Dx \Rightarrow \text{Slope: } -D$$

Thus, if we have the measure of S_0 and the measure of S at a given value of b the previous problem is uniquely determined, and we could easily estimate the value of the slope D of the straight line with a linear regression. The same analysis could be done by starting from a different initial point. In this case we need to start from the relation between the value of S and the value of b in the considered voxel. Hence, if we take the logarithm to both side of the equation we obtain:

$$S = S_0 e^{-bD} \Leftrightarrow \ln S = \ln S_0 - bD$$

In the previous relation we know the value of $\ln(S_0)$ and also the value of b . Hence, the only thing needed to be estimated is D . This estimation could be solved by using a linear estimator. In practice in each of the two representation proposed the measures that are considered are not only two but are more because the reliability of the estimation of just one measure is not the best one that we could obtain. However, if we would like to understand the minimum number of measures that we need to solve the previous problem is just two. One for the reference signal or an acquisition without the application of any gradient and one measure of the whole volume with a specific gradient applied.

Inside the previous analysis and estimation we have discussed on how we could estimate the value of D in a situation where the water molecules present an isotropic movement. In particular, inside a considered voxel there is not only grey matter but also a little portion of white matter and some other cell types. Hence, the diffusion coefficient that it is estimated through this analysis is not the real value of the diffusion coefficient, but it is an apparent value or an average value over the considered voxel. Therefore the diffusion coefficient is called the *Apparent Diffusion Coefficient* or *ADC*.

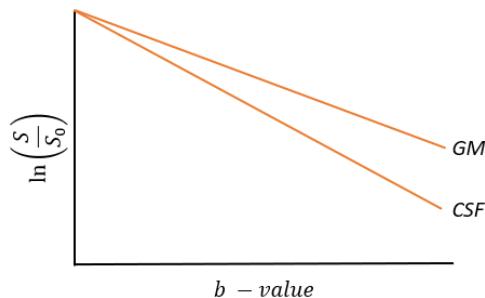
With the hypothesis that diffusion time are short, exchange between voxels can be considered negligible and the attenuation of the measured signal is given by the next relation:

$$\frac{S}{S_0} = \sum_i p_i e^{-bD_i}$$

Where D_i and p_i are the diffusion coefficient and the molecules fraction diffuses into the i -th voxel. In this case the ideal approach would be to separate the several D_i performing a multi exponential fitting. Unfortunately, D_i values are often not high and not different one from each other. Therefore, it is preferable to estimate the value of the *Apparent Diffusion Coefficient* as previously said:

$$\frac{S}{S_0} = e^{-ADCb} \Rightarrow \ln\left(\frac{S}{S_0}\right) = -ADCb \Rightarrow y = -ADCx$$

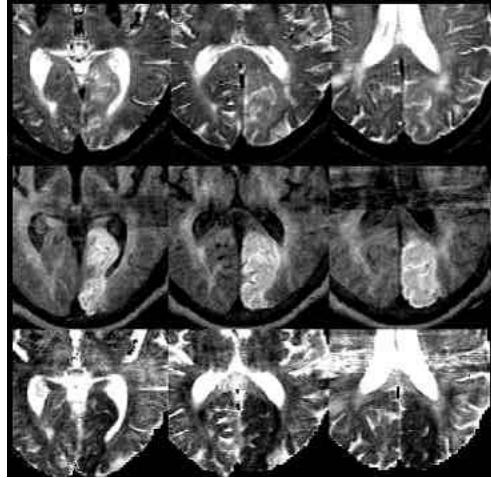
This coefficient is the most used in the literature and represent only the average value of the diffusion coefficient inside the analysed voxel. This represents the fundamental relation to measure diffusion with MRI techniques. Specifically, it has been observed that between the grey matter (GM) and the cerebral spinal fluid (CSF) the two straight lines identified are the next ones:



With this analysis and estimation processes we could perform different acquisition of our brain volume at different b -value. This leads to the generation of different images. Because of the signal acquired is related with b -value and this last one with the Δ term. Hence, higher is Δ , higher is b , higher is the time interval between two acquisition and thus the final acquired signal is more related with slow movement. What we could appreciate inside our analysis is that higher is the value of b darker is the final image.

In general once measured and acquired the different images at different b -values we could compute the *ADC map* involving in the estimation process all the information available. Down below is reported a comparison between a T2w images, the diffusion MRI images and the ADC maps post computed. In this case

we could clearly see the presence of three slices (columns) where there is enhanced with the ADC map in the right occipital region the presence of a higher diffusion coefficient due to its darker signal.



Diffusion MRI: Diffusion Tensor Imaging

If the tissue is isotropic as instance the grey matter only the *ADC* is sufficient to describe the characteristic of the tissue (mean diffusivity). Instead if the tissue is anisotropic as instance the white matter, the *ADC* is not enough informative hence we need to introduce the *Diffusion Tensor Imaging* known as *DTI* technique. In this case, it is removed the constraint that a specific voxel contains just grey matter, and the overall diffusion coefficient could be described with the *ADC*. With this technique instead of a single value is estimated a whole tensor. With this approach we still assume that we have random walk inside protons. Down below is reported the mathematical relation of the signal measured inside an anisotropic condition.

$$S = S_0 e^{-\sqrt{b}D\sqrt{b^T}}$$

In this case the value of D and b are not just scalar but they are represented with respectively a vector and a tensor. In fact, when we have introduced the acquired signal from the MRI scanner, we have just discussed it along the x direction. This is a generalization of the previous analysis and the most of time it is also more informative because we could compute again the *ADC* coefficient from the *DTI* imaging.

The value of S_0 , also in this case, represents the MRI signal application without the gradient used to get the diffusion tensor imaging. For this reason the formula needs to consider some scalar product between the vector b and the tensor D . We can move all out from the exponent of the previous equation with the computation of the logarithm of the ratio between the signal S and the signal S_0 . Hence the mathematical representation is the next one:

$$\ln\left(\frac{S}{S_0}\right) = - \begin{bmatrix} \sqrt{b_x} & \sqrt{b_y} & \sqrt{b_z} \end{bmatrix} \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yz} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \begin{bmatrix} \sqrt{b_x} \\ \sqrt{b_y} \\ \sqrt{b_z} \end{bmatrix}$$

After some algebra calculation and by exploiting that the diffusion tensor D is symmetric we could obtain the next relation:

$$-\ln\left(\frac{S}{S_0}\right) = D_{xx}b_x + D_{yy}b_y + D_{zz}b_z + 2D_{xy}\sqrt{b_x}\sqrt{b_y} + 2D_{xz}\sqrt{b_x}\sqrt{b_z} + 2D_{yz}\sqrt{b_y}\sqrt{b_z}$$

If from above, we define the following vectors:

$$\bar{b} = [b_x, b_y, b_z, \sqrt{b_x}\sqrt{b_y}, \sqrt{b_x}\sqrt{b_z}, \sqrt{b_y}\sqrt{b_z}]$$

$$\bar{D} = [D_{xx}, D_{yy}, D_{zz}, 2D_{xy}, 2D_{xz}, 2D_{yz}]^T$$

Inside a specific analysis the paradigm requires to set the value of the vector b along a specific direction and then solve the problem with a linear estimator to get information about the tissue. With the previous relation and representation of the vector b and D we could obtain the next mathematical expression.

$$\ln\left(\frac{S}{S_0}\right) = A = -\bar{b}\bar{D}$$

Remember that for each vector b we could obtain just one single image of our brain. Let consider now the following main direction $[1,0,0]$, $[0,1,0]$, $[0,0,1]$ which are the canonical vector of the orthogonal cartesian plane. Along these three directions we have oriented and applied a gradient. We would like to describe the situation mathematically.

$$\begin{aligned} \ln\left(\frac{S_1/S_0}{S_2/S_0}\right) &= -\sqrt{b} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yz} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \sqrt{b} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= -b \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = -b \begin{bmatrix} D_{xx} \\ D_{yy} \\ D_{zz} \end{bmatrix} \\ -\frac{1}{b} \begin{bmatrix} \ln(S_1/S_0) \\ \ln(S_2/S_0) \\ \ln(S_3/S_0) \end{bmatrix} &= \begin{bmatrix} D_{xx} \\ D_{yy} \\ D_{zz} \end{bmatrix} \end{aligned}$$

The information that could be retrieved in this case are just information about the diagonal terms. In fact, during the different acquisition we have fixed the value of b but we are only changing the direction of this. The pitfall of this analysis is that we could not obtain information about the off – diagonal term. For this reason we need to add some other direction to improve the analysis performed. This could be seen from the relation reported below and in the next page.

$$\ln \begin{pmatrix} S_1/S_0 \\ S_2/S_0 \\ S_3/S_0 \\ S_4/S_0 \\ S_5/S_0 \\ S_6/S_0 \end{pmatrix} = -b \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & 1 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix}$$

By solving the previous relation we could get all the necessary information for our diffusion matrix. Hence, from the previous example we could understand that to determine completely the diffusion tensor we need to analyse and apply different b vector along the three main principals non colinear directions. This means that the different b vector applied are not random but are directed through specific directions. If we consider a higher number than 6 main direction, we could obtain information on the diffusion coefficient along that specific additional directions.

From a practical point of view the tensor D is known as a symmetric matrix with 6 elements and to estimate all of them we need to acquire six distinct signals along 6 not colinear directions and obviously with a non – diffusion weighted image. It is usually a good practice to use more than 6 direction to uniformly sample the three – dimensional space.

Let come back to the main mathematical relation reported below:

$$\ln(S/S_0) = A = -\bar{b}\bar{D}$$

If \bar{b} is squared than the value of D could be easily calculated with the inverse relation: $D = -\bar{b}^{-1}A$.

In general \bar{b} is not squared and thus to solve the problem we could apply a linear least square estimator where \bar{D} is obtain through the relation:

$$D = -(\bar{b}^T \bar{b})^{-1} \bar{b}^T A$$

If we want to introduce information on the measurement errors, we need to implement them inside the estimator structure and for this reason it is used the weighted linear least square, where the measurement error information is summarized inside the matrix S which contains the inverse of the variance of the measurement error.

$$D = (\bar{b}^T S^{-1} \bar{b})^{-1} \bar{b}^T S^{-1} A$$

Moreover if we would like to consider the data S without computing the ratio with S_0 and taken the logarithm we could use some non – linear weighted or non – weighted least squares.

Once D is correctly estimated we could compute the eigenvalues and eigenvectors of the matrix. In particular with these knowledge we could describe the overall diffusion process with just information on these three preferential directions. To introduce the different metrics that could be computed for each voxel once the tensor is completely identified we need to define the symbols that we are going to use. These are v_1, v_2, v_3 for the three main eigenvectors and λ_1, λ_2 and finally λ_3 for the three main eigenvalues.

If the three distinct eigenvalues estimated are equal the diffusion could be represented with a sphere. In this case the diffusion is representative of an isotropic tissue. Instead, if the three main eigenvalues are different, we are inside an anisotropic condition.

In general we could perform some particular analysis that describe the diffusion through specific coefficient. These are the one listed below:

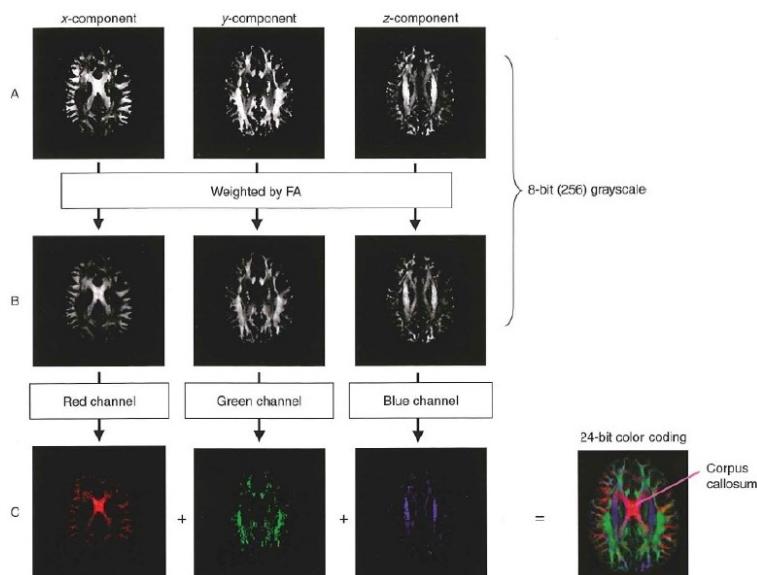
- *Longitudinal diffusion: LD = λ_1*
- *Fractional Anisotropy: FA = $\sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}$*
- *Radial diffusion: RD = $(\lambda_2 + \lambda_3)/2$*
- *Mean diffusivity: MD = $(\lambda_1 + \lambda_2 + \lambda_3)/3$*
- *Mean diffusivity: MD = $(\lambda_1 + \lambda_2 + \lambda_3)/3$*

The *Fractional Anisotropy* is the best index to represent the anisotropy characteristic movement during the analysis of the diffusion of different particles. Specifically, it is one of the best indices that could be used to describe the impact of a certain pathology inside a given specific tissue. The *mean diffusivity* instead calculates a diffusion coefficient like if the diffusion movement happens inside an isotropic tissue.

In general it is needed to perform the eigenvalues and eigenvectors decomposition because we would like to obtain some base that are invariant to rotations and translations.

Before we have analysed, the possible impact of the eigenvalues on some indexes that are needed to describe the diffusive movement. However we have not still analysed the meaning of the eigenvectors. If the eigenvalues represent the diffusion coefficients, we are finding with the different eigenvalues the rotation and inclination respect to the initial reference system that have the diffusion surface. Hence the eigen vector brings information on the orientation of the diffusive movement.

From the practical point of view only the first eigenvector is considered because describes the principal direction of the flow of the water molecules and inside those voxels of white matter just the main eigenvector is relevant. All this information could be combined together and computed a final colormap to detect the different fibers inside the final image.



LESSON 17: TRACTOGRAPHY

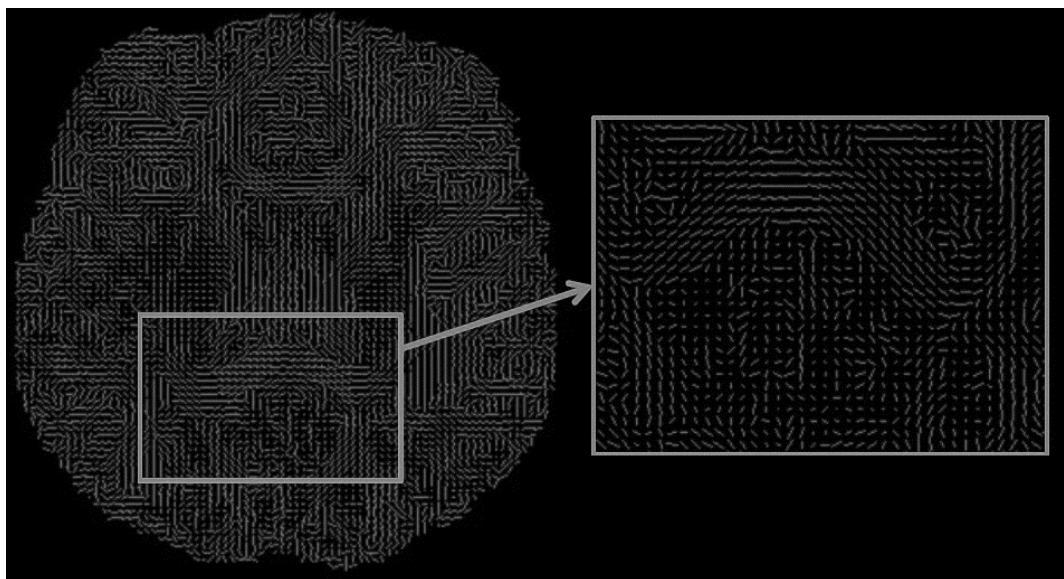
Today we will talk about the fibers tracking. The model that could be used are a lot and someone are very complex. We will see and deal with the principal tractography methods.

Tractography: DTI Fiber tracking

We move from the previous lesson where we have talked about the different models used to estimate the diffusion coefficient as a number or as a tensor and we pass to deal with the fibers tracking. The first approach to the construction of the axon bundles is derived from the diffusion tensor model.

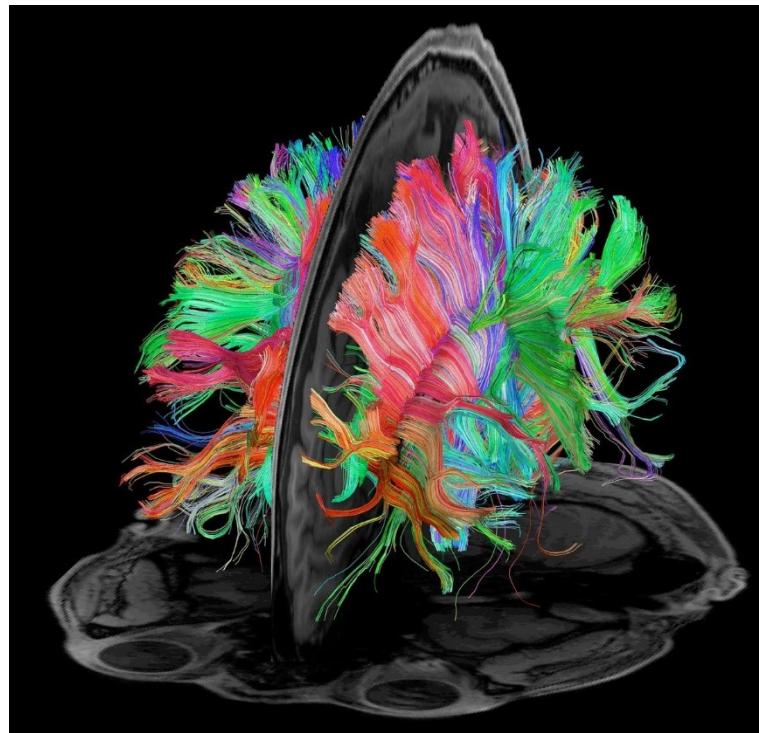
Diffusion principal directions can be used in each voxel to visualize the directional information about axonal fibers. We have seen in the previous lesson the specific methods with which we could combine the eigenvectors and in particular the direction that could be assumed from the ellipsoids by considering just the first of the main eigenvector. Especially, we could determine the three mains direction of the axons inside the brain by combining information coming from the main eigenvector and the fractional anisotropy to compute three different brain maps representative of the three main direction. From a practical point of view this information contained inside the maps are then represented with different colours and finally combined.

To do so, we can draw maps with each voxel having a vector representing such a direction. This could be seen in the next image and in its enlargement. In this case we could distinguish between the dots and oriented arrows. In the first case the flow of the information is directed through the direction of the arrow, instead with the dot this means that the flow is equally distributed among the three main direction. Thus in the first case the voxel is mainly made up of white matter while in the second case the voxel is mainly made up of grey matter.



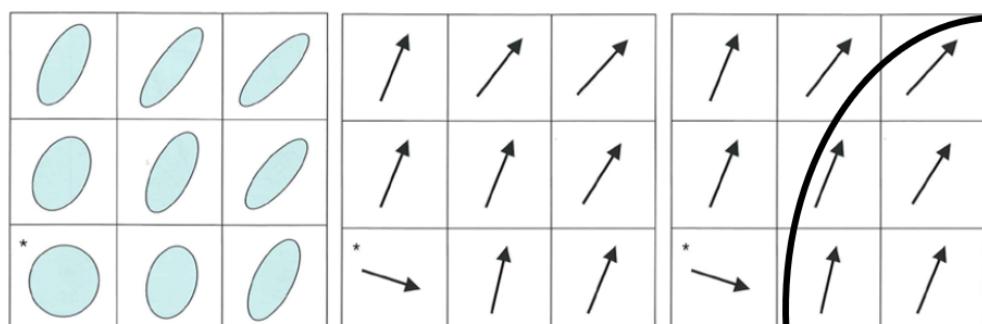
Representing voxel by voxel with the directional information we could appreciate a structure. An alternative way to visualize it, is to represent the entire diffusion ellipsoid which is related with the diffusion tensor. By using this information we need to use some method to link the different vector that follows the same flow.

These methods need to work in a three – dimensional space and not only a bidimensional space. This will give us a continuous representation of the situation. Some possible results by using these algorithms are the images reported below:



In this case we could see the presence of the corpus callosum. The information, that we could get by arriving to the previous representation is very consistent because we could reconstruct the fibers connection without applying any cerebral biopsies or any invasive technique. With these images we could detect the direction of the fibers, and this is an important information because it is not easy by just looking on images the real connection of the brain and its axonal organization without any invasive measurements. The Diffusion approaches could give us a lot of these information both *in vivo* and in clinical trials with different degree of accuracies and by focalizing to different spatial scales.

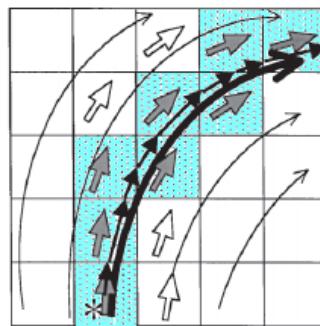
In the tractography methods we have the next representation for the diffusion coefficient and its directionality information. Now we need some pragmatical information and methodology to link the different voxels that follows the same fibers. In the image reported below we could see the presence in the left lower corner of an isotropic diffusion coefficient representative of the situation. While over the main diagonal (the squares between the left lower corner and the right upper corner) the flow of the information presents a higher directionality characteristic.



From the first image reported on the left we could move to the image reported on the right where it is represented the same matrix considering the directionality information and in particular the direction of the flow. In this case the different arrows also present a final arrowhead, but this does not carry any additional information. In the down left voxel there is any information to print the direction of the flow with a specific orientation and therefore we could use this representation.

Tractography: Deterministic approach

The fundamental assumption common to deterministic methods is the equality between the voxel wise principal diffusion eigenvector and the line tangent to the axonal fibers belonging to the same voxel. The important things is that when we detect a specific direction we could not distinguish between the different axons but just a bundle of axons. This method is known as deterministic tractography and was the first method introduced to give a solution to this problem. The axonal fiber is represented as a smooth line known as *streamline* where in the figure below is represented with the continuous black big arrow. This is originated form the centre of a voxel and with its direction equal to the one of the principal eigenvectors. Such line then changes its direction when it passes through a different voxel. Different interpolation techniques can then be applied to ensure this line is neither segmented nor presents non – physiological curvature.



Tractography: Deterministic approach – Streamline

The first thing that we need to perform before going to apply a deterministic tractography and in general a tractography analysis is the estimation of the diffusion tensor for each voxel and then compute for each tensor its eigenvalues and eigenvectors. By using this information we could generate the streamline. This is based on the mathematical relation between the principal diffusion eigenvector and the line tangent to the axonal fibers. The axonal fibers, represented as a continuous streamline, can be described as a curve in a three – dimensional space, $r(s)$, parametrized by its arc length s along its trajectory. Following the Frenet equation, given a curve $r(s)$ its evolution along its length is described by the following differential equation:

$$\frac{dr(s)}{ds} = t(s)$$

Where with $t(s)$ we are representing the unitary tangent vector to $r(s)$ evaluated in s . Thus from the previous differential equations we need to derive the description of $r(s)$. Especially we know the value of the tangent to $r(s)$ in s thanks to the information contained inside the main eigenvector. As seen before, we assume that the tangent direction to the streamline is represented by that of the voxel wise diffusion direction:

$$\frac{dr(s)}{ds} = \epsilon_1[r(s)]$$

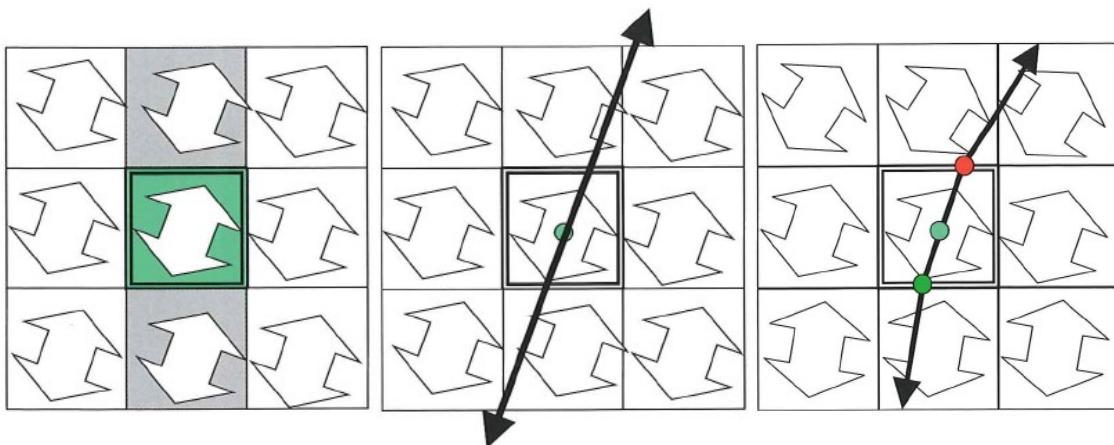
Where ϵ represents the major eigenvector in the specified point $r(s)$. Such equation can then be resolved with numerical methods by specifying the initial condition $r(0) = r_0$, which represents the starting point of the streamline, which we call the *starting seed*. The initial condition is the reason why we have cited before that for the deterministic tractography need some starting point from where to start the evolution of the system. We will generate the streamline related to the specific starting seed and then we could see the stability of the solution by moving the starting point.

The differential equation is often resolved by using either the Euler's method, or more advanced one such as the Runge – Kutta's method. Before considering the solution, though, it must be kept in mind that the streamline s defined in the continuous Cartesian Space, while data are acquired on a (x, y, z) discrete grid. In particular, we have a single measurement for each voxel.

We can for example hypothesize that the voxel wise value is the same across the entire Cartesian space occupied by the voxel. This method has an error propagation which most likely disrupts the fibers reconstruction. This is due to the multiple presence in the same voxel of different tissue types.

Another thing and alternative that we could consider is to perform a smooth interpolation, where we allow a variable contribution from neighbour's voxels when deciding the streamline direction.

The next example is the best one to clarify the concept of fibers tracking with deterministic tractography. In this case, the first things that we could appreciate is the presence of bidirectionality between the different principal eigenvector computed. In the image on the left, we could think to solve the previous differential equation by solving this by looking on the voxel level. In the central case we could see the presence of a non-weighted condition in which the direction of the fiber could be clearly seen. Instead, in the opposite situation or by using the assumption to weight the influences of neighbour voxel to compute the final streamline to solve the differential equation we could think to subdivide the voxel into sub voxel. Specifically in this case the number of sub voxels are three. One before entering in the middle area, the middle area and finally the region above the middle area. In this case the eigenvector that needed to be used in centre area is the same obtained from the analysis and the solution of the differential equation. While if we move close to the border, we are close to the previous and the next voxel. The eigenvector in this position is the weighted version between the two successive voxels. The weighting process is not fixed and could be changed by incorporating information coming from generalized neighbour voxel of a considered one.



To conclude we could solve the differential equation by using only information coming from a single voxel or by subdividing the voxel in smaller subunit and considered in each one information which are a weighted version of the neighbour subunits of the one considered. In this last case the direction of the streamline is not constant inside the same voxel but varies through the voxel itself.

Tractography: Deterministic approach – Streamline: Three dimensional case

At each step, we spatially evaluate the coordinates $[x_r, y_r, z_r]$ of the point $r(s)$ which our streamline has reached during its propagation. Therefore, we consider the 26 adjacent voxels, thus evaluating the next direction on a cube $3 \times 3 \times 3$ voxels (27 elements in total). This could be done since the fiber tracking would like to reconstruct a whole fibers with a three – dimensional representation.

We then associate a weight to each of the 27 voxel wise tensors which is linked to relative distance between the streamline point $r(s)$ and the centre of the voxel itself. The interpolated tensor is the result of the weighted average of the 27 tensors.

$$D_i = \frac{\sum_{j=1}^{27} p_j D_j}{\sum_{j=1}^{27} p_j}$$

Every time we change the considered voxel and the position under analysis, we could compute the weight version of the diffusion tensor that needed to be used to compute and solve the previous differential equation. Moreover, this will be the tensor on which all useful parameters will be computed for the tractography algorithm and thus the computation of the eigenvectors, eigenvalues, and the fractional anisotropy coefficient.

The differential equation is often resolve by using either the Euler methods or more advanced one such as the Runge Kutta's as previously seen.

Tractography: Deterministic approach – Streamline: Euler forward method

Starting from the seed point, we evaluate the diffusion tensor in that position and its principal direction. Once chosen an integration step α , which as a rule of thumb should be inferior to half the dimension of the voxel, the update rule is:

$$\begin{aligned} r(s_1) &= r(s_0) + \alpha \epsilon_1 [r(s_0)] \\ r(s_2) &= r(s_1) + \alpha \epsilon_1 [r(s_1)] \\ &\vdots \\ r(s_{i+1}) &= r(s_i) + \alpha \epsilon_1 [r(s_i)] \end{aligned}$$

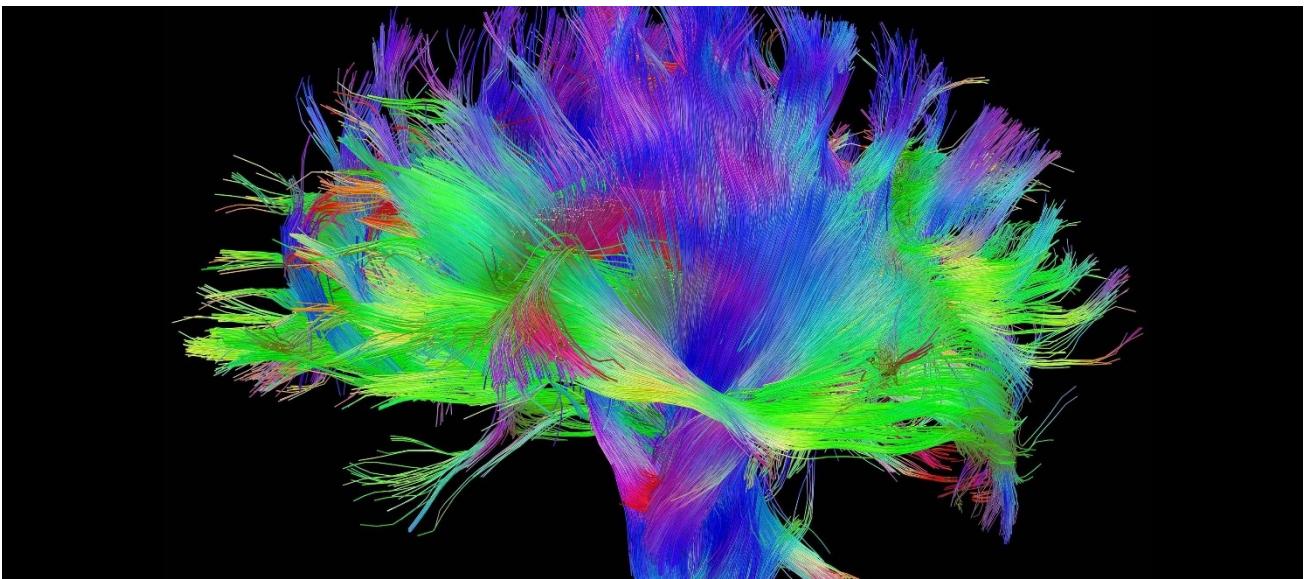
The propagation direction of the streamline is then defined, at each point, from the principal eigen value computed after the tensor interpolation principle proposed before. This could be repeated for each seed.

Tractography: Deterministic approach – Streamline: Stopping criterion

The problem is how to stop the algorithm. Because we have decided how we could start the algorithm and how we could progress with it. This method stops if:

- The streamline curvature exceeds a fixed threshold. In fact if the threshold is of about 40° with an integration step α equal to 0.4 voxel means that we are allowing directions changes with a maximum of 40° in the arc with a length of 0.4 voxels. Usually these thresholds are set up to enhance some physiological constraints.
- The anisotropy (FA – it's the one of the interpolated tensors) is lower than a fixed threshold, for which we assume there is no preferential direction for the water molecules. In literature the FA is of about 0.18. This value allows for the distinction between the anisotropic white matter and the isotropic grey matter.

What we obtain by using this algorithm is as instance the next results. Obviously, the next results is obtained by fixing a seed for each fiber. Once all the streamlines are estimated we could give them a specific colour maps and we could reach some specific tractography representation.



There are some limitations is using the principal eigenvalue as the indication for the fiber direction.

- The noise of the diffusion image affects the direction of the principal eigenvector, which results in a directional error which propagates along the fiber reconstruction.
- As FA goes down, the uncertainty about the estimation of the principal eigenvector goes up. Thus when we approach an isotropic environment the estimates are affected more from bias than the anisotropic environment. This fact may have an impact in erroneously tracing the fiber in zones where is no strong directional indication. An example where this may happen is the thalamus, the arcuate fasciculus, and the subcortical white matter.
- Partial volume effects may influence the tracking.

Tractography: Tensor deflection

As the name implies, the tensor deflection technique estimates the propagation direction by using the entire Diffusion Tensor, rather than just identifying it with only the principal eigenvector. The output direction v_{out} will then be computed as the deflection on the entering direction v_{in} operated by the tensor D .

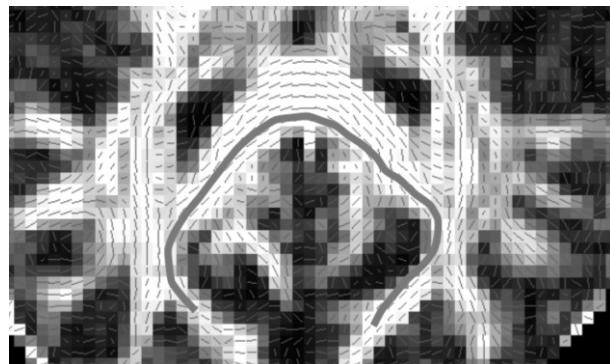
$$v_{out} = Dv_{in}$$

Thus we can link the input and the output direction from the voxel. Higher is the anisotropy higher the deflection will potentially be. If D is isotropic, the deflection will be close to zero. This algorithm uses the same previous stopping criterion that we have seen based on the fiber maximum deflection and the maximum fractional anisotropy index admissible.

If in the same voxel we have fibers with different orientations, for example in cases of crossing fibers and/or high curvature where the algorithm stops, this technique fails as the previous one. Specifically if we have a situation like the one reported below where there is a cross between different bundles along perpendicular direction. This will be considered inside the analysis as an isotropic diffusion along that specific voxel. Thus, it may be necessary to use local reconstruction methods which go beyond the diffusion tensor to reconstruct the true path of the axonal fiber.



This was one of the first big problem that has been detected when using the deterministic approaches. We are able to define the well define path by using the diffusion deflection as in the situation reported below:



But when we have complex pattern exempli gratia crossing, we are not able more to identify and solve the previous problem. In particular when fibers cross each other, in voxel containing two crossing bundles, the tensor ellipsoid is pancake shaped known as *oblate* or *planar* tensor. In voxels containing three crossing bundles, the tensor ellipsoid is spherical. Down below are reported the main shapes.



Spherical Tensor



Oblate Tensor



Prolate Tensor

In particular in these areas, the DTI is meaningless, and the FA evaluated is biased. In the previous images is reported also a prolate situation where is represented the real ellipsoid surface descriptive of a singular fiber inside the analysed voxel.

Tractography: Ball and Sticks model

To solve this problem we use and introduce a different model to link the different information. We think to find a better and complex model to describe what happens inside the brain. The ball and sticks model relay on the concept that the tissue that constitute the considered voxel is made up of a percentage of isotropic and anisotropic tissue. From the mathematical point of view this could be written as:

$$Tissue = (1 - f)\% \text{ isotropic} + f\% \text{ anisotropic}$$

Thus we can describe the measured signal for the j -th gradient applied as the next analytical expression:

$$s_j = s_0 \left[(1 - f) \exp(-b_j d) + f \exp \left(-b_j d (x_j^T v)^2 \right) \right]$$

Where $(1 - f)$ represents the fraction of anisotropic volume inside the j -th voxel, d represents the diffusivity, x_j represents the unit vector which identify the direction of the gradient j and finally v which is the fiber orientation. The first term of the sum contains the information of the isotropic tissue while the second the information coming from the fraction f inside the voxel of the anisotropic tissue.

We can make even more complex the model by considering the subdivision of the given volume. In this case the signal s_j could be represented with the next analytical expression:

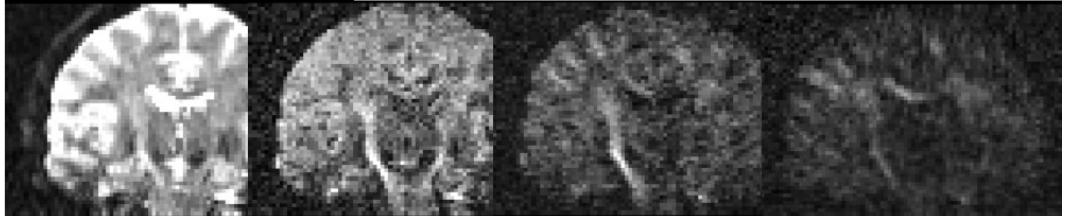
$$s_j = s_0 \left[(1 - \sum f_n) \exp(-b_j d) + \sum f_n \exp \left(-b_j d (x_j^T v_n)^2 \right) \right]$$

Where in this case s_j represents the signal measured inside the j -th voxel while f_n represents the anisotropic volume fraction inside the n -th subdivision of the considered voxel. In this case n represents the maximum number of sticks that could be used inside the ball and sticks model.

With these types of model to correctly estimate all the possible parameters we need to have a lot of data. In general are necessary more measure of s instead of its basal condition s_0 . Specifically, the gradient applied, or the b -values applied could assume value inside a given set of elements. We are switching from a mono shell approaches where we are changing only the direction of application of the b -value to a multi shell approaches where we change the magnitude of the b -value too and not only the direction. As instance we start from the value of b equal to 1000 s/mm^2 and we acquire as instance 32 principal direction. Then we change the value of b and we perform the same acquisition along the same 32 principal direction identified.

We need to have more acquisition to estimate all the parameters in the previous mathematical description of the ball stick model. One important thing that we need to remember is that higher is the value of b higher is the angular contrast. Thus, with the measurements we could resolve the previous problem of crossing

fibers. Of course as we improve the b value lower is the SNR finally obtained as could be seen in the images reported below. From the left to the right the value of b applied are 300, 1000, 2000 and 3000 s/mm^2 .



Tractography: Multishell diffusion DTI

Nowadays the multishell diffusion DTI is largely used and the model based is much more complex than the one seen before with the ball and stick applied on a parcelled voxel. These are as instance the:

- *CHARMED (Composite Hindered And Restricted Model of Diffusion)*. This protocol needs to have information along 63 different directions and 5 b values such as 82, 413, 1150, 2250 and 4000 s/mm^2 . The high number considered is because the number of parameters that we need to estimate is very high. This means that we need to acquire 70 volumes in total including the 7 b_0 and the 63 diffusion weighted images.
- *NODDI (Neurite Orientation Dispersion and Density Imaging)*. This is one of the methods nowadays mostly used. It is the simplification of the charmed paradigm because it estimates just the macro parameters. In this case the protocol needs to consider the 34 direction, 2 b value. The optimized protocol considered just 30 direction with a value of b equal to 700 s/mm^2 , 60 direction with b equal to 2800 s/mm^2 and 9 b_0 images.

In all the two previous mentioned methods the acquisition of multiple value of b_0 is necessary because after all we could compute the average between all the images acquired and we could get a better SNR of the final results. In the following paragraph we will go deeper inside the NODDI algorithm and we will do just a little overview of the CHARMED one.

Tractography: NODDI

The idea under the NODDI is that we could split the isotropic and the anisotropic information inside a given voxel. It is a little bit different the subdivision between the isotropic and the anisotropic information. This approach would like to consider the presence inside the voxel of three types of tissue. These are the next three microstructural environments. The first one is the intra cellular environment, the second the extra cellular environment and finally the cerebral spinal fluid. The full signals S and could be written as:

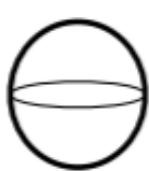
$$S = (1 - \nu_{iso})[\nu_{ic}S_{ic} + (1 - \nu_{ic})S_{ec}] + \nu_{iso}S_{iso}$$

Where the terms that made up this equation have the following meaning:

- S_{ic}, ν_{ic} : These two terms refer to the intra cellular microstructural environment.
- S_{ec} : This term represents the extra cellular structure.
- S_{iso}, ν_{iso} : These two terms represent the presence of cerebral spinal fluid.

The first part of the tissue represents the portion on the anisotropic information while the second represents the isotropic tissue. Each compound affects water diffusion within the environment in a unique way and gives rise to a separate normalized MR signal.

The three different environments used in the model are represented in graphical term with the next geometrical shapes to describe the diffusion movement inside that specific environment.



Cerebral Spinal Fluid



Extracellular environment



Intracellular environment

An example of how we can move from the graphical representation to the mathematical formulation we need to follow some of the down below reported example. Inside the NODDI algorithm we have to remember that are largely used the *stick*, the *ball* and finally the *zeppelin*. The first one is used to model the intra cellular compound, the second the presence of CSF and finally the third to describes the extra cellular compartment.

Tractography: NODDI - Parameters

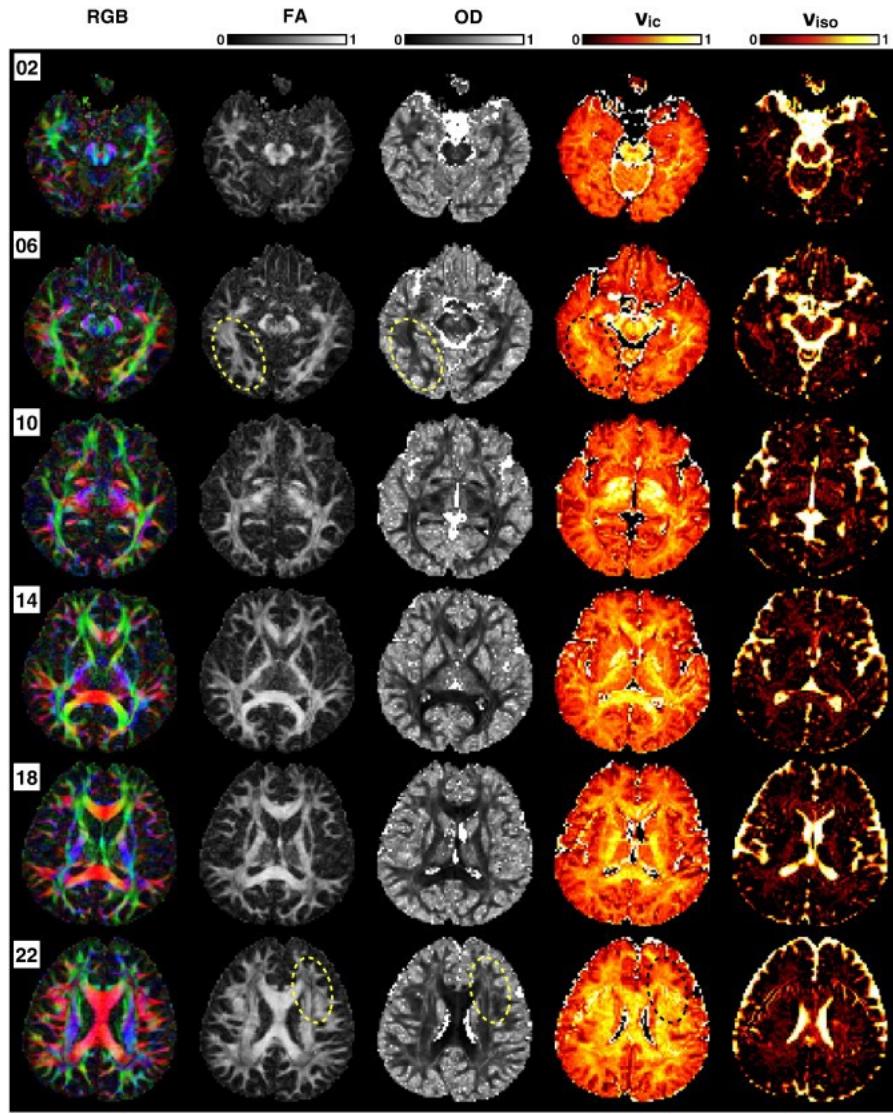
The complete set of parameters for the NODDI model we need to introduce the next terms:

- v_{ic} : Intra cellular volume fraction
- $d_{||}$: Intrinsic free diffusivity
- κ : Concentration parameter of Watson distribution
- μ : Mean orientation of Watson distribution
- v_{iso} : Isotropic volume fraction
- d_{iso} : Isotropic diffusivity

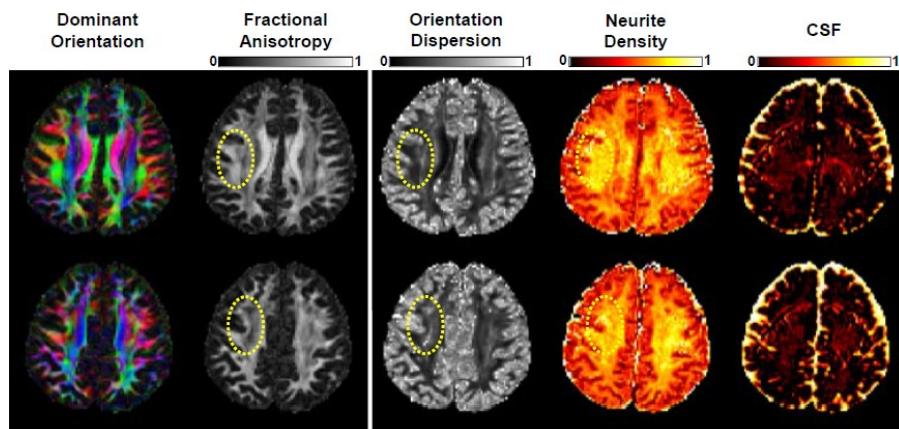
All the parameters do not have to be estimated because some are fixed to certain value that are found through specific experiment. Specifically the value of $d_{||} = 1.7 \times 10^{-3} \text{ mm}^2 \text{s}^{-1}$ and $d_{iso} = 3 \times 10^{-3} \text{ mm}^2 \text{s}^{-1}$. The remaining parameters needed to be estimated. We can think also to fix the mean orientation of Watson distribution by using also some a priori knowledge. The fitting routine and procedure determines the maximum likelihood estimates of the parameters, using a Rician noise. This last version of the measurement error is very close to the gaussian distribution.

Tractography: NODDI - Results

Putting together all this information what we can obtain are three different parameters that are so relevant to describe the brain anatomical and histological structures. We can estimate three different parameters, the κ which is used to compute the orientation dispersion and the v_{ic} and v_{iso} . The first one represents the intracellular volume fraction or the neurite density and the isotropic volume fraction. The parametric maps reported below are generated through the usage of NODDI. The pathological information could be found mainly in the v_{iso} map.



Now we need to create some algorithm that follows the orientation contained inside the orientation dispersion maps (OD maps). In the figure below we could see a pathological tissue mainly related inside the fractional anisotropy. The orientation dispersion maps contain different information from the fractional anisotropy. In fact, in this map we have some information about the quality of the pack of the bundles inside our brain.



Tractography: CHARMED

The CHARMED is a little bit more complex and we could get some other important information. In this case the number of compartments used are just two. The first represents the water diffusing inside the axons. Here we recognize the restricted compartment. Instead, in the second represents the water diffusing outside the axons. This is known as the hindered compartment.

Tractography: NODDI and CHARMED - Disadvantages

When we decide to add compartment to our analysis it is not always a good idea. This because we need more measures, and we have more parameters to estimate. We need to be sure that the addition of certain compartments is needed and necessary only when we could better underline the physiology of the pathology or the state of the patient. Moreover, sometimes linear fitting is often impossible because of the high nonlinearity inside the model adopt.

Tractography: Spherical deconvolution

Until here we are always used model driven approaches based also on the physiology of the body. Instead, now we would like to introduce a new method called the spherical deconvolution which is a data driven approaches and express the diffusion signal s by the modelling expression reported below:

$$s(\nu) = FOD(\nu) \otimes h(\nu)$$

Where in this equation we have the ν which represents the unit vector of the diffusion gradient direction, $h(\nu)$ represents the system impulsive response function which is known and finally what we would like to really estimate is the $FOD(\nu)$. This represents the weight of the fibers response along each direction. We need to have some parametric distribution of the FOD . This is not physiological based instead is a parametric deconvolution problem that could be simply estimated with generalized spherical deconvolution by using a Bayesian approaches.

We are not going in detail of this approach, but we are really moving from even difficult model driven approaches to the completely data driven. However, when we move to this last class of analysis, we need to remember that we have enriched the model just because there are some problems that needed to be resolved.

The choice one makes regarding the description of white matter fibers heavily affects the results one may obtain. Using a single dominant direction per voxel can somehow work in zones where anisotropy is high and fibers have high axonal diameter exempli gratia the corpus callosum but has been demonstrated to fail where the configurations are more complex.

By using more advanced reconstruction methods as instance multiple tensors, probabilistic methods, and others we can track multiple fibers going in different directions, all in the same voxel. This is impossible with DTI. Advanced techniques require extended datasets and lengthy acquisition times. This makes the clinical translation where time and cost are important variables of such methods highly impractical.

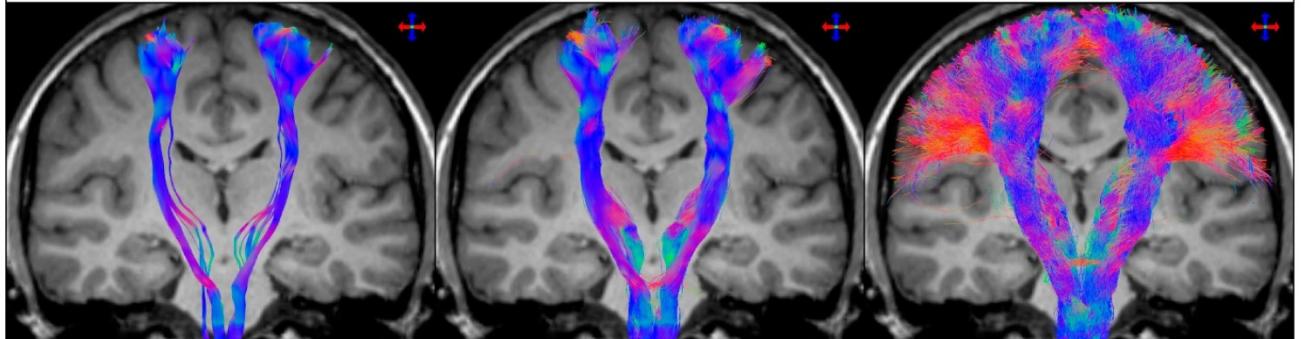
Tractography: Probabilistic approach

Using a probabilistic approach, the fiber reconstruction error can be computed by representing the estimation uncertainty as a function of the available diffusion directions inside the voxels. The possible presence of multiple fibers with different orientation is then explicit.

This approach allows to continue with the tracking even where the deterministic tractography is terminated, by simply associating a higher error to the reconstructed fibers.

The uncertainty associated to the multiple directions inside a white matter voxel can be represented by means of the so-called Orientation Distribution function (ODF) which can be quantified in many possible ways. As each MR image is affected by noise, the tensor estimation itself is affected by noise. Along the reconstructed fiber, the error is propagated leading to always increasing differences with the ground truth, and eventually to wrong cerebral pathways reconstruction. The tract directions can be represented with a cone whose aperture angle represents the uncertainty with which the directions are estimated (for example, we can represent the directional confidence intervals at 95%). The quantified uncertainty varies depending on the nature of the voxel (as a general rule, less anisotropy, more errors), and can be estimated by simulation or other statistical techniques.

By using this approach we could reach better results than the deterministic analysis. In fact, inside the images reported below we could see the clear difference between the Conventional Diffusion Tensor and the probabilistic tractography. We could see that the second approach seems to be superior because we can see that fiber reconstruction is much more complete.



*DTI and Deterministic
Algorithm*

*DTI and Probabilistic
Algorithm*

*CDS and Probabilistic
Algorithm*

```
1 %
2 % Title: Laboratory 7
3 %
4 % Note:
5 % dMRI Analysis and generation of the fiber reconstruction
6 %
7 % Author: Matteo Martin
8
9 % THEORY -----
10 %
11 % dMRI
12 % Today we discuss and see some of the possible data manipulation that
13 % needed to be done to achieve a good data visualization of some dMRI
14 % analysis. In particular we know that all the information that come from
15 % this technique are based on the movement of the different water
16 % molecules and how they could moves inside the tissue. In particular the
17 % analysis that could be done are of two types.
18 %
19 % dMRI - ISOTROPIC DIFFUSION
20 % The first approach is based on estimate the diffusion coefficient
21 % assuming that the water diffusion coefficient is equal inside the three
22 % main direction. This condition is known as isotropic analysis.
23 %
24 % dMRI - ANISOTROPIC DIFFUSION
25 % Beacuse of assuming that the different tissue structure that are
26 % available inside the brain is too a strong assumption considering the
27 % movement of the water molecules inside the three main direction equal.
28 % Thus beacuse as instance inside the white matter the flow of the water
29 % molecules follows the different fibers we could differentiate this
30 % analysis and introduce the anisotropic one. In this case instead of
31 % estiamting as the previous situation just one single value for all the
32 % direction we need to estimate an entire tensor.
33 %
34 % dMRI - CLINICAL APPLICATION
35 % This exam is very important beacuse enhance the possible modification of
36 % the diffusion coefficient inside the brian structure. The eventual
37 % modification could enhance pathological condition of which the patient is
38 % affected of.
39 %
40 % dMRI - DATA ORIGIN
41 % The origin of the data that comes from this type of MRI methodology is
42 % strictly related with a particular MRI sequence that needed to be
43 % applied. In particular we have that by applying addition gradient with
44 % specific direction we could excite the tissue along a specific direction
45 % and thus we could obtain information from it. The dMRI data are strictly
46 % related with the direction of the gradient applied.
47 %
48 % DATA -----
49 %
50 % dMRI_volumes: Inside this matrix are contained all the information coming
51 % from the MRI scans. The last size of the 4 Dimensional
52 % structure is equal to the size of the bvals and bvecs.
53 % bvals      : Data relative to the isotropic diffusion analysis.
54 % bvecs      : Data relative to the anisotropic diffusion analysis.
```

```
55
56 %% CCC
57 % Clear all, close all, clear the command window
58
59 close all; clear all; clc
60
61 %% DATA - LOAD
62 % Load of the data that are contained inside the folder data
63
64 load('dMRI_labData.mat')
65
66 dMRI = double(dMRI_volumes);
67 [nVOXX, nVOXY, nSLICE, nB] = size(dMRI);
68 clear dMRI_volumes;
69
70 %% SO - GENERATION
71 % Extraction of the reference volume activity detected with a value of b
72 % exactly equal to 0.
73
74 S0 = dMRI(:,:,:,:bvals == 0);
75
76 %% HISTOGRAM - VISUALIZATION
77 % Visualization of the histogram of the SO volume to choose the threshold.
78 % This last one will be used to obtain a mask used to identify those voxels
79 % with intensity greater than a threshold.
80
81 figure, histogram(S0), grid on, grid minor, box on,
82 title('HISTOGRAM - SO'), xlabel('Intensity []'), ylabel('A.F. []')
83
84 TH = 125;
85
86 figure, hold on, histogram(S0), xline(TH,'r--','LineWidth',2), hold off
87 grid on, grid minor, box on,
88 title('HISTOGRAM - SO'), xlabel('Intensity []'), ylabel('A.F. []')
89
90 %% MASK - GENERATION
91 % Generation of the mask to understand which are the voxels that we should
92 % need to consider in our analysis and where we should use the fit.
93
94 MASK = S0 > TH;
95
96 %% MASK - VISUALIZATION
97 % Visualization of the mask generated at the previous step by the usage of
98 % the implay function.
99
100 implay(MASK)
101
102 %% ISOTROPIC ANALYSIS - FIT
103 % Fit of the data by computing the logarithm of the ratio between the
104 % signal S and the signal SO. To correctly perform this analysis we need to
105 % extract first all the possible voxel that are active inside the mask
106 % matrix and then we need to extract all the possible values from the 4D
107 % object related with that voxel in the same position of the three
108 % dimensional object for each element in the fourth dimension. After that
```

```

109 % we compute the diffusion coefficient with a linear regression estimation.
110
111 S      = double(dMRI);
112 Slog = log(S./repmat(MASK.*S0,[1 1 1 nB]));
113 B      = bvals(2:nB)';
114 D      = zeros(nVOXX,nVOXY,nSLICE);
115
116 for idS = 1:1:nSLICE
117   for idx = 1:1:nVOXX
118     for idY = 1:1:nVOXY
119       if MASK(idX,idY,idS)
120         VoxSignal      = squeeze(Slog(idX,idY,idS,2:nB));
121         D(idX,idY,idS) = -inv(B'*B)*B'*VoxSignal;
122       end
123     end
124   end
125 end
126
127 % It is important to look at the contrast in the image. In the ventricles
128 % there is zero restriction to move and thus the signal intensity is very
129 % very high. The maximum value that we obtain from this analysis is higher
130 % than 3.5 mm/s. Thus the diffusion coefficient is much higher than the
131 % maximum physiological value of the diffusion coefficient of the water at
132 % 37°C. This is due to some non physiological artifact and the bias induced
133 % from the model
134
135 %% ISOTROPIC ANALYSIS - VISUALIZATION
136 % Visualization of the slice number 25 of the matrix D computed to check if
137 % the elaboration until this point is correct or not.
138
139 figure, imagesc(imrotate(D(:,:,25),90)),
140           colormap 'hot'; colorbar
141           title('SLICE 25 - ADC Distribution')
142
143 %% ANISOTROPIC ANALYSIS - B MATRIX
144 % First generation of the B matrix to correctly compute the fitting process
145 % by the usage of the linear regression for each voxel inside the matrix.
146
147 BV  = bvecs(:,1:nB);
148 B   = [BV.^2; BV(1,:).*BV(2,:); ...
149          BV(1,:).*BV(3,:); ...
150          BV(2,:).*BV(3,:)]';
151 B   = B.*bvals(1:nB)';
152
153 %% ANISOTROPIC ANALYSIS - FIT
154 % Evaluation of the diffusion tensor matrix to correctly perform the
155 % anisotropic analysis. In this case we need to first extract the voxel of
156 % the mask that are not related with infinite values, then we need to
157 % compute with a linear regression the whole possible element of the
158 % diffusion tensor. These elements need to be reorganize and identify the
159 % diffusion tensor. We compute then the eigen values and eigen vector of
160 % the same tensor and we take the absolute value of the negative components
161 % if all the element are negative but if just someone are negative we put
162 % them to zero. We sort the eigenvalues inside the vector and the eigen

```

```

163 % vector too. The next step consists in computing the fractional anisotropy
164 % coefficient for each voxel, the mean diffusivity and a 4 dimensional
165 % vector in which in each fourth dimension is stored a specific directional
166 % information.
167
168 FA = zeros(nVOXX,nVOXY,nSLICE);
169 MD = zeros(nVOXX,nVOXY,nSLICE);
170 V = zeros(nVOXX,nVOXY,nSLICE,3);
171
172 for idS = 1:1:nSLICE
173     for idx = 1:1:nVOXX
174         for idY = 1:1:nVOXY
175             if MASK(idx,idY,idS)
176                 VoxSignal = squeeze(Slog(idx,idY,idS,1:nB));
177                 if all(VoxSignal ~= -Inf)
178                     DS = -inv(B'*B)*B'*VoxSignal;
179                     DD = [DS(1) DS(4)/2 DS(5)/2;
180                            DS(4)/2 DS(2) DS(6)/2;
181                            DS(5)/2 DS(6)/2 DS(3)];
182
183                 [VEC, VAL] = eig(DD); VAL = diag(VAL);
184
185                 if all(VAL < 0), VAL = abs(VAL); end
186                 if any(VAL < 0), VAL(VAL < 0) = 0; end
187
188                 [VAL, ID] = sort(VAL);
189                 VEC = VEC(:,ID);
190
191                 % FRACTIONAL ANISOTROPY
192                 DENFA = VAL'*VAL;
193                 NUMFA = (VAL(1)-VAL(2))^2 + (VAL(2)-VAL(3))^2 ...
194                             + (VAL(3)-VAL(1))^2;
195                 FA(idx,idY,idS) = sqrt(1/2)*sqrt(NUMFA/DENFA);
196
197                 % MEAN DIFFUSIVITY
198                 MD(idx,idY,idS) = mean(VAL);
199
200                 % PRINCIPAL VECTOR
201                 V(idX,idY,idS,1) = abs(VEC(1,3)).*FA(idX,idY,idS);
202                 V(idX,idY,idS,2) = abs(VEC(2,3)).*FA(idX,idY,idS);
203                 V(idX,idY,idS,3) = abs(VEC(3,3)).*FA(idX,idY,idS);
204                 %V(idX,idY,idS,4) = sum(VEC(:,1).*FA(idX,idY,idS));
205             end
206         end
207     end
208 end
209 end
210
211 %% ANISOTROPIC ANALYSIS - VISUALIZATION
212 % Visualization of the information contained inside the principal vector
213 % matrix computed at the step before.
214
215 figure
216 for idS = 1:1:nSLICE

```

```
217 tmp = reshape(squeeze(V(:,:,idS,:)),nVOXX,nVOXY,3);
218 image(imrotate(tmp,90))
219 title(sprintf('SLICE %i - Directional information',idS))
220 pause(0.2)
221 end
222
223 %% ANISOTROPIC ANALYSIS - VISUALIZATION
224 % Visualization of the 25th slice inside the four - dimensional volume
225 % computed at the previous steps.
226
227 tmp25 = reshape(V(:,:,:,25,:), nVOXX, nVOXY, 3);
228 figure, image(imrotate(tmp25,90)),
229 title(sprintf('SLICE 25 - Directional information',idS))
```

LESSON 18: FNIRS INTRODUCTION

functional Near Infrared Spectroscopy: Introduction

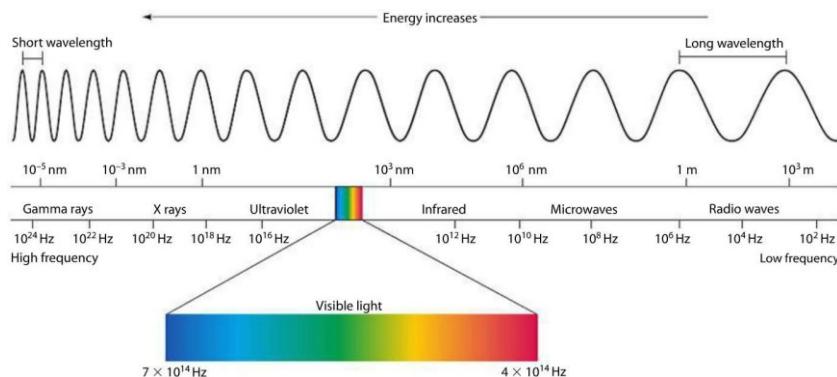
In this lesson and the further we will talk and introduce the *functional Near Infrared Spectroscopy (fNIRS)*. This is an optical technique that could measure the brain activity using light. If we perform an experiment and we place over the light of our smartphone our finger we could notice that only the red light will pass through it. This means that the other waves of the light are completely absorbed from the tissue and thus the intensity attenuation could not be measured. Moreover, another important thing that we could notice from this experiment is that the light will diffuse all over our finger. This is quite important because it means that the light beam it is scattered over all the directions.

Electromagnetics waves: Interaction and characteristics

The light interacts with matter via two fundamental mechanism. These are the *absorption* where the energy of the photon is absorbed by the medium, hence a certain quantity of energy is loss by the photon. While the second mechanism is known as *scattering*. In this case the photon energy remains unchanged but what varies is the direction of propagation of the considered photon.

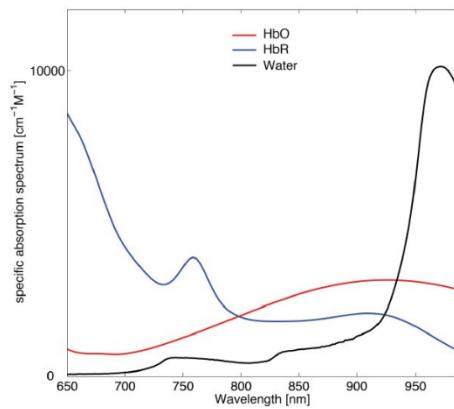
The likelihood of these two interactions within a given biological tissue depends on the scattering μ'_s and absorption μ_a coefficients of the given tissue, which in turn, are dependent on the wavelength of the light. These coefficients have units of $(\text{length})^{-1}$ and therefore represent the inverse of the average distance that the photon will travel through a material before experiencing an absorption or scattering interaction. An example of scattering and absorption coefficient that are null is the water. In fact if we point a laser through the water this beam passes unchanged. Instead, if we change the material, and as instance we consider a black tea glass cup the coefficients changes. Specifically, it appears a non – null absorption coefficient. To let the black teacup be characterized of a scattering coefficient not null we could insert inside it some milk which will satisfy this request. Hence, from these three simple examples we could appreciate the different electromagnetic properties of the different materials.

Down below is reported the electromagnetic main scheme of the light spectrum. Here we could notice how much close are the infrared light to the visible spectrum and to the red light. From this scheme we could see how the wavelength and the frequency of the light changes by varying the energy contained inside the radiation.



In general as the wavelength decreases and thus the frequency increases the energy associated to that light beam increases too. One more additional information is that even the infrared light close to the visible one could be clearly see visually.

To apply these concepts of absorption and scattering and how we could exploit them inside our body is reported in the image down below. This chart report how much is the specific absorption rate inside some specific biological compound at different wave lengths. From this chart we could appreciate that the oxygenated and deoxygenated haemoglobin, respectively HbO and HbR are characterize of a different behaviour. This is very meaningful, and we will discuss later in this lesson of the possible application.



Cytochromes are biochemical molecules that if studies represent a direct measure of the brain's metabolism. However these compounds have a very low concentration inside the tissues compare to the one of the oxygenated and deoxygenated haemoglobin. The biological tissues such as scalp and skull are not completely transparent to red and near infrared light. Specifically, in fact, they are characterized from a such strong absorption rate which will decreases a lot the light beam intensity. However, thanks to the knowledge of the two previous compounds, HbR and HbO , we could appreciate that these two components depending on the light wavelength presents a specific absorption rate. Specifically, if we get a measure at 600 nm and another one at 850 nm, thanks to the knowledge of the absorption coefficient of these two elements we could determine the concentration of the two compounds. From the same chart we could notice that we could not increase too much the wavelength because it will be more absorbed from the water instead of these two substrates.

Haemoglobin: Role and states

The haemoglobin is a biological molecule which is very important because it is the main transporter inside our body of oxygen O_2 and carbon dioxide CO_2 . Usually, the haemoglobin could be found as already previously said in two main structures, the oxygenated (HbO) and the deoxygenated states (HbR). The red light has been successfully used for the detection of blood vessel to perform blood test in a safe environment and also to train the new nursery staff. The first technology that was developed was the Pulse Oximeter in 1972 from the Takuo Aoyagi. This technology was spread all over the world in the last year of pandemic because was used to measure the amount of oxygen that it is present inside our finger.

NIRS: Beer Lambert Law

The pulse oximeter is based on a very simple concept. This relies on emitting a certain light beam through our finger and measure the intensity of the exiting beam. The attenuation of the signal is due to the absorption of mainly the two compounds known as the *HbR* and *HbO*. The mathematical description is formalized from the *Beer Lamber Law* where it is assumed that there is no scattering phenomenon activated inside our finger.

Thus if we substitute our finger to generalize the situation with a general medium by modelling it with a certain rectangle described with certain parameter of width and absorption coefficient μ_a . This last one, just to remember, is strictly related with the material. The mathematical relation between the input and the output intensity of the light beam signal is the next one:

$$\frac{I_{out}}{I_{in}} = e^{-x\mu_a}$$

The absorption coefficient of our material could be rewritten with the next analytical description. Here we account for the different compound which will generate the attenuation inside our finger:

$$\mu_a = \tilde{\mu}_{HbO} C_{HbO} + \tilde{\mu}_{HbR} C_{HbR} + \tilde{\mu}_{H_2O} C_{H_2O} + \tilde{\mu}_{Fat} C_{Fat} + \dots$$

In this case we have just highlighted the different contribution to the absorption coefficient by the concentration of oxygenated haemoglobin, the concentration of the deoxygenated haemoglobin, and finally the concentration of the water the fat and other components. The value of μ for which each concentration is scaled is called the *specific absorption coefficient* which are measured in $\frac{1}{mm\ Molar}$. These coefficients change its values by changing the wavelength considered. From the Beer Lamber Law we could estimate the value of the absorption coefficient.

$$\frac{I_{out}}{I_{in}} = e^{-x\mu_a} \Rightarrow \mu_a = \frac{1}{x} \ln \left(\frac{I_{in}}{I_{out}} \right)$$

By measuring optical attenuation at a given wavelength we can calculate μ_a at that wavelength. By measuring optical attenuation at many wavelengths, we can measure the absorption spectrum of a material, which provides information on the components of itself.

In 1977 Frans Jöbsis sees that it could be used the light in the brain to measure of the biological activity. In fact the red light could pass through the scalp and could let us be able to measure the brain oxygenations changes during time. In fact, Jöbsis performed the first transcranial spectroscopy experiment. It was also observed that the source and the detectors couple could not be placed in diametral position inside the head. This because the intensity of the light beam needed to travel such a through such a long and complex structure is very high. This will overcome the limit superimposed from IEEE in terms of electromagnetic waves maximum amplitude. Hence, the position of the sources and the detectors should be tiny controlled.

The first application of the NIRS technique is the *cerebral oximeter* for neonatal subjects which is still nowadays used. This technique works suddenly because of their thin width of the scalp. Hence, the intensity of the light beam would not be so of a high value. The technology however it has been adapted also for elder subjects which are characterize of a thicker scalp than the child. This leads to the possibility to study the changes in blood oxygenation in some pathologies.

Electromagnetic waves: Biological tissues

Inside biological tissues the scattering and the absorption relative to the near infrared light is dominated from the scattering instead of the absorption. In fact, the scattering coefficient of the skin is typically 100 times greater than the absorption coefficient meaning that a photon is 100 times more likely to be scattered than absorbed per unit distance travelled. As results the path of a near infrared photon in tissue resembles to a random walk.

Thus, if we place a source of near infrared light over our scalp, we could detect the spread over all our head of the light emitted. This is because the skin scattering coefficient is much higher than the absorption one. Specifically, not only the scattering coefficient of the skin but also of the most inner structures are characterize of a higher scattering coefficient than the absorption one. A possible example of propagation of the light beam emitted is reported down below.



In this case the absorption is higher closer to the position where the source is placed. As we go further from the source position, or equivalently we go deeper inside the brain the absorption decreases a lot and if we detect photons this is only due to the scattering effect that predominate. Specifically, as we could see from the images reported above, after a certain settling time if the light beam is maintained active it will activate a whole photons distribution.

NIRS: Modified Beer Lamber Law

Since scattering predominate inside the head and the brain tissues, we need to modify the Beer Lamber Law to let this equation be used also in this condition. In fact, if we apply just the Beer Lamber Law the values detected will not satisfy the condition of poor scattering situation. Thus in our medium we could not more considered only the μ_a but, we need also to consider the μ_s . Hence, because of this last request we must consider the possible light direction where the beam will be spread inside our head.

As previously said, photons could come out from different directions. Thus the first problem that we need to account for is that the photons that after the scattering phase assume a direction which does not point more towards the detector will be lost.

The second main problem is that, if a certain photon is absorbed partially and enters inside our head it could face with different scattering activity due to the different tissue which it will interact with. This increases the travelling distance causing a loss of energy that could be then detected and altered the measure.

The third problem is that we have an intensity captured also from our hair which absorption coefficient is just unknown. Thus the lamber beer law could be rewrite in the next way:

$$\frac{I_{out}}{I_{in}} = e^{-x\mu_a} + \text{Losses}$$

The unknowns value inside the previous relation are a lot, while the measured are a few. In fact, inside the previous equation we could only know the intensity of the light that exits out from the scalp, while the unknowns are represented from the intensity in input, the losses and the travel distance due to the random walk.

To solve this problem the first trick that could be used is to measure the concentration changes instead than absolutes values. This because the value of μ_a varies with time due to the variation of the compound's concentration. The mathematical relation could be written in the next form:

$$\ln \frac{I_2}{I_{in}} - \ln \frac{I_1}{I_{in}} = \ln \frac{I_2}{I_1} = -x\Delta\mu_a \Rightarrow \Delta\mu_a = \frac{1}{x} \ln \frac{I_2}{I_1}$$

Where x represents the distance travelled from the photons. Further assumptions are that the losses and the intensity in input do not change between t_1 and t_2 . The distance travel from the proton is not measured directly but we could assume that X is equal to the distance between the source and the detector multiplied by a simple scaling factor that keeps the random walk of the photons into account.

$$x = mD$$

Where D is known as the *Differential Pathlength Factor* and depends on *wavelength* and *age* according to the next analytical equation:

$$DPF(\lambda, age) = \alpha + \beta age^\gamma + \delta \lambda^3 + \epsilon \delta^2 + \xi$$

Another assumption is that the DPF does not change between t_1 and t_2 . The assumption requested is very important because means that the value is stable over time during the overall acquisition. By using these last assumptions inside the variation of the absorption coefficient variation we could write:

$$\Delta\mu_a = \frac{1}{mDPF} \ln \frac{I_1}{I_2}$$

As absorption coefficient was linearly related to concentration, now the changes in the absorption coefficient are linearly related to changes in concentration of the absorbing molecules at a given wavelength.

$$\Delta\mu_{a|\lambda} = \tilde{\mu}_{HbO|\lambda}\Delta C_{HbO} + \tilde{\mu}_{HbR|\lambda}\Delta C_{HbR} + \tilde{\mu}_{H_2O|\lambda}\Delta C_{H_2O} + \dots$$

With NIRS, we can measure the change in absorption coefficient of a tissue, which is dependent on the change in concentration of chromophores such haemoglobin, water, fat and other components.

During an acquisition, thanks to the fact that we are measuring the variation of the compound's concentration, we could assume that the static compounds as fat, water and other more not significant ones do not change in time. By this last assumption inside the previous mathematical relation we could disregard all the terms apart from the oxygenated and deoxygenated haemoglobin.

$$\Delta\mu_{a|\lambda} = \tilde{\mu}_{HbR|\lambda}\Delta C_{HbO} + \tilde{\mu}_{HbR|\lambda}\Delta C_{HbR}$$

The previous assumption is confirmed also because inside the NIRS the absorption coefficient of the water changes just a little and thus this variation could be neglected.

By inspecting the previous equation we could clearly see that the unknowns in the previous equation are respectively two. These are represented from the variation of the two states haemoglobin concentration. This problem could overcome by taking two different measures at two given wavelengths. In this way we could solve the following system and obtain the so – called *modified Beer Lambert Law (MBLL)*:

$$\begin{aligned} \Delta\mu_{a|\lambda_1} &= \tilde{\mu}_{HbO|\lambda_1}\Delta C_{HbO} + \tilde{\mu}_{HbR|\lambda_1}\Delta C_{HbR} & \ln\frac{I_{1|\lambda_1}}{I_{2|\lambda_1}} &= mDPF(\tilde{\mu}_{HbO|\lambda_1}\Delta C_{HbO} + \tilde{\mu}_{HbR|\lambda_1}\Delta C_{HbR}) \\ \Delta\mu_{a|\lambda_2} &= \tilde{\mu}_{HbO|\lambda_2}\Delta C_{HbO} + \tilde{\mu}_{HbR|\lambda_2}\Delta C_{HbR} & \ln\frac{I_{1|\lambda_2}}{I_{2|\lambda_2}} &= mDPF(\tilde{\mu}_{HbO|\lambda_2}\Delta C_{HbO} + \tilde{\mu}_{HbR|\lambda_2}\Delta C_{HbR}) \end{aligned}$$

NIRS: Modified Beer Lamber Law - Wavelengths

The problem now is to select the right value of λ to choose. In the NIRS general application there are two standard choices that are developed on the same concept. This consists in taking one light at a wavelength lower than the isosbestic point and the other higher than this last one. In this way we could exchange the absorption coefficient of the two species and get more informative and accurate results. The common choices are the next two set of wavelengths:

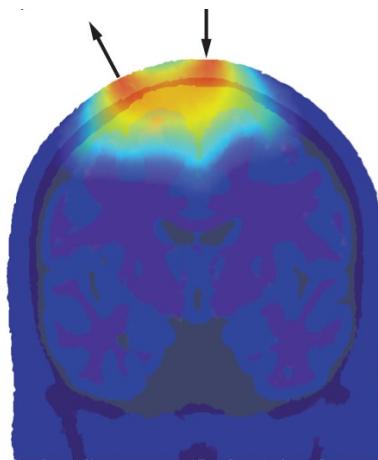
[780,850][nm]

[760,830][nm]

These two wavelengths are delivered at the same time and in the same position of our head. Then are retrieved from the detector and are analysed computationally.

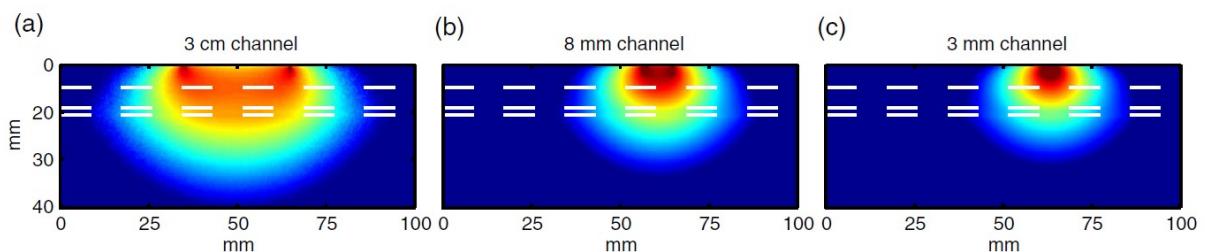
NIRS: PMDF

In the brain as previously we have enhanced, if we place a certain source that emits a certain amount of light this will be spread all over our head. Now, we need to find a specific position where to place the detector to measure the intensity of the light beam delivered. It has been observed that by placing a detector close to the source this will generate a certain banana shape pattern of the light detected. A pair made up of a detector and a source it is called a *channel*. The banana-shaped path followed by photons leaving the source and being measured at the detector is called *Photon Measurement Density Function (PMDF)* and has been computed using probabilistic approaches simulating the propagation of photons in biological tissues (*Monte Carlo simulations*). The *PMDF* can be thought of as giving the probability that a detected photon will travel through a given region of tissue. The greater this probability, the more sensitive the NIRS measurement will be to changes in the optical properties of that region.



In the figure is reported a possible probability maps where we could measure and detect the signal intensity attenuation. Inside the figure are highlighted even the source and the detector position and the high banana shape made up.

NIRS: Source and detectors separation



Another thing that needed to be controlled is the distance between the source and the detector. Let assume that we model the head with a four layers structure, respectively, the scalp, the skull, the grey matter and finally the white matter. If the distance of the two electrodes is not enough, we will be more sensitive for the most external layer. Vice versa, if the separation of the source and the detector is enough it will generate a good banana shape that will be sensitive to the most inner structure as the grey matter. Obviously,

increasing the separation will affect the intensity of the signal measured. Consequently further are placed the two fibers lower is the intensity of the signal measured and less reliable is the measurement detected.

The biological tissues are transparent to near infrared light, and this last type of light is selectively absorbed by oxy and deoxy haemoglobin. The modified Beer Lambert Law can be used to relate optical intensity with absorption coefficient and consequently concentration of absorbing molecules. Hence, NIRS can be used to measure the concentration changes of oxy and deoxy haemoglobin and as we know from the fMRI theory these two compounds will change its concentration during task related activities.

Neurovascular coupling principle

When our brain is doing something, the metabolic demand is increased. The vasculature responds by dilating and increasing blood flow and particularly, to the brain is send an higher amount of blood flow than needed to do not incur in some less oxygen concentration in brain area. This produced a lot of localized response that are expressed in changes of HbO and HbR . The main difference between the signal that we could measure with the fMRI and the signal that we could measure with the fNIRS is that the first one could only estimate the peaks while with the second we could estimate more properly the HRF. The disadvantage of this last technique is that could reconstruct only the most external layers and not the internal one as the fMRI.

NIRS: Single channel and Multi channel

Single channels NIRS provide just an information of a local region of the brain and not on a wider one. In this case to solve this problem is used the multichannel NIRS where there are a lot of multiple sources and multiple detectors where we could take different information.

NIRS: Hardware and commercial devices

The NIRS hardware is very close to the hardware used in EEG technologies. In fact we have a box of 60 x 60 cm that is connected with optical fibers. These last ones in turn are connected with a cap which will be positioned inside the head of the subject. The cable connected to the cap could be either sources or detectors.

NIRS devices can substantially differ in the type of sources, the type of detectors, the number of sources, the number of detectors. There is not a standard configuration as for the EEG and depends heavily on the commercial brand of the element. System moreover could also different one between another for the sensitivity to longer or closer distances. Hence, the choice of one commercial instrumentation or another could differ due to the application that we need to develop. Nowadays are developed also some types of instruments that are fibreless that contain the whole electronics for the NIRS. All the data that are acquired are then send directly to a computer.

LESSON 19: NIRS COMPOSITION

NIRS: Hardware

The hardware that are characteristic for the NIRS technologies are a lot and depends on the typology of the study that it is needed to be performed. In particular, if we would like to have a very high detailed spatial configuration of the different superficial HRF activated we need to use some multi – channel instrumentation that is made up of a lot of fibers. These last ones are characterized of a very heavy weight and to hold all the cables are build up specific supports. Nowadays the most important NIRS devices are also portable and for this reason NIRS in infants is much better and more reliable in infants thanks to the portability of the technology. In this case of portable devices, inside the cap are placed some small hexagons that contains the whole electronics of the fNIRS system. This is a real breakthrough the NIRS technology and the whole system once positioned remain fixed on subject's head. Hence, this decreased a lot the significance of the movement artifact. The company that produces this type of devices is *Lumo*. This system is very dynamic because we can place wherever we want inside the cap the different hexagons to get much more signals coming from a certain brain section instead of another. Some examples of studies where this technology has been used involves child that moves and plays or interact with their moms.

NIRS: Resolution

The fNIRS technology has the next collocation inside the measurement system. This is characterized from a good *temporal resolution* compared to the fMRI. This because this last technique gets samples with a sampling frequency equal to 6 Hz while our technology with 200 Hz. The *spatial resolution* is not one of the best and this is mainly due to the different physical dimension of each channels placed over the scalp. Compared to the information retrieved from the fMRI which have a spatial resolution of 3 mm this is much lower and approximate equal to 1.5 cm. Thus get information from the fNIRS technology could let us estimate better the HRF activation but we could not detect the HRF of the inner part of the brain. From this last sentence we could understand that the *depth* is another important property of the fNIRS technology that is reachable with that specific array configuration. In fact, as inside our array of channel we increase the separation between sources and detectors we could get information in deeper structure. But to obtain this last one concretely we need to increase the intensity of the emitted signal which should not overcome a certain threshold of security established from the IEEE community.

NIRS: Difference between fMRI and fNIRS - Characteristic

To summarize the possible differences between the fMRI and the fNIRS technology we report below a summarizing table.

	<i>fMRI</i>	<i>fNIRS</i>
<i>Sampling Frequency</i>	Low	High
<i>Spatial Resolution</i>	High	Low
<i>Depth Resolution</i>	High	Low
<i>Anatomical Information</i>	Yes	No
<i>Detect HbO and HbR</i>	No	Yes

<i>Portability</i>	None	High
<i>Noise</i>	High	None
<i>Invasivity</i>	Low	Low
<i>Motion Sensitivity</i>	High	Low
<i>Movement Constraints</i>	High	Low
<i>Interference with other devices</i>	High	No

One of the most important difference between the two considered system that is also reported inside the table before is that the fNIRS could detect the changes in HbR and HbO while the fMRI technology not. This because this last technology could only detect the HbO that is highly related with the BOLD signal. Obviously, because of the fNIRS is a portable device could be used to get information even during movement of the subject. This will corrupt a lot the fNIRS signals. The movement artifact instead could not be detected in fMRI because the movement space is very tiny and for the acquisition to have some relevant results is asked explicitly to maintain the same position.

NIRS: Difference between fMRI and fNIRS - Usage

Diffuse optical methods should not be used to map human brain function if fMRI is readily available and suitable for the experimental population and the chose paradigm. There are other studies where the fNIRS technology is well suited and this happen when the experiment involves children, infants, or vulnerable adults as well. As instance, when we need to deal with claustrophobic subjects and when we need to study the HRF activation in involved motor task activities.

An example where the fNIRS technology is used is to study if the cochlear implants works correctly or not and to determine eventual disfunctions inside the auditory system. Other studies are the one where we involve movement of the subject as instance to inspect the HRF activated during walking and running activities. Another possibility is to use the fNIRS to study the social interaction between a group or two people. There are a higher number of studies that we do not cite yet and they could be done with fNIRS instead of fMRI because of the portability of the instrumentation.

The fNIRS is less know than fMRI because these two technologies were born exactly in the same year or the 1992. Then thanks to the clinical acceptance of the fMRI technologies this last field becomes wider and enrich instrumentation and techniques a lot. Instead, fNIRS is majorly used in research for non – clinical application also nowadays and not only in the past.

In the same year, in the 1992, came out four different papers that declare the birth of the fMRI and the fNIRS technologies. During the first years of application the problematic shown by the fNIRS technology were too complex and was not available any possible solution with the tool of that time. Hence, fMRI was initially push much further because it was ready for being used inside the everyday clinical practice. The solution of the fNIRS problem was proposed, found, and developed in the last ten years of research. Moreover, this technique has been confirmed during the last decade to perform some clinical evaluation in infants thanks to the lower motion artifact susceptibility of the technologies and the thin scalp layer.

The challenges that are nowadays available for this technology are listed in the next paragraph.

- *Lack of information about anatomical structure.* This was also the main reason because the community choose to push further the MRI field instead of the NIRS one.
- *Limited coverage.* This because we could get information about a very restricted brain area and we could not measure the whole brain structure. This has been partially solved with the same configuration of the high – density EEG build up with the NIRS technology.
- *Optical fibers are bulky, above all for high – density montages.* This prevents the whole scalp coverage. This has been partially solved with the advent of the high – density wearable devices.
- NIRS only provides relative quantification of the concentration changes and not the absolute values. This also was partially solved with some technological advances in other optical techniques.
- Optodes are placed on the scalp, which means that NIRS signals are highly sensitive to superficial layers.

Optical techniques: Continuous Wave

As there are many different MR modalities that can be used to quantify different physiological indexes there are other different techniques based on NIRS that can quantify similar indexes. The NIRS system we have seen so far are called *continuous wave (CW)* system and they measure only the intensity of light incident at the detector. They cannot separate the effects of absorption and scattering, and they cannot estimate the absolute measure the chromophore concentration.

Optical techniques: Spatially Resolves Spectroscopy

With the same technology we could perform some *spatially resolves spectroscopy (SRS)*. This technique let us be able to measure the tissue oxygen saturation. This is done thanks to the usage of some sources and some detectors in the same side of the tissue. The oxygen concentration extracted is just related with the most peripheral information. In fact, the light beam is emitted from the source and thanks to the deflection of the photons which energy is then measured from the detectors. The blood oxygen saturation could be described as the ratio reported below:

$$SO_2 = \frac{HbO}{HbO + HbR}$$

This is the principle where the cerebral oximeters work.

Optical techniques: Frequency Domain NIRS

Another possible optical technique that uses the previous consideration on the light beam is the *frequency domain NIRS (FD - NIRS)*. In this case the measurement system is detecting both the amplitude and the phase of the signal. Thanks to this information we could estimate the real DPF. The advantages of using the FD – NIRS instead the fNIRS is that the signals measured are not just relative changes but absolute values.

Optical techniques: Time Domain NIRS

One of the last NIRS technique is the *Time domain NIRS (TD – NIRS)*. In this case instead of measuring the intensity and the phase we are measuring the time of flight of individual photons. Hence, the photons that will be captured first are the one that have interacted fewer with the biological tissue. Instead the photons that will arrive last travelled a lot inside the tissue. Thanks to this measure we could estimate the *Temporal Point Spread Function* and it will be helpful to estimate the absolute value of the light intensity.

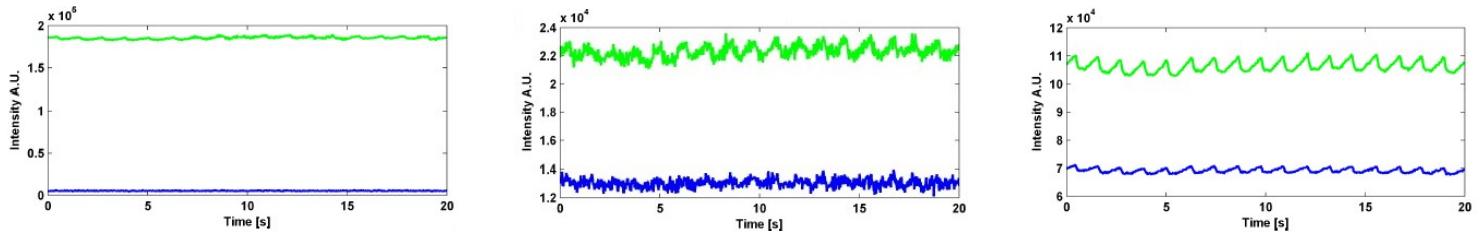
Optical techniques: Diffuse Correlation Spectroscopy

The last but not the least important optical technique is the *Diffuse Correlation Spectroscopy (DCS)*. This technique is very similar to the one of ASL proposed inside the MRI technique. With this we could estimate the blood flow index and a perfusion index which are very close to the one measured with the equivalent MR technique. In this case we could detect the fluctuations due to the different speed of the blood flow that will produce changes in the movement of the photons belonging to the beam emitted.

By combining different optical techniques we could get different and various information from the brain such as concentration of chromophore.

NIRS: Physiological and Non Physiological components

One other problem that we could detect and highlight inside the fNIRS technology is the possibility to detect certain noises and different pattern between channels that are very close one with each other. This is due to the possible noises that is affected a given channel. The types of noises could be physiological related and not physiological related. If of the first type we could find heartbeat or breath noises which usually enhance the reliability of the measurement, while if of the second type, they should be removed from the signal before inspecting them. In the figure reported below we could appreciate some different patterns where is activated the heartbeat or some breath noises.



From the picture reported above the image more located on the left represents just white noise and a pattern of any informative signal. The image in the center presents some physiological artifact such as heartbeat and some high frequency components that corrupt the signal. Finally the image on the right shows a signal which is mainly made up of heartbeat. In general, if we find physiological artifact inside our signals this is a marker of good measurement process. Instead, if the acquisition does not show any physiological artifact inside the pattern it means with a certain degree of accuracy that the measurement process does not end up well. When we inspect a given NIRS signal we need to know that this component is made up of extra cerebral and cerebral contribution. The first one is mainly related with the skin and the skull which are tissues highly vascularized. Instead the second contribution is what we want to measure.

NIRS: Physiological and Non Physiological components – Classification

If we would like to understand which are the main component of the NIRS signal, we should refer to the scheme reported below.

Systemic Component: Vasomotor waves (≤ 0.1 Hz), Respiration (≈ 0.2 Hz), Cardiac (≈ 1 Hz).

Neural Component (Cerebral Component): Haemodynamic Responses (≈ 0.1 Hz).

Extracerebral Component: Mainly related with blood and noises.

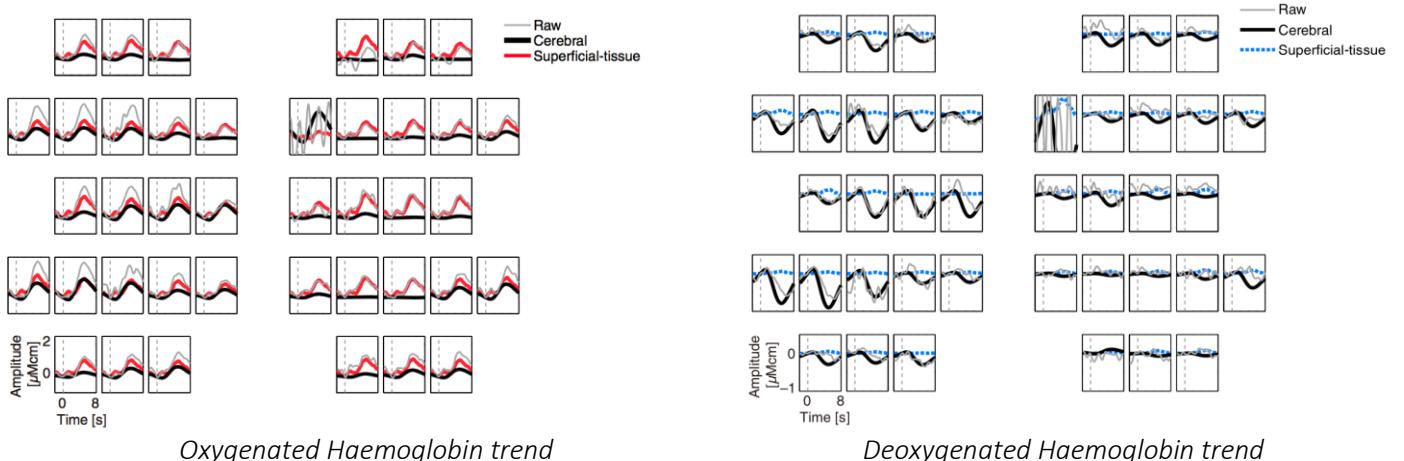
Measurement Noises: White noise overlapped with the signal.

The *system components* are the one that should always be found inside the NIRS signal to determine that the measurement is performed well. Mainly, the vasomotor waves are related with the alteration of blood volume that is carried from blood vessel inside specific brain area. The *neural component* instead is the one that we should detect and measure, finally we have some *extracerebral components* that we need to remove from our signal. This process will be described in the next lessons.

NIRS: Evoked and Non Evoked responses

The signal composition of the NIRS could be subdivided more specifically also in *evoked* and *non – evoked* components. We refer to the first one as the situation where thanks to the activation of task a certain brain region is stimulated, and this will change the blood flow versus that region of the brain. The second one instead refers to the situation and activities that are always active and the blood flow is maintained almost constant during this situation. This last type of activity is much more difficult to detect and deal with. Specifically it has been shown that the even in the *non – evoked* activities we could detect some system changes. These last one are more enhanced in the NIRS analysis instead of the fMRI one because of the supine position of the subject. Researchers, in fact, has demonstrated that in supine position the system compounds intervene more than in a lay down position (like the fMRI one).

In the example reported before we could appreciate the representation and measurement of a systemic evoked signal task related with in this case the squeezing of the right hand. We could see the activation of a given HRF in the left hemisphere while inside the right hemisphere we could appreciate the contralateral activation (HRF in the lower right channels represented of the scheme). The superficial signals have a systemic HRF that could mimic the haemodynamic response.



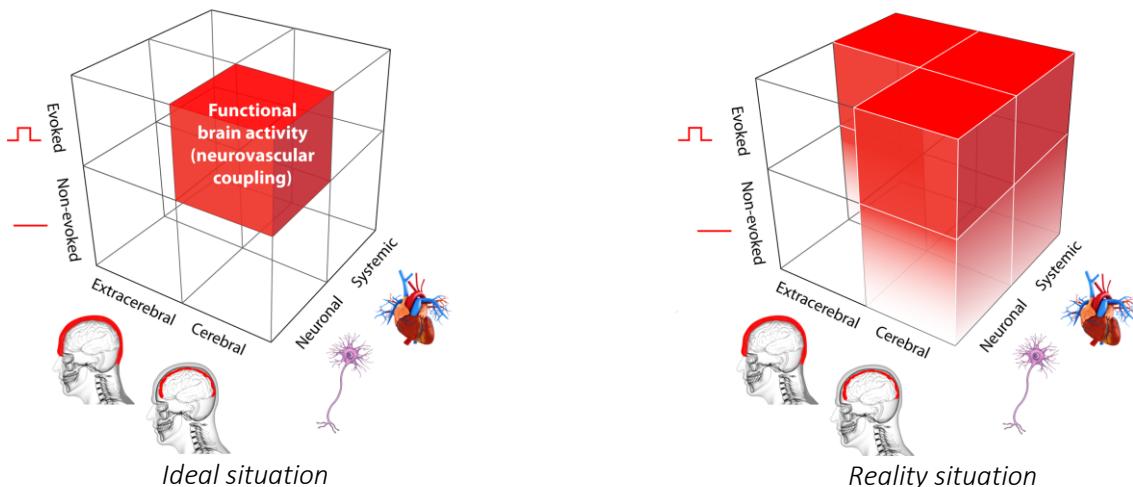
In the picture on the right side of the page we could see the different pattern that could be measured once we consider some superficial tissues and some cerebral one. Here inside the superficial one we could not appreciate any HRF that mimic the HRF of the cerebral tissue. Moreover, in general the deoxygenated haemoglobin is usually less contaminated by physiological confounds.

NIRS: Physiological components – Mayer waves

One of the most problematic compounds detectable in the NIRS signal that could corrupt a lot the informative content is the *blood pressure waves*. These are also called the *Mayer waves*. They are the most challenging compound inside the detection of HRF. This because they have a frequency which is very close to the one seen for the HRF specifically 0.1 Hz. Thus the discrimination between HRF and Mayer waves is very difficult. It has been shown that the Mayer waves have a different contribution depending on the scalp head and this is mainly related with the different position of major arteries. To understand better the role of the Mayer waves we need to understand their effect on the accuracy reconstruction of the HRF. In this case, if the Mayer waves are strong enough, they will reduce the accuracy in reconstruction of HRFs. Instead, if the Mayer waves effect is not so strong it won't be a problem reconstruct the different HRF.

NIRS: Signal composition

In the next image we reported how the NIRS signal is made up. What we would like to highlight is that the only compound that is significant for our application is the cerebral neuronal evoked HRF. This because we could determine which brain area is activated during a task and we could distinguish between pathological and physiological condition in this response.



The fact is that we would like to determine only the functional brain activity due to the neurovascular coupling phenomenon activated in response to electrical activity of the neurons.

NIRS: Channel position

The channel position in the NIRS technology is not well defined as with the EEG one. This because are two different techniques and in the second case the optodes are in brain area that we would like to inspect and study. To have also some physiological and anatomical relevance of the result we need to use some MNI parcellation to correctly position the different channels. Thus, because of the MNI space is always useful, to perform correctly the NIRS analysis we should also have some anatomical information retrieved as instance with T1w or T2w images with fMRI.

NIRS: Channel connection

The second problem that we need to solve is how we can connect the detector and the sources to create our acquisition grid. Generally there is not yet a standard that we could use to determine the array configuration. Thus usually we need to set up our own by thinking on what we should measure in that specific analysis. Down below are reported some possible array configuration. During the laboratory we will use an array configuration which has been developed from the research team of the Department of Information Engineering in Padova.



Where the red dots indicate the sources while the blue dots indicate the detectors. The simpler and commonly used approach consist in design an array over the ROI, run photon migration simulations to obtain the brain sensitivity of the array (the PMDF of each channel) and manually and iteratively move the array until the wanted ROI is covered. In this case if we sum all the PMDFs of individual channels we could get information about the sensitivity of different brain region. After having placed all the channels we need to connect the sources and the detectors in different way as instance by using squares or by using triangles. This depends on the type of overall PMDF that we would like to generate.

One very recent approach is to try to solve algorithmically in an automated way the array design problem to yield an optimum array covering the given ROI. In this way we could automate the process by finding those positions that are relevant to obtain a good map of the ROI considered.

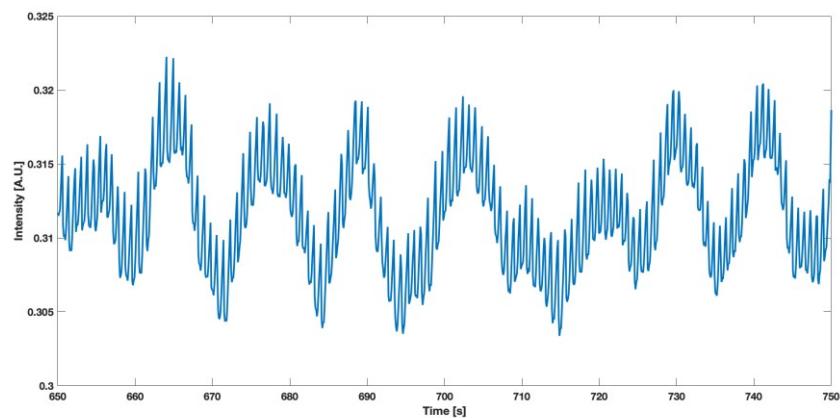
Once the optodes placement has been defined, the NIRs cap can be created. Usually a soft black tissue cap of a similar size of the participant's head is used. This depends on the subject where we need to apply this cap and needs to have some specific region and support to place the different optodes in correspondence to the brain region of interest and related with the sources and detectors position chosen to map this last one. As fMRI there are different types of paradigm that could be used, the most common are *event related designs* and *block design*. Usually approximately 40 trials per condition are presented for event related designs and 10 blocks per condition for block design to achieve enough signal to noise ratio to obtain a reliable estimate of the haemodynamic response.

LESSON 20: NIRS STANDARD PROCESSING PIPELINE

Today we start to see how we could analyse the fNIRs signal. As we have seen in the last lesson there is a methodology that we could follow to acquire the different fNIRs signals. Specifically the number of signals that are acquired depends on the quantity of channels that are available inside the configuration set up.

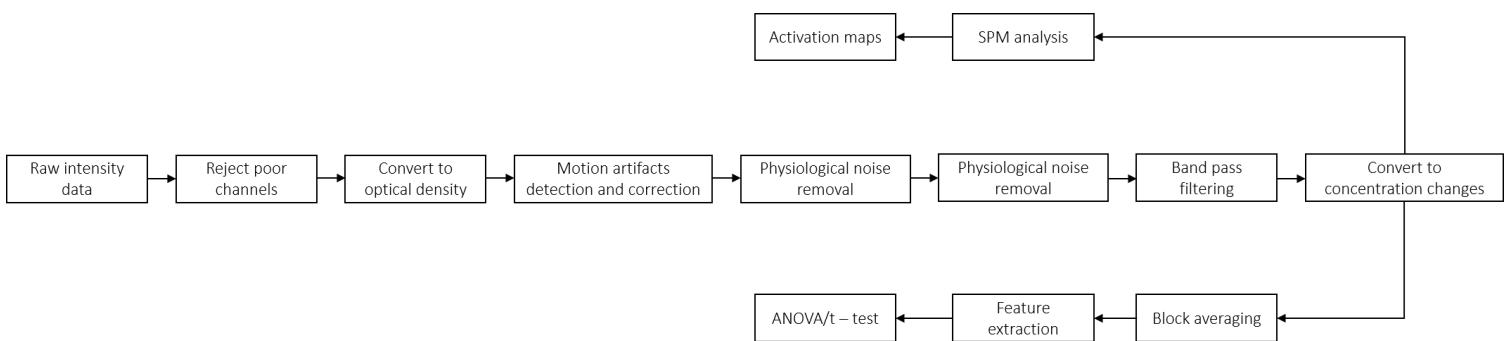
NIRS: Ideal signal

An ideal signal that we could measure with this technique is the one reported below inside the image. In this case we could see the intervene of some physiological compounds as respiration and heartbeat. In this case reveal some physiological components is not a bad thing because means that the measurements have a lot of meaning and we are really measuring the blood flow. The slow component and oscillations are related with the respiration while the higher frequency spikes are mostly related with heartbeat.



NIRS: Standard processing pipeline

The standard process is summarized with the scheme reported below. Inside the pre – processing phase we have some relevant steps like the motion artifacts detection and correction and the physiological noise removal. The first cited operation is the most important because as we have seen in the previous lecture the NIRS signal are heavily corrupted from motion artifacts. The second operation mentioned instead, is necessary to enhance just the HRF inside the signal. Instead, in the last part of the NIRS analysis is represented from the statistical analysis section. With the SPM method we could generate some relevant activation maps which could be used to enhance the possibility to extract activation maps as the fMRI results.



The most famous software that are available are *Homer2*, the *Homer3*, *NIRS – SPM*, *NAP*, *NIRS Toolbox* and *NeuroDOT*. All these software have a GUI that could be used from the clinician to analyse the data. However

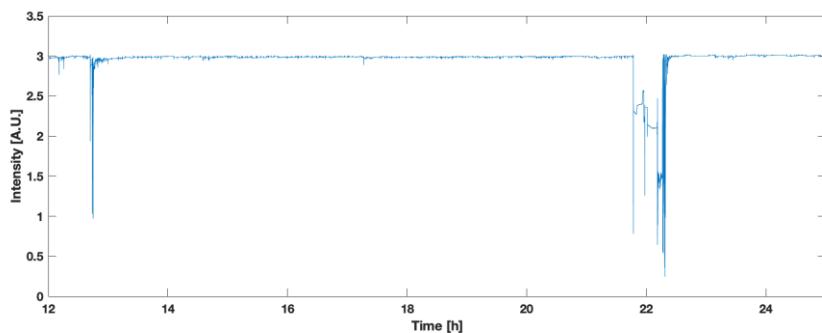
software has also some functions that could be used inside own code which is for engineers the preferable ways because we could control all the possible variable.

NIRS: Standard processing pipeline - Rejection of poor channel

After having acquired all the possible signals we could face some problems with specific channels. We could have issue with hair thickness, colour, skull thickness and source detector distances that can influence the amount of light that is being measured at a detector. Moreover, in addition to this, the photodiodes convert the light into an electrical current, which can be measured as voltage. This instrumentation and electronical devices have a characteristic response curve. Here we could appreciate that the linear characteristic of the electronical device is contained inside two bounds. One is an upper and one is a lower. Thus, the power of the light incident to the photodiodes needs to be contained inside a certain interval. In fact if the power is too high the photodiodes response shows a characteristic saturation limit. While if the signal measured could be associated to the noise floor this is always associated to a unique value. Thus we need to not consider some specific channels depending on the amplitude of the signal measured.

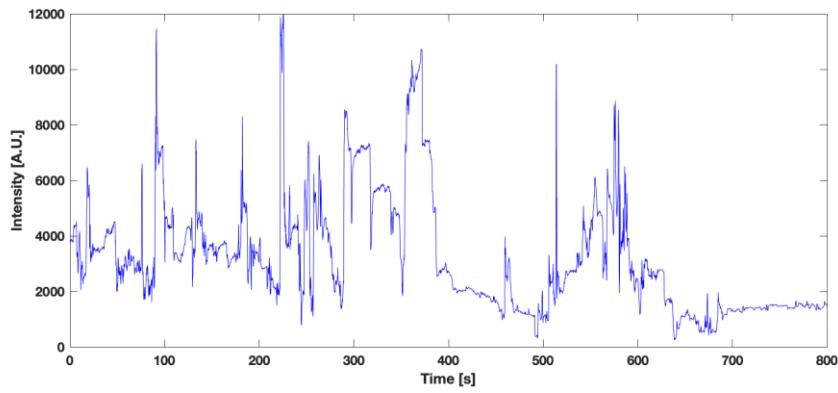
If the number of photons reaching the detector is smaller or equal to the noise floor of the system, hence the measured signal is indeed white noise, the channel did not bring any useful information and should be rejected.

Likewise, if the number of photons reaching the detector is approximately equal to the saturation limit, the channel did not measure any useful information and should be rejected. One reason this might happen is when the coupling between the source and the scalp is not perfect, causing what is called *light piping* (light emitted by the source and directly measured at the detector without travelling through the head). Another cause of this problem is the small distance between the source and the detector. The signal measured is the one reported below:



In general the signal presents some downwards deflection when the *light piping* is activated. Both noise floor and saturation limit depend on the system used and the commercial devices chosen.

Another reason why channel could be rejected is because of low signal to noise ratio when the signal is unstable due to bad scalp – optode coupling. Sometimes optodes could not be fix very well to the scalp through the cap and thus creates a such variation of the signal pattern. A characteristic shape of example is the situation reported below.



This is the first step that every NIRS analysis need to undergo. This because looking first carefully to the data is important to understand if such compounds are relevant or not if the acquisition was performed well or not and if the successive processing steps are executed correctly or not.

NIRS: Standard processing pipeline – Convert to optical density

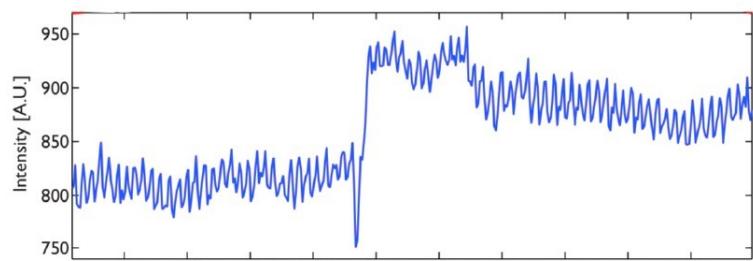
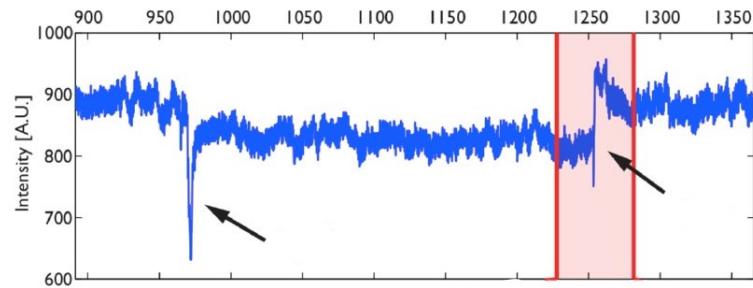
The second step is to convert all to optical density. To perform this operation we need to follow the next equation. In this case, usually, I_1 is set as the mean across time of the measured intensity which is considered a good normalization baseline. While I_2 represents the signal considered. The relation that needed to be used is the one reported below.

$$OD = -\ln \frac{I_2}{I_1}$$

This step is necessary because the optical densities are normalized as zero mean signals while the initial signals are not. Thus, starting from channels recording with different and multiple amplitudes thanks to the application of the ratio and the logarithm we could get a normalization of all the optical density signals which will show a much more restricted range of variation.

NIRS: Standard processing pipeline – Motion artifacts: Introduction

The following step consists in the motion artifacts correction and detection. Usually the reason why the motion artifact are presents in the signals that we analyse is because the fibers that connects each element inside the cap reaching its extensional limit and after this threshold could move the different sources and detectors. This will cause inside the signal a lot of artifacts that changes dramatically the amplitude of the signal. This type of artifact inside the signal could be presented as spikes or steps with a certain amplitude different from the baseline of the signal. Down below are highlighted two phenomenon of motion artifacts.



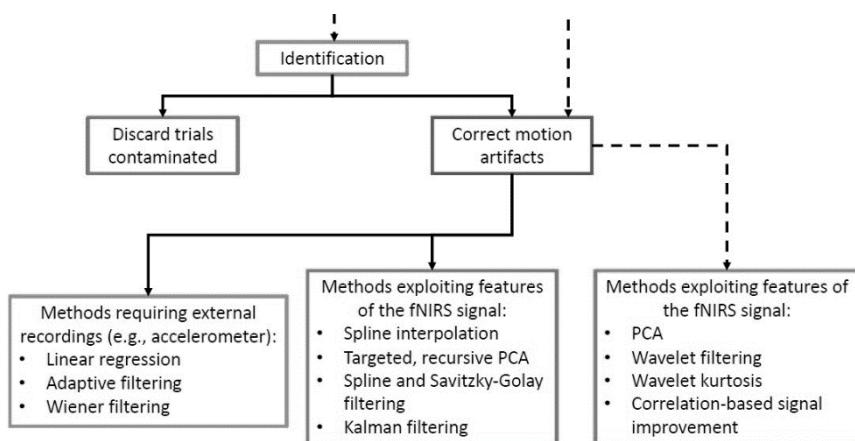
From the two images reported above we could appreciate in the image on the left the optical density signal that shows two motion artifacts highlighted thanks to the usage of an arrow. While in the image reported on the right, we could see the presence of the zoom on the red marked area of the signal reported on the left. The first motion artifact is a spike while the second is a step. In the first case we measure a sudden change in the position of the detector. Instead the step could let us be able to reach a higher level of intensity detected. In the image reported before we could see that there is a first spikes movements and then due to the movement of the detector or the channel this is then maintained.

NIRS: Standard processing pipeline – Motion artifacts: Possible sources

The motion can also induce slower changes which are more difficult to discriminate from physiological changes and very challenging to reduce. The worst movements during fNIRs acquisitions are those creating a relative displacement of the cap optodes from the scalp. Some instances of these actions are: reading aloud, nodding up down (annuire), nodding sideways, twisting right, twisting left, shaking head rapidly and finally the worst is raising the eyebrows.

NIRS: Standard processing pipeline – Motion artifacts: Techniques Overview

Inside the situations where we could detect motion artifacts, we would like to correct them instead of removing all the possible signals from the analysis. The strategy is not to exclude the whole signal, but we should use some variability extraction inside the time interval where the motion artifacts are detected to remove these from signal. The first thing that we could try to correct and handle the motion artifacts after the identification phase, is that we could discard trials that are contaminated from these or we could correct the same by using also some additional information coming from some head movement detection instrumentation. In general the instrumentation is made up of gyroscope and accelerometer that detect the acceleration and the direction of the head movement. These data before being used need to be processed inside the computer. Some of the possible methods that could be used to reduce the motion artifacts inside the NIRS signals are the ones reported below in the scheme.

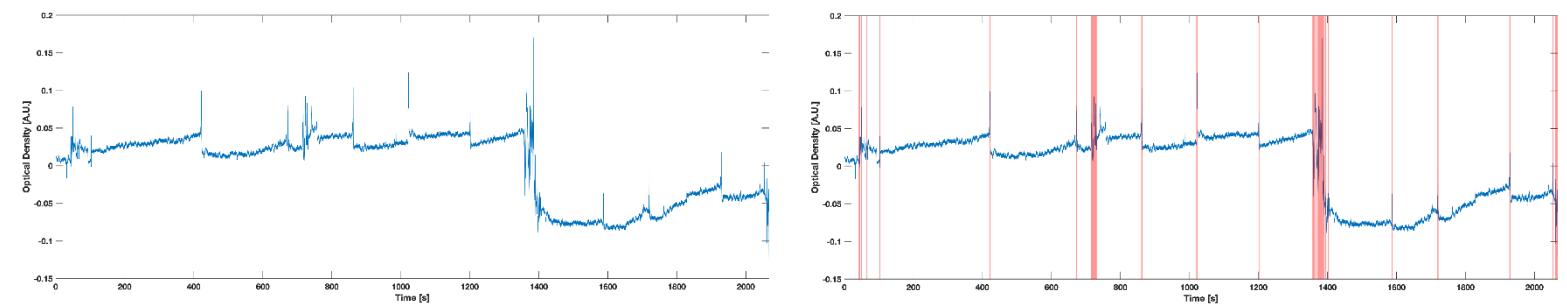


NIRS: Standard processing pipeline – Motion artifacts: Identification

As reported in the scheme highlighted before the motion artifacts have first to be identified. Then they could be corrected by the two previous main methodology that we have mentioned. Now we need to discuss on

how we could perform the identification steps. This could be either done manually or automatically thanks to the identification and quantification of the magnitude changes inside the signal. Generally the different motion artifacts could be seen as rapid intensity signal changes. An example of automatic algorithm that could be used is to compute the standard deviation of a finite interval of the signal before the instant of time considered and compare the value obtained with the standard deviation of the signal in a certain interval of time. If the value is higher than this threshold then we could say with a certain degree of accuracy that the signal in that instant of time presents a motion artifact and thus we marked the instant of time for which the artifact is activated. In general this comparison of the standard deviation could be enriching with also some intensity thresholds.

To recap an automatic way to implement this inside an algorithm, we need first to set the size of the sliding window, then we set the threshold for the amplitude and the threshold for the standard deviation computed in the signal compared with the overall condition. Down below just to highlight some results we could appreciate the application of this algorithm to the signal reported on the left.



NIRS: Standard processing pipeline – Motion artifacts: Discarding

The quickest approach is to discard the different trials contaminated by motion artifacts. By doing so the problem is that the reliability of the signal to reconstruct the final HRF is much lower than correct these particular time intervals affected by these artifacts. This is however unfeasible when few trials are available as instance with infants and clinical populations and entails losing useful information.

NIRS: Standard processing pipeline – Motion artifacts: Correction

The best approaches are the ones that try to correct the motion artifacts instead of discarding the interval of time where the motion artifacts are detected. Specifically every signal can be corrected with several methods which different one requiring pre identification of motion artifacts or not. The two main classes are the one reported below.

Requiring pre – identification of motion artifacts

Spline interpolation, target recursive PCA, Spline and Savitzky – Golay filtering, Kalman filtering.

Not requiring pre – identification of motion artifacts

Principle component analysis, Wavelet filtering, Wavelet kurtosis and Correlation based signal improvement.

All these techniques do not differ only for the a priori information that are reported previously but also because some techniques work more efficiently in presence of specific motions artifacts than another. In the next paragraph we will treat deeply the spline interpolation, the target recursive PCA and finally the wavelet filtering. As we will see the best method to achieve the best results available between these three techniques is the wavelet filtering.

NIRS: Standard processing pipeline – Motion artifacts: Correction – Spline interpolation

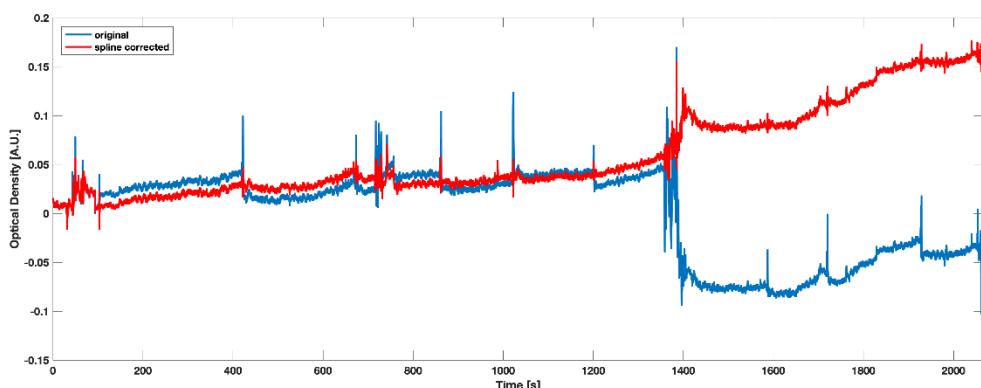
The spline interpolation motion correction method works in five different steps that we will see in detail in the next paragraphs. The first one is to identify the motion artifacts in signal $x(t)$. The second step of the spline is to segment the original signal with part of motion artifact and part not affected from motion artifact. Thus at the end of this procedure we get two distinct matrixes, the one containing only motion artifacts and the one that does not contain any motion artifact. The mathematical representation of these two structures is the one reported below.

$$x_{MA}(t) = \{x_{MA,k}(t)\}, k = 1, \dots L \quad x_{NMA} = \{x_{NMA,k'}(t)\}, k' = 1, \dots L'$$

Where L represents the number of motion artifacts identified, while L' represents the number of segments without motion artifacts. The acronym MA stand for motion artifact affected, while NMA non affected from motion artifact segment. The third step is to apply a cubic spline interpolation of each k -th segment stored in x_{MA} . The spline function is reported below:

$$\min p \sum_{j=1}^n w_j |y_j - f(x_j)|^2 + (1-p) \int \lambda(t) |D^2 f(t)|^2 dt$$

Where the first part of the written represents the error in the measure between the estimation and the real data, while the second represents a roughness measure. This last term is representative for the regularization matrix. The final step is to subtract the spline interpolated function to the original signal in the interval considered and interpolated with this technique. Then to connect the two section of the signal before and after the interval of time where is performed the interpolation, each segment is parallel shifted with respect to its mean value and the mean value of the previous segment.

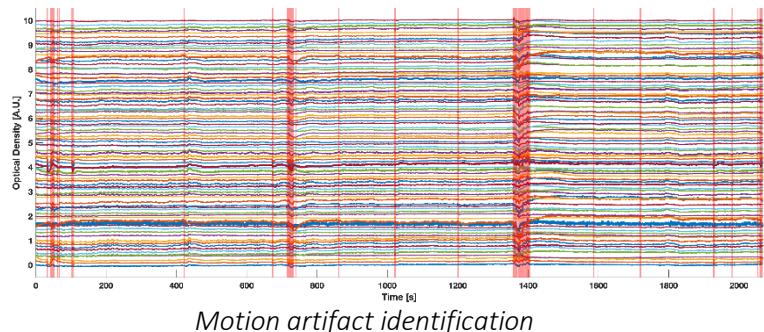
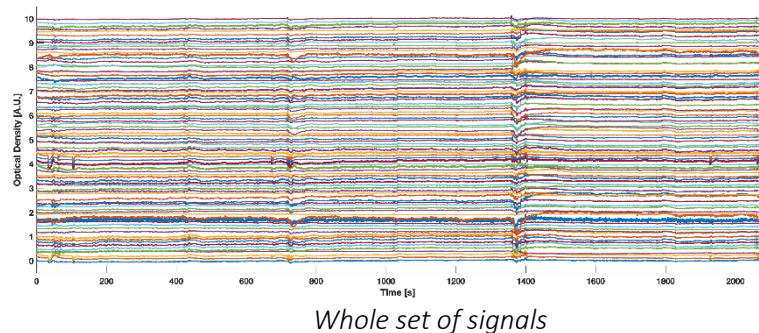


As we could see from the image reported above after having applied the spline correction, we could insert some drift of the baseline due to the mean correction before and after the artifact detected.

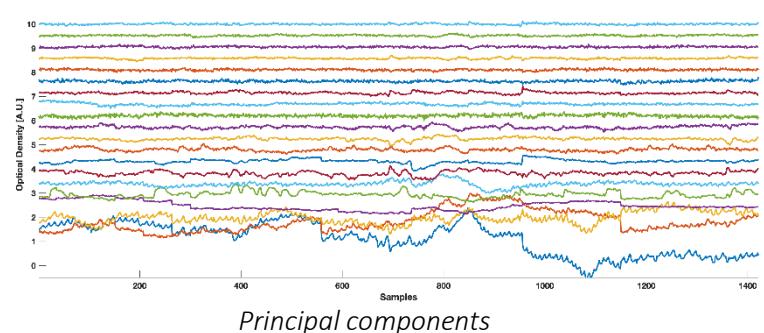
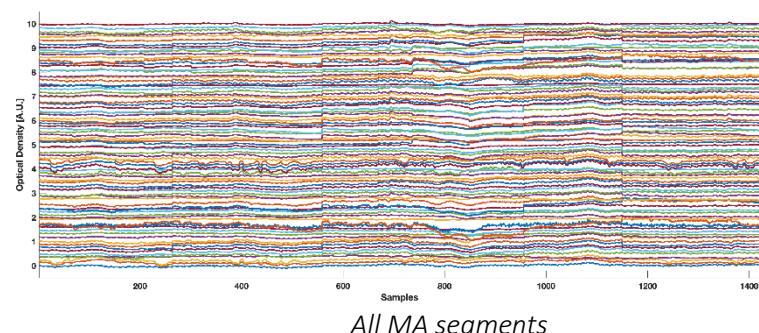
The advantage of this technique is that it is simple, it is fast, it works at the single channel level because we need to consider once channel per time and only when we conclude this analysis we could move further. It could correct only the possible motion artifacts and anything else. It also corrects the eventual signal offset due to the movement of the optodes in the scalp. However there are some limitations as instance it requires reliable technique to identify motion artifacts. Usually a fixed parameter of the cubic spline function might be suboptimal for all the motion artifact segments and the entire information of the signal during the motion artifact is lost due to this correction. The last limitation of this method is that it leaves some high noises spikes inside the result.

NIRS: Standard processing pipeline – Motion artifacts: Correction – Target recursive PCA

This method exploits the PCA and is made up exactly of seven steps. The objective of this algorithm is to apply the PCA and to do so we exploit the information coming from the whole set of signals measured. The first step consists in identify the different artifacts inside the signals. Once do this we need to build up the matrix where we would like to perform the PCA. Thus, assume that a certain instant t a certain movement artifact could be detected in a certain subset of channels. In this case, we consider the time interval in the whole set of channels available where it is in the subset detected the motion artifact. Thus, we could already understand that this method will be perfect in those cases where the motion artifacts are presents inside whole signals and not just in a subset.



The second step consist in concatenate the different motion artifact segments that made up the considered signals and we could in this way identify a matrix of elements where we would like to perform tPCA. The step number three consist in compute the different principal components. Down below are highlighted the matrix with all the motion artifact components and the matrix with just the principal components computed.



The principal components are computed via eigen vector decomposition of the covariance matrix. The number of principal components that are computed is equal to the number of channels that we considered.

In this case if we would like to formalize the principal components decomposition, we need to use the next analytical relation. First, we consider the matrix Y of dimension $n \times p$ of data. Where n represents the number of samples while p the number of channels. The covariance matrix of dimension $p \times p$ is exactly equal to:

$$C = Y^T Y$$

Then we could represent the decomposition of this matrix with the next analytical structure.

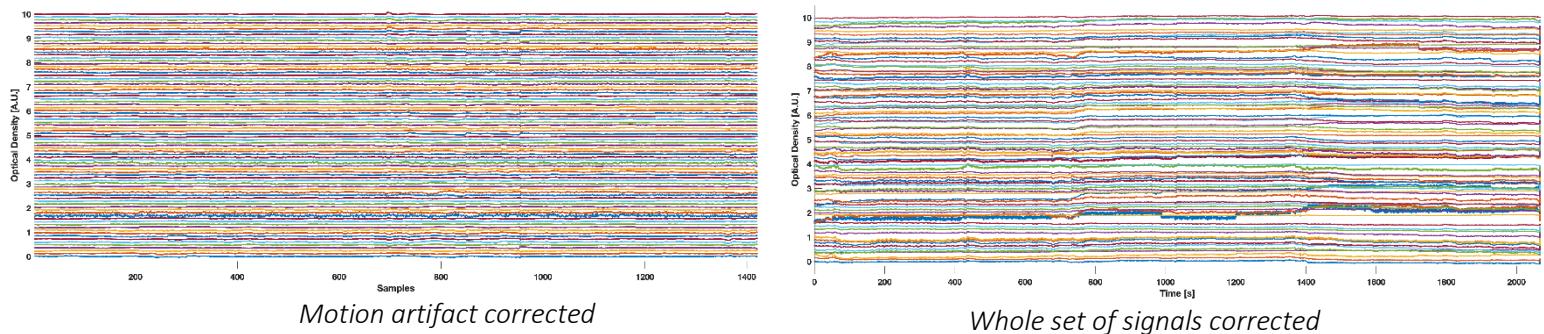
$$C = V L V^T$$

Where V is the matrix of principal eigenvectors while the matrix L is a square diagonal matrix. The diagonal elements are equal to the eigenvalues associated to the eigen vectors. Thus the principal component matrix defined with the acronym PC could be calculated as:

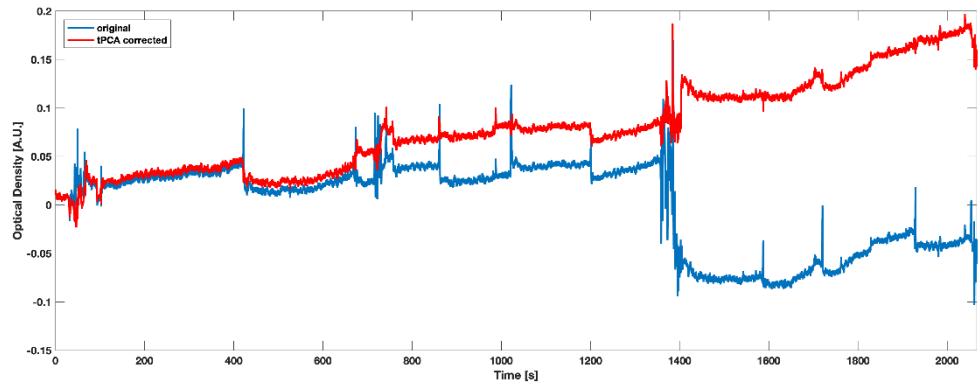
$$PC = Y V$$

Once computed all we order the PC according to the variance of the data that they explain. The idea of the tPCA is that motion artifacts should be the main source of variance, therefore they should be present in the first couple principal components. The idea of the PCA is to remove the first M components to remove a certain percentage of variance. The number M depends on the subject considered. Usually M is chosen to let the value of variance be at maximum 97%.

Once performed this filtering phase we need to reconstruct the different segments. The procedure is quite easy. From the PCs that we delete a certain number M of components we reconstruct the initial matrix by the inverse procedure. Now we have the corrected motion artifact matrix which contains all the time interval of the whole set of signals where were identified different artifacts. Hence, the only step further is to replace the same time interval with the corrected interval of time filtered. This procedure has to be redone until the motion artifact are not more presents inside the final results. Down below we report an image of the corrected motion artifacts and the final matrix of the signal corrected from these.



We first have to identify all the MA segments. This procedure continues until a certain number of iteration or until a certain change. In this case there is a high change. Part of the signal that following the segment there is changes in the baseline of the signa could not be identified. To compare the results obtained with the tPCA and the spline method we report the results obtained by using the tPCA technique on the same signal elaborate with spline.



The tPCA algorithm for motion correction it is simple, fast and very efficient when motion artifacts are present in almost all channels. Moreover, because we correct only the motion artifact of the different signals, we need to perform some off set corrections. To work correctly the technique need to use recording that presents a lot of motion artifact inside the whole set of channels. Another limitation is that the motion artifact identification algorithm needs to be very reliable. This method required also to determine the amount of variance that should be tried to explain.

```
1 %
2 % Title: Laboratory 8
3 %
4 % Note:
5 % fNIRS analysis visualization and data manipulation
6 %
7 % Author: Matteo Martin
8
9 % THEORY -----
10 %
11 % HOMER3
12 % Homer3 could be run also with a GUI but we will write our own code to
13 % perform the different and more accurate analysis. This because could be
14 % more easy to run specific data processes and could be specialized
15 % for every type of dataset. In general Homer3 could be also open with the
16 % GUI. To run in this way this software we just need to write Homer3 inside
17 % the command window of Matlab.
18 % Inside the pop up window, specifically on the right we could see the
19 % presence of measurement probes and non measurement probes which are
20 % highlighted with two distinct colours.
21 % Inside the chart reported on the left we could clearly see the presence
22 % of just vertical bars if no channels is visualized and selected. The
23 % vertical bars represents when the stimulus onset. In particular in this
24 % case the subject undergo through three different main condition.
25 % To visualize specific time courses we need to select in the right map one
26 % given channel where we would like to perform our analysis.
27 % When we visualize a selected channels we could see in the left chart the
28 % presence of different time courses developed in time. These are of
29 % different colours and refers to a specific connection between the
30 % detector and the measurements probes. Those signal measured could be of
31 % high intensity or low but in general we could see the presence of
32 % physiological and not physiological components inside them.
33 % Another important part of this program is the possibility to select which
34 % of the wavelength to visualize if obviously the file loaded contains
35 % information for two distinct wavelength.
36 %
37 % SNIRF
38 % The .snirf file format is worldwide adopted to store information about
39 % the fNIRS analysis. In general, the .snirf file contains an object, in
40 % Matlab at least, that contains other object in turn. These are:
41 %
42 % - formatVersion: String element that contains information about the
43 % version of the snirf file format.
44 %
45 % - metaDataTags : It is an object of the class MetaDataTagsClass and
46 % contains information about the measurements
47 % procedure and other addition information.
48 %
49 % - data : It is an object of the class DataClass and contains
50 % information about the time series measurements, the
51 % time vector and the measurement list. In the first
52 % object previously mentioned (DateTimeSeries object)
53 % are stored all the time series information. In the
54 % second the time vector and finally the measurements
```

```

55 %
56 %
57 %
58 %
59 %
60 %
61 %
62 %
63 % - stim : Contains information about the events that the
64 % subject undergo during the analysis. It contains
65 % the different type of activity inspected, the onset
66 % of the stimulus, the duration and the consecutive
67 % amplitude once fixed a given scale.
68 %
69 % - probes : This object contains information about the three
70 % dimensional position and localization of the
71 % different probes inside the configuration
72 % considered. It contains both the three dimensional
73 % and the two dimensional position of all the probes
74 % in the set up.
75 %
76 % - aux : This contains some additional information about the
77 % NIRS signals acquisition.
78 %

79 % In Matlab when we need to handle object to see which methods could be
80 % used with a given one we just need to type the name of the object and
81 % then click TAB to find the possible list of methods to use.
82 %

83 % DATA
84 % The data that we use are characterize from a lot of physiological
85 % knowledge and information inside them. In fact, even from the inspection
86 % with the Homer3 software it is possible to see that the different signal
87 % presents an oscillation due to the heartbeat and a major oscillation due
88 % to the breath. The probes configuration is shows inside the PDF contained
89 % inside the same folder of this file. In particular here in the image it
90 % could be seen the differences between the different probes marker in red.
91 % Someone in fact shows blue circle that surrounds the red dot. These one
92 % are the so called Short Separation channels. This means that the distance
93 % between the probe and the detector is lower than usual. Sometimes this is
94 % necessary because help a lot to perform and control some given
95 % measurements. They have additional information that we will see during
96 % successive lectures.

97
98 %% CCC
99 % Clear all, close all, clear the command window
100
101 close all; clear all; clc
102
103 %% DATA - LOAD
104 % Load of the data that are contained inside the folder data. The way that
105 % we could use to load data inside Matlab are various. In particular when
106 % the file to load are located in different folder than the one of
107 % correctly work we need to use the next syntax.
108

```

```

109 % pathnameLoad = 'path';
110 % SNIRF          = SnirfLoad(fullfile(pathnameLoad,'nameOfFile'));
111 %
112 % In this way the loading processs is safer beacuse with the fullfile
113 % function we could let the path written for a given OS be indepedent from
114 % the OS itself. The other method is the one wider used that specify just
115 % the file that needed to be load beacuse it is contained inside the same
116 % work folder.
117
118 SNIRF = SnirfLoad('subj02.snirf');
119
120 %% OBJECT - EXTRACTION
121 % Extraction of the different main object that are needed to correctly
122 % perform the analysis inside our case.
123
124 NIRSDATA = SNIRF.GetData(); % Information of the data
125 STIM      = SNIRF.GetStim(); % Information of the stimulus
126 PROBE     = SNIRF.GetSd();   % Information on the probes configuration
127
128 % the next object so created have specific meaninfull fields such as:
129 %
130 % - Data Class:      dataTimeSeries: [21423x168 double]
131 %                      time: [21423x1 double]
132 %                      measurementList: [1x168 MeasListClass]
133 %
134 %                      The first object contains information about the
135 %                      Time series measured and acquired through NIRs
136 %                      technology. The second object is simply a vector
137 %                      that hold information about the time instant and
138 %                      finally the measurementList which we have described
139 %                      before in the THEORY section.
140 %
141 % - Stim Class:       name
142 %                      data
143 %                      dataLabels
144 %                      states
145 %
146 %                      The first filed contains information on the type of
147 %                      stimulus given to the subject involved in the
148 %                      analysis. The data field instead contains
149 %                      information about the time when the stimulus is
150 %                      given and then the other two some additional
151 %                      information about the stimulus given to the
152 %                      patient.
153 %
154 % - Probe Class:      wavelengths: [2x1 double]
155 %                      wavelengthsEmission: []
156 %                      sourcePos2D: [16x3 double]
157 %                      detectorPos2D: [23x3 double]
158 %                      landmarkPos2D: []
159 %                      sourcePos3D: [16x3 double]
160 %                      detectorPos3D: [23x3 double]
161 %                      frequencies: 1
162 %                      timeDelays: 0

```

```

163 %                               timeDelayWidths: 0
164 %                               momentOrder: []
165 %                               correlationTimeDelays: 0
166 %                               correlationTimeDelayWidths: 0
167 %                               sourceLabels: {16×1 cell}
168 %                               detectorLabels: {23×1 cell}
169 %                               landmarkLabels: []
170 %
171 % The whole information contained inside the object
172 % of this class does not need any explanation
173 % beacuse they are really well characterized from the
174 % name of the different properties object name.
175
176 %% TIME SERIES - EXTRACTION
177 % Extraction of the time series from the Data object load in the previous
178 % step. One additional information that we calculate is the sampling
179 % frequency. This last quantity is simply derivable from the time vector.
180
181 I = NIRSDATA.GetDateTimeSeries();    % Information on the intensity
182 T = NIRSDATA.GetTime();            % Information on the time instant
183
184 %% DATA - VISUALIZATION (FULL)
185 % Visualization of the data inside dedicated graphical window. The
186 % visualization is differentiated between the two distinct wavelength
187 % analysed.
188
189 figure, subplot(1,2,1),
190     plot(T,I(:,1:end/2)),
191     title('Wavelength 1'),
192     xlabel('Time [s]'), ylabel('Intensity [A.U.]')
193     xlim([T(1) T(end)]), grid on, grid minor, box on
194
195     subplot(1,2,2),
196     plot(T,I(:,end/2+1:end)),
197     title('Wavelength 2')
198     xlabel('Time [s]'), ylabel('Intensity [A.U.]')
199     xlim([T(1) T(end)]), grid on, grid minor, box on
200     sgttitle(sprintf('Intensity signals (TOT = %i)',size(I,2)))
201
202 % WAVELENGTH 1
203 % From this visualization we could see that we have a very widespread
204 % dataset. Specifically, we have a lot of signals that are also currupted
205 % from the noise, physiological artifacts and motion artifacts. Inside the
206 % signals we could also appreaciate some physiological componenents like
207 % the heart beat and the respiration trend.
208
209 % WAVELENGTH 2
210 % On the other hand, or in the second signal chart visualziation we could
211 % see signals of a larger amplitude. This is very important to be notice
212 % beacuse reflects a situation where the tissue is healthy beaucse the
213 % light to an higher wavelength is less absorbed from the biological
214 % tissue. Also in this chart we could see some physiological components
215 % like the heartbeat and the respiration.
216

```

```

217 %% DATA - VISUALIZATION (SINGLE)
218 % Visualization of a single channel signal at the two distinct wavelengths.
219 % This is needed to compare the behaviour of the tissue at correspondence
220 % to the two wavelenght in time.
221
222 chSel = 65; % Channel to visualize
223
224 figure, plot(T,I(:,[chSel, chSel + 84])), %
225     title(sprintf('Channel %i',chSel))
226     xlabel('Time [s]'), ylabel('Intensity [A.U.]')
227     xlim([T(1) T(end)]), grid on, grid minor, box on,
228     legend('Wavelength 1','Wavelength 2')
229
230 %% DATA - SELECTION
231 % Selection of those channel that presents a mean value inside the range
232 % between 0.001 and 2.5 and with a minimum signal to noise ratio equal to
233 % 2. These information are passed through a function where are computed all
234 % the values.
235
236 meanTH = [0.001 2.5]; SNRTH = 2; % Options
237
238 selCh      = noiseFreeChannel(I, meanTH, SNRTH); % Signal validation
239 nCHS        = sum(selCh)/2; % Number of selected
240 nCH         = length(selCh)/2; % Number of selected
241 selChW1    = find(selCh,nCHS,'first'); % Index wavelength 1
242 selChW2    = find(selCh,nCHS,'last'); % Index wavelength 2
243 mlAct{1}   = selCh; % Homer notation
244
245 %% DATA - VISUALIZATION (FULL)
246 % Visualization of the data filterd by channel inspecintg the properties
247 % previously highlighted.
248
249 figure, subplot(1,2,1),
250     plot(T,I(:,selChW1)),
251     title('Wavelength 1'),
252     xlabel('Time [s]'), ylabel('Intensity [A.U.]')
253     xlim([T(1) T(end)]), grid on, grid minor, box on
254
255     subplot(1,2,2),
256     plot(T,I(:,selChW2)),
257     title('Wavelength 2')
258     xlabel('Time [s]'), ylabel('Intensity [A.U.]')
259     xlim([T(1) T(end)]), grid on, grid minor, box on
260     sgttitle(sprintf('Intensity signals (TOT = %i)',sum(selCh)))
261
262 % We could appreciate that the total number of signals that are considered
263 % is decreased thanks to the previous data visualization. In particular any
264 % signal is rejected beacuse presents a mean value higher than 2.5 and any
265 % signal is rejected beacuse presents a lower value of SNR than 2. The only
266 % signals that are not more considered are the one that presents a mean
267 % value below 0.001. Which is the threshold to distinguish a measure and a
268 % while noise acquisition of the introstrumentation.
269
270 %% OPTICAL DENSITY - COMPUTATION

```

```

271 % Computation of the optical density by first taking the absolute value of
272 % the whole measurements and than by taking the logarithm of the ratio
273 % between the absolute value of the record over its mean value.
274
275 OD      = -log(abs(I)./mean(I,1));    % Optical density computation
276
277 NIRSOD = DataClass();                  % Object generation
278 NIRSOD.SetTime(T);                    % Object generation: Time vector
279 NIRSOD.SetDataTimeSeries(OD);          % Object generation: Time series matrix
280 NIRSOD.SetM1(NIRSDATA.GetM1());        % Object generation: M1
281 NIRSOD.SetDataTypeDOD();              % Object generation: Optical Density
282
283 %% DATA - VISUALIZATION (FULL)
284 % Full data visualization of the optical density calculated at the step
285 % before by open up a dedicated graphic window.
286
287 figure, subplot(1,2,1),
288     plot(T,OD(:,selChW1)),
289     title('Wavelength 1'),
290     xlabel('Time [s]'), ylabel('Intensity [A.U.]')
291     xlim([T(1) T(end)]), grid on, grid minor, box on
292
293     subplot(1,2,2),
294     plot(T,OD(:,selChW2)),
295     title('Wavelength 2')
296     xlabel('Time [s]'), ylabel('Intensity [A.U.]')
297     xlim([T(1) T(end)]), grid on, grid minor, box on
298     sgttitle(sprintf('Optical Density Signals (TOT = %i)',sum(selCh)))
299
300 % Even with these visualziation we could see the presence of different
301 % physiological component inside the optical density signal. These are
302 % related with the heart beat and with the respiration.
303
304 %% DATA - SELECTION
305 % In this case we need to select those time points that are affected from
306 % motion artifacts and those time points which are not. In this situation
307 % we first evaluate the possible time points that are affected from motion
308 % noise, we then register them in a whole matrix and inside a summary
309 % vector. The paramters that are used to compute correctly the motion
310 % artifacts extraction and detection and the tMotion, the tMask, the
311 % SDThreshold and the AMPThresh. By changing them we could see the changing
312 % in the output of the algorithm.
313
314 tMotion = 0.5; tMask = 1; SDThresh = 12; AMPThresh = 0.5;    % Options
315
316 [tInc, tIncCh] = hmrR_MotionArtifactByChannel(NIRSOD,...           PROBE,[],mlAct,[],...
317                                         tMotion, tMask,...           SDThresh, AMPThresh);
318
319
320
321 %% DATA - VISUALIZATION (FULL)
322 % Visualization of the full data that we have computed the optical density
323 % and that we have selected inside the analysis. Additional information are
324 % proposed beacuse over these signals are represented some additional

```

```

325 % information to see those interval of time neglected.
326
327 MOTION = (tInc{1} == 0)*min(OD(:)) + (tInc{1} == 1)*max(OD(:));
328
329 figure, subplot(1,2,1),
330     hold on,
331     plot(T, OD(:,selChW1)),
332     h = plot(T,MOTION,'r','LineWidth',2);
333     hold off,
334     h.Color = [1 0 0 0.05];
335     title('WAVELENGHT 1')
336     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
337     xlim([T(1) T(end)]), grid on, grid minor, box on
338     ylim([min(OD(:)) max(OD(:))])
339
340     subplot(1,2,2),
341     hold on,
342     plot(T, OD(:,selChW2)),
343     h = plot(T,MOTION,'r','LineWidth',2);
344     h.Color = [1 0 0 0.05];
345     title('WAVELENGHT 2')
346     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
347     xlim([T(1) T(end)]), grid on, grid minor, box on
348     ylim([min(OD(:)) max(OD(:))])
349     sgttitle(sprintf('Motion Artifcat Corrected (TOT = %i)',sum(selCh)))
350
351 %% DATA - VISUALIZATION (SINGLE)
352 % Visualization of the single data that we have computed before in term of
353 % optical density. Visualization of the different instant of time where it
354 % is activated a motion artifact correction.
355
356 chSel = 65;
357
358 m      = min(OD(:));
359 M      = max(OD(:));
360 MOTION = (tIncCh{1}(:,[chSel, chSel + nCH]) == 0)*m + ...
361           (tIncCh{1}(:,[chSel, chSel + nCH]) == 1)*M;
362
363 figure, hold on,
364     plot(T,OD(:,[chSel, chSel + nCH])), 
365     h = plot(T,MOTION,'r','LineWidth',2);
366     h(1).Color = [1 0 0 0.05];
367     h(2).Color = [1 0 0 0.05];
368     hold off
369     title(sprintf('Channel %i',chSel))
370     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
371     xlim([T(1) T(end)]), ylim([m M]),
372     grid on, grid minor, box on,
373     legend('Wavelength 1','Wavelength 2')
374
375 %% DATA - SPLINE INTERPOLATION
376 % Motion correction through the spline interpolation technique.
377
378 NIRSDOMCS = NIRSDO.copy(); p = 0.99;

```

```

379
380 NIRSODMCS = hmrR_MotionCorrectSpline(NIRSODMCS, mlAct, tIncCh, p, 1);
381 ODMC      = NIRSODMCS.GetDataTimeSeries();
382
383 %% DATA - VISUALIZATION (SINGLE)
384 % Visualization of the effect of the spline interpolation on the data that
385 % are available. These data are contained inside the NIRSODMC structure.
386 % This is a snirf object that contains the usual information. To compare
387 % the results of the spline interpolation we put together two distinct
388 % figures.
389
390 figure, subplot(1,2,1)
391     hold on,
392     plot(T,OD(:,chSel)),
393     plot(T,ODMC(:,chSel)),
394     h = plot(T,MOTION(:,1),'r','LineWidth',2);
395     h.Color = [1 0 0 0.05];
396     hold off
397     title('WAVELENGHT 1')
398     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
399     xlim([T(1) T(end)]), ylim([m M]),
400     grid on, grid minor, box on,
401     legend('Non Corrected','Corrected')
402
403     subplot(1,2,2)
404     hold on,
405     plot(T,OD(:,chSel + nCH)),
406     plot(T,ODMC(:,chSel + nCH)),
407     h = plot(T,MOTION(:,2),'r','LineWidth',2);
408     h.Color = [1 0 0 0.05];
409     hold off
410     title('WAVELENGHT 2')
411     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
412     xlim([T(1) T(end)]), ylim([m M]),
413     grid on, grid minor, box on,
414     legend('Non Corrected','Corrected')
415     sgttitle(sprintf('Channel %i',chSel))
416
417 %% DATA - SPLINE INTERPOLATION
418 % Motion correction through the spline interpolation technique.
419
420 NIRSODMCP = NIRSOD.copy();
421 tMotion   = 1;    tMask      = 2;
422 SDThresh  = 10;   AMThresh   = 0.3;
423
424 NIRSODMCP = hmrR_MotionCorrectPCArecurse(NIRSODMCP, PROBE, [], mlAct, ...
425                                         tInc, tMotion, tMask, SDThresh, AMThresh, 0.97, 5, 1);
426
427 ODMC      = NIRSODMCP.GetDataTimeSeries();
428
429 %% DATA - VISUALIZATION (SINGLE)
430 % Visualization of the effect of the spline interpolation on the data that
431 % are available. These data are contained inside the NIRSODMC structure.
432 % This is a snirf object that contains the usual information. To compare

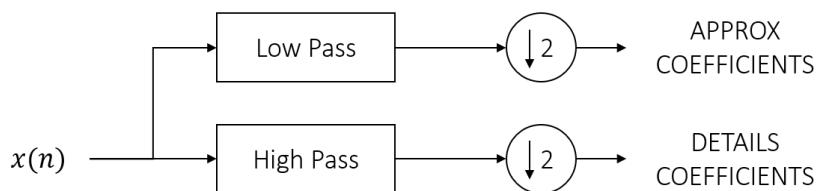
```

```
433 % the results of the spline interpolation we put together two distinct
434 % figures.
435
436 figure, subplot(1,2,1)
437     hold on,
438     plot(T,OD(:,chSel)),
439     plot(T,ODMC(:,chSel)),
440     h = plot(T,MOTION(:,1),'r','LineWidth',2);
441     h.Color = [1 0 0 0.05];
442     hold off
443     title('WAVELENGHT 1')
444     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
445     xlim([T(1) T(end)]), ylim([m M]),
446     grid on, grid minor, box on,
447     legend('Non Corrected','Corrected')
448
449     subplot(1,2,2)
450     hold on,
451     plot(T,OD(:,chSel + nCH)),
452     plot(T,ODMC(:,chSel + nCH)),
453     h = plot(T,MOTION(:,2),'r','LineWidth',2);
454     h.Color = [1 0 0 0.05];
455     hold off
456     title('WAVELENGHT 2')
457     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
458     xlim([T(1) T(end)]), ylim([m M]),
459     grid on, grid minor, box on,
460     legend('Non Corrected','Corrected')
461     sgttitle(sprintf('Channel %i',chSel))
```

LESSON 21: NIRS STANDARD PROCESSING PIPELINE

NIRS: Standard processing pipeline – Motion artifacts: Correction – Wavelet filtering

The last method is the wavelet filtering approach and is based on the discrete wavelet transformation of the data. This is a time frequency analysis and visualize the data on a global representation thanks to these two distinct terms, the approximation coefficient, and the details one. To filter the different signals we use a low pass and a high pass filter and by using these models we could estimate the approximation and the detail coefficients. Thus, once defined the impulsive response of the high pass filter with $h(n)$ and the impulsive response of the low pass filter indicated with $g(n)$, we could express the details and the approximation coefficient respectively as the convolution between the input signal and these two impulsive responses. This could be formalized in the next way:

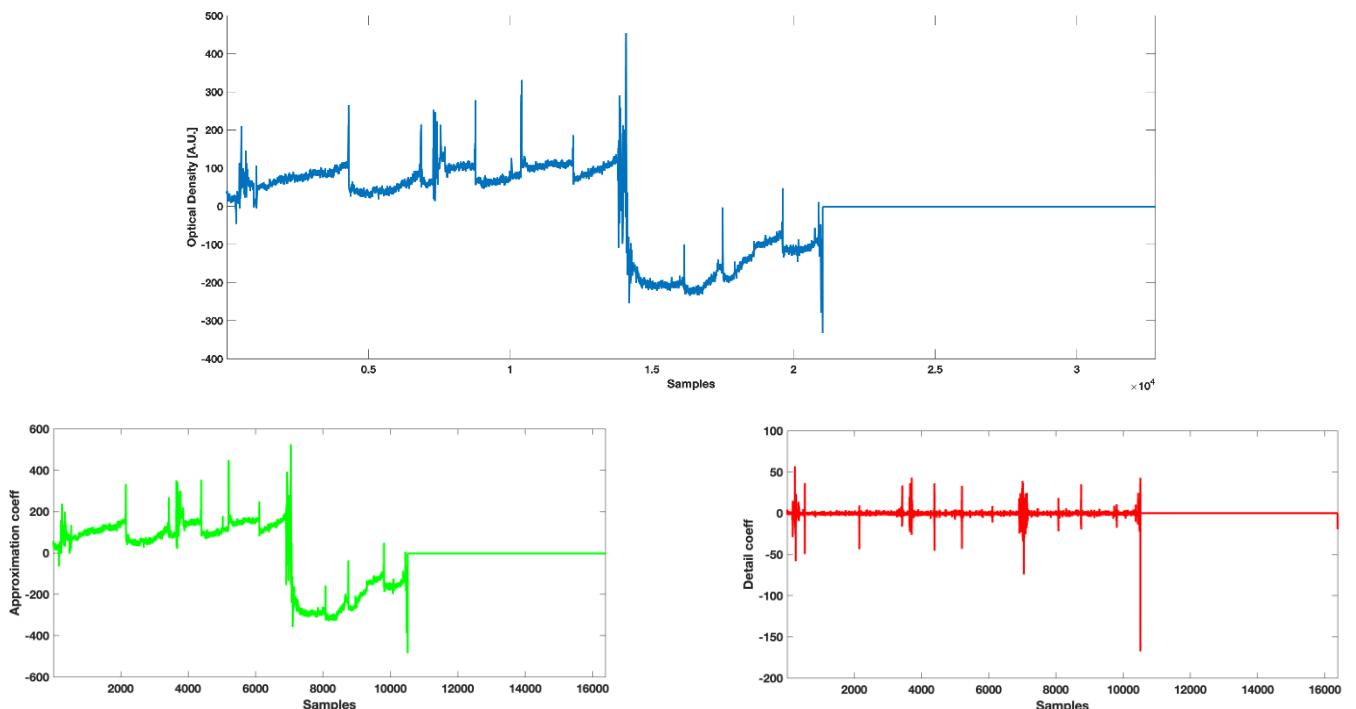


$$y_{low}(n) = x(n) \otimes g(n) = \sum_{k=-\infty}^{+\infty} x(k)g(n-k) \quad y_{high}(n) = x(n) \otimes h(n) = \sum_{k=-\infty}^{+\infty} x(k)h(n-k)$$

Approximation coefficient

Details coefficient

The two signal keeps just a certain frequency content of the signal. The output of the high pass filter is called the *details coefficient* while the other are called the *approximative coefficient*. The cut off frequency of both filters is half the sampling frequency used to sample the signal. In the following image we could see the presence of the approximation and the details coefficient decomposition of the signal mainly considered.



This subdivision could be applied repeatedly thanks to the usage of the same approach and operation chain. This is quite useful because we are decomposing the signal always in an approximation coefficient and detail coefficient. By using this procedure we need to consider the cut off frequency always the half of the frequency band associated to that considered approximation coefficients. This operation could be iterated until a certain splitting criterion is reached (usually the pattern of the signal could be related with the heartbeat and the respiration only). When this happens the process is arrested. Specifically, at this point the approximation coefficients found during the decomposition are tough as be the major contributor to the motion artifacts. All the spikes inside the detail coefficient if they satisfy a certain condition are set to zero to delete the motion artifact from the signal.

To determine if a certain spike should be place to zero or not, we need to estimate the inter – quartile range to choose which spikes comes out from the distribution of the detail coefficient and which one not. This component could be computed easily with the next relation:

$$IQR = q_3 - q_1$$

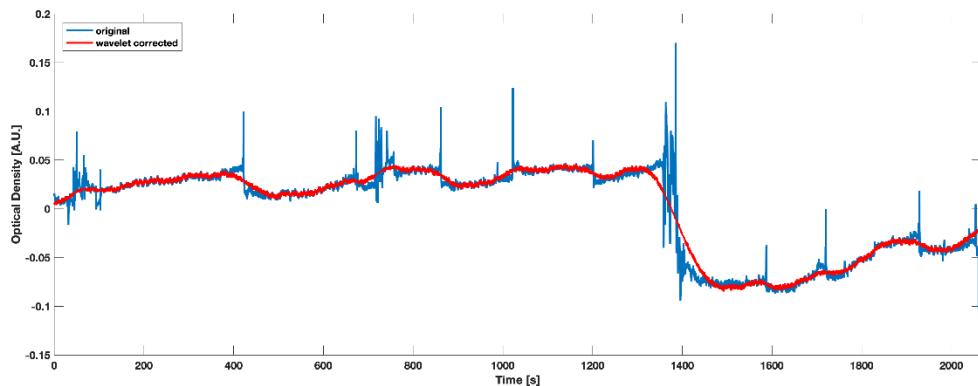
After having compute this element we need to determine the outliers by calculating the upper and lower threshold. The next relation assume that we could represents our detail coefficient distribution through a box plot by visualizing where the different coefficient is placed and if they needed to be discarded or not depending on the two next threshold and the parameter α .

$$THR_1 = q_3 + \alpha IQR$$

$$THR_2 = q_1 - \alpha IQR$$

α is a tuneable parameter for the user and the threshold depends as previously reported also on it. Higher is α more conservative are the two distinct thresholds, while lower is the value of α less conservatives are the two given thresholds. Thus, once select the value of α we need to compute the different thresholds for each detail coefficient and the spikes outside the defined range are placed to zero.

Because of the wavelet transformation is a reversible process, once applied the filtering we need to proceed backward the decomposition to reconstruct the initial filtered signal. This last one has a lower frequency content respect the initial one. To enhance what is the result of applying a wavelet filtering process to a given signal we report down below an example.



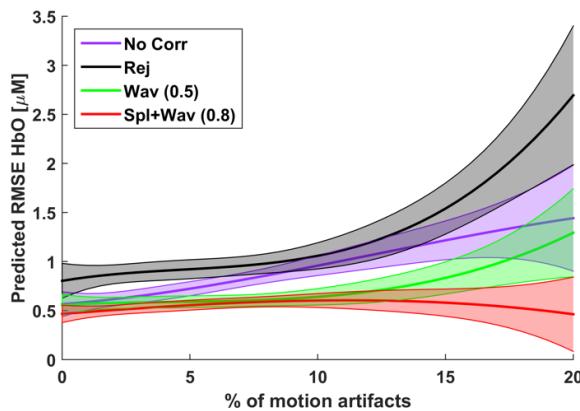
In this image we could appreciate a pattern which represents the wavelet corrected signal and the original one. As we could see from this chart there is not needed any mean adjustment before and after the

identification of the motion artifact interval. The pitfall of this technique is that keeps the changes in baseline.

The advantages is that works at the single channel level instead of PCA and as spline. This is sometimes better because we are correcting the signal only when we detect some motion artifact and moreover there is an identification and correction performed in one step. In term of advantages we have also a better maintaining of the low frequency content of the signal, specifically the physiological content. Spline and PCA instead removes all the possible information. The downside of this technique is that it is much slower than the PCA, and we need to set up manually the threshold to keep certain spikes or not and it is strictly related with the design used. In fact, we need to set the value of α according to the experimental design that we choose. Due to the choice of bad value of α we could even underestimate the haemodynamic responses amplitude. This is not a problem because the decrease in the amplitude happens in all the condition and thus when we perform some statistical analysis the result would not be different. When we have a change in the baseline this could not be corrected because of the not physiological reason to that changes.

NIRS: Standard processing pipeline – Motion artifacts: Correction – Technique to choose

The first thing to do is to qualitatively look at your data and individuate the types and amount of motion artifacts that are presents. There are also some several studies that have compared the performances of the different motion correction techniques on the accuracy of HRF recovery. In the image reported below we could appreciate the different techniques and the different concentration of haemoglobin that we could reconstruct over different channels once applied different motion artifact correction techniques.



In this study we could appreciate some different techniques and the predicted RMSE in HbO once inside the signal measured are presents a low or a high percentage of motion artifacts. In this case we could appreciate that at a low percentage of motion artifact considered inside the signal the wavelet filtering perform worse than the spline and the wavelet technique. This order relation is maintained also in situation where we have higher percentage of motion artifacts. In this case the wavelet performs almost equal to not correct the data.

NIRS: Standard processing pipeline – Band pass filtering: Noise removal

Until now we have kept inside the different signals the physiological information that help us to discriminate between a waveform rich of information and a waveform which is not. In this step of the standard processing

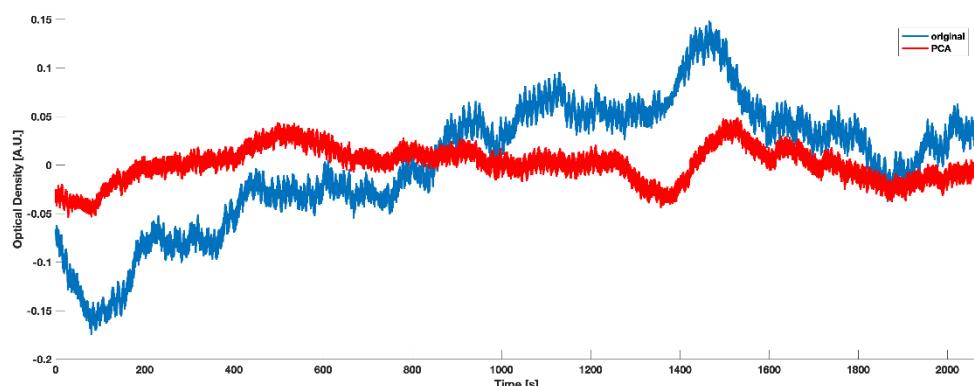
pipeline we need to perform some physiological noise removal which is a step coupled with a band pass filter. The objective of our analysis is that we would like to measure is the haemodynamic response which is task related. Anyway even if the HRF is task related we need to account that there are mainly three task evoked contributions inside our measurements. These are due to the *neural evoked changes in the cerebral compartment* which is we want to detect. The *systemic evoked changes in the cerebral compartment* related with the systemic changes to answer to higher demand of oxygen and finally the *system evoked changes in the extra cerebral compartment*. This last one does not have any contribution inside the fMRI instead. After all these we need to remember that there are also some *non – evoked contribution* that could corrupt the result, and they are changes in cerebral and extra cerebral compartment. These last ones are mainly related with Mayer waves.

In this step of the pipeline for the processing of the fNIRS data we need to remove the noise by applying some band pass filter to the signal. From the noises that have a physiological origin is very important in fNIRS removing *systemic non evoked physiological changes* and *evoked physiological changes*. This last type of physiological artifact inside the signal could cause both false positives and false negatives respectively by mimicking or masking the actual HRF inside the brain. The non – evoked systemic noise, instead, could be reduced and prevented with a good paradigm design as instance using a jittered inter stimulus interval. Thanks to the non – constant inter – stimulus time presentation the systemic responses are characterized of a different value and thus comprehensive in the result they are neglected thanks to the non in – phase contribution of these noises.

Before we have talked to apply a certain band pass filtering. This is the best way to reduce all the possible physiological and not physiological noises that are activated inside our measured signals. The haemodynamic response function is in the most of cases between the 0.03 and 0.1 Hz. This depends also on the stimulus that are presented to the subject. We could easily remove the high frequency content of the signal measured by using a *low pass filter* using cut off frequency of 0.5 Hz. The other filter that could be used is a *high pass filter* where there is a cut off frequency in correspondence of 0.01 Hz. This will delete the very low frequency oscillations and instrumental drifts exempli gratia due to the laser heating.

NIRS: Standard processing pipeline – PCA: Mayer waves removal

With band pass filtering however we cannot remove the entire Mayer waves since they share the same frequency range of the haemodynamic response. The non – evoked systemic changes and, in particular Mayer waves could be reduced by using PCA and removing the first M components. PCA works exactly as the PCA we have seen for motion correction, but instead of applying it to motion corrupted segments, we apply it to the whole signal of all channels. The assumption here is that physiological noise is the main source of variance in the signals and is common across channels.



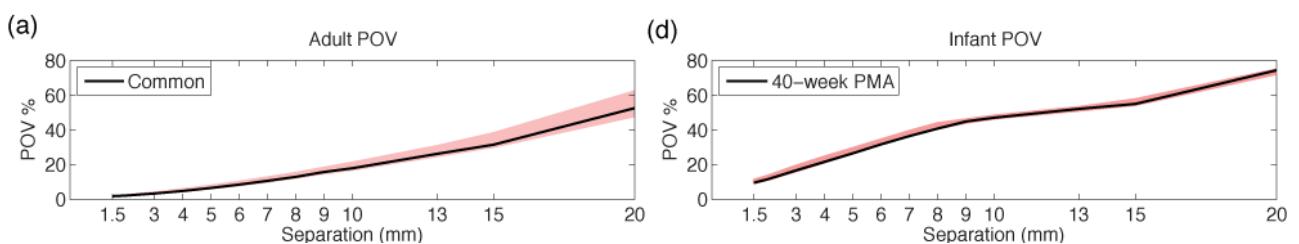
In the image reported above we could appreciate the application of the PCA algorithm to the previous clue signal. We could see that the most of low frequency oscillations due to physiology have been highly reduced. Not all physiology could be reduced or removed since they are not perfectly synchronous across channels and has also been demonstrated that some physiological changes are variable across the head due to vasculature distribution for example.

The PCA technique has some problem because will remove the components that are common across channels. Thus if a haemodynamic response is common in all channels, we could remove also the haemodynamic response as well. Thus the PCA is never used if all channels are expected to be active. When we except activation between different regions, we could not apply this technique. With band pass filtering and PCA we can reduce the non – evoked systemic contribution. This last contribution has some frequency content that it is time locked. To remove the systemic evoked responses other methods has been developed.

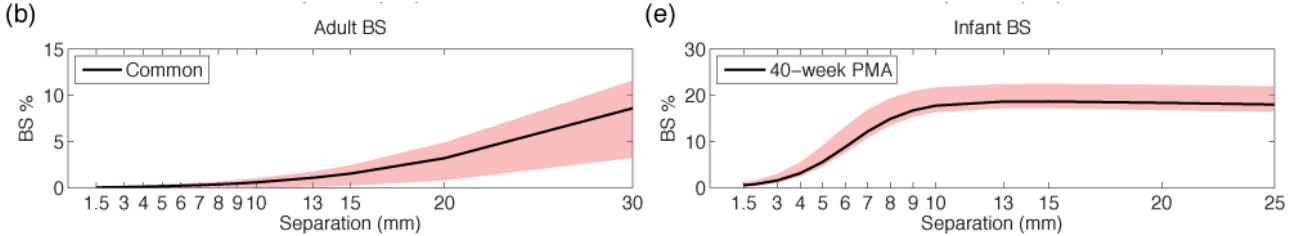
NIRS: Standard processing pipeline – Short Separation Channels: Systemic evoked changes removal

The systemic evoked response instead is solved by looking on the distance between the channels. Specifically if the maximum distance between a source and a detector is of about 3 cm the brain sensitivity is of about 8 % while moving closer the source and the detector the sensitivity of the brain is lower and finally if we move the source and the detector one close each other of about 3 mm the brain sensitivity is of the 0.1 %. In our signal once performed all the possible elaboration explained until now, we could distinguish between the cortical and the extracellular contribution measure. In particular if we place a detector near the source, we have a signal that is related with the extracellular brain contribution. These channels are called the *short separation channels (SS)*. The idea is to use these signals to remove and reduce the final contribution of these to the considered signal. The problem is to find the right shortest distance between the source and the detector because it depends on some specific factor. This last one is the previous mentioned brain sensitivity. The SS should exhibit zero sensitivity to brain because we would like to regress the extracellular components. Moreover, they needed to be characterized of same sensitivity in scalp and skull tissue as a standard channel.

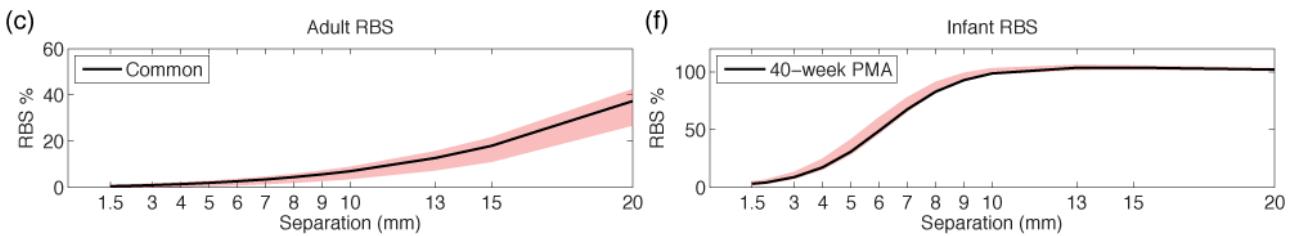
There were some studies of photon stimulation where it was studied even the propagation in head tissues by changing the distance between the source and the detector. It has been possible to study the relationship between the *PMDF (Photon Measurement Density Function)* of standard and SS channels. As expected, increasing the source detector separation increases the *percentage of overlapping voxel (POV)* between SS channel and the channel that are placed to a higher distance. In infants has been observed that increasing the separation between SS channel will induce to increasing the value of POV. This is shown in the figure reported below.



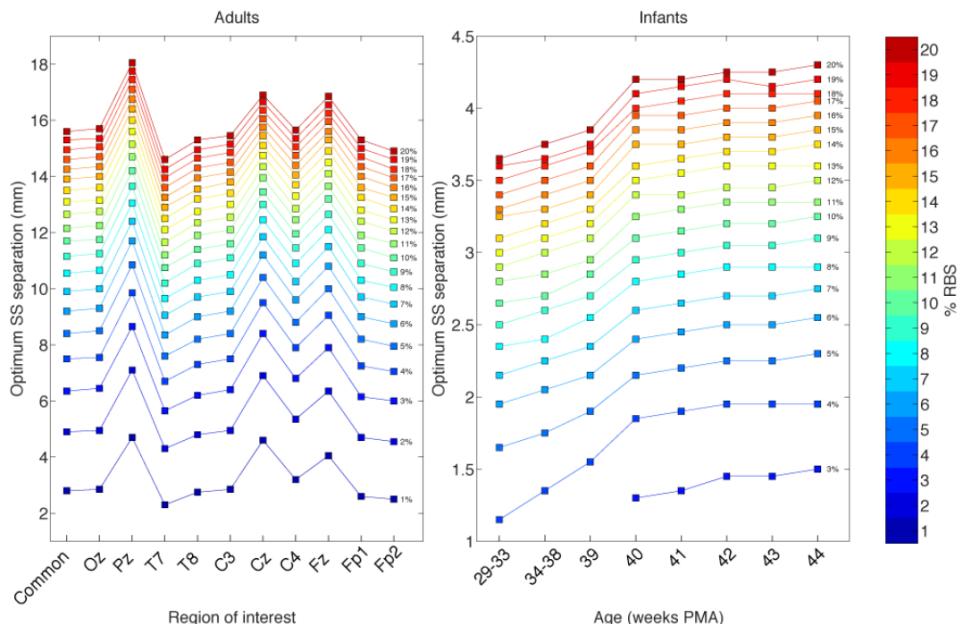
The *brain sensitivity* (*BS*) of the *SS* channel increases as source detector distance increases. In the adult the brain sensitivity increases as the separation between the source and the detector of the short separation increases. What we could notice is that the brain sensitivity increases quite steeply after 15 mm of separation between *SS*. For the infants instead for standard channel the brain sensitivity increases as the separation increases. But the astonishing thing is that even 10 mm of separation between source and detector of *SS* let the brain sensitivity reaches the maximum value.



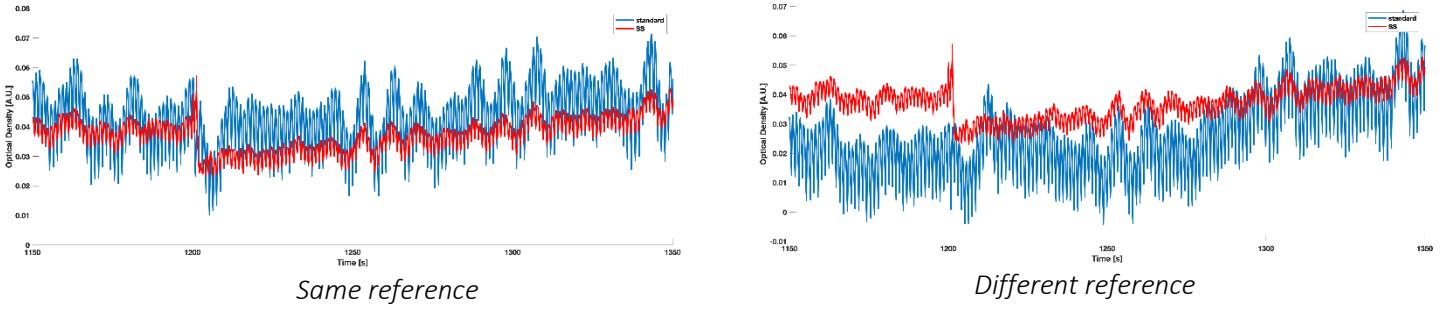
Another index that has been proposed and computed in this study is the *relative brain sensitivity* (*RBS*). This is defined as the brain sensitivity of a short separation channel over a standard one. This confirms the results previously found. They are more informative than anything else.



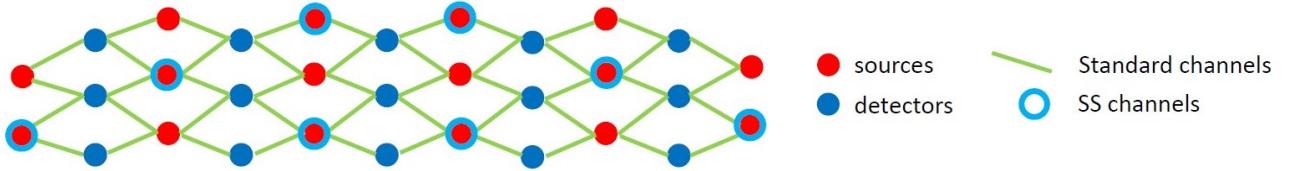
We could create a table that tells which the optimum distance between the *SS* channels varying and depending on the brain location. In the next chart we have reported the *relative brain sensitivity* that can be considered as the average proportion of true functional brain signal that may erroneously be removed when regressing the selected *SS* channels signal from the standard channel signal. The minimum distance between a source and a detector is of about 2 mm and this is due to the physical hardware limitations. The idea to use this chart is to identify where we need to place our electrodes and then choose the value of RBS that we would like to achieve. This positioning will give a certain separation. The values changes because the scalp thickness changes too by selecting different position.



The signal acquired from the short separation channels shows the same physiological variation of the standard channel and they also present the different motion artifact. If we take the SS channel and a standard one that are placed over two distinct places of the head, we could see that the physiological components of the two signals acquired are similar, but the motions artifacts are not correlated because placed in two different location of the head.



Ideally, we would like to have a specific SS channel for each channel. This is not possible currently due to the limited number of fibers that commercial devices have. Therefor, usually, the short separation channels are spread in the layout to try to cover different areas of the head and have similar pattern with the standard channels that are surrounding them. In the image reported below, representative of an array configuration, we could see the presence of the short separation channels.



In this case we have spread the short separation channels between all the possible location of the array. There exist several methods to regress the SS channel signal from the standard channel one. Some examples are:

- *Least square approaches*
- *State space modelling*
- *Bayesian approaches*
- *Compressive sensing approaches*
- *Empirical mode decomposition approaches*
- *General linear model*

We will go deeper inside the *general linear model framework*. In this method we need to convert to contraction changes the optical density measures and then we apply the GLM. The conversion is quite easy thanks to the usage of the modified beer lambert law. This has the following definition:

$$\ln \frac{I_1|\lambda_1}{I_2|\lambda_1} = mDPF(\tilde{\mu}_{HbO|\lambda_1} \Delta C_{HbO} + \tilde{\mu}_{HbR|\lambda_1} \Delta C_{HbR})$$

$$\ln \frac{I_1|\lambda_2}{I_2|\lambda_2} = mDPF(\tilde{\mu}_{HbO|\lambda_2} \Delta C_{HbO} + \tilde{\mu}_{HbR|\lambda_2} \Delta C_{HbR})$$

Where the term $mDPF$ represents a correction factor and is dependent on the value of the wavelength and other subjective characteristics.

NIRS: Standard processing pipeline – HRF recovery

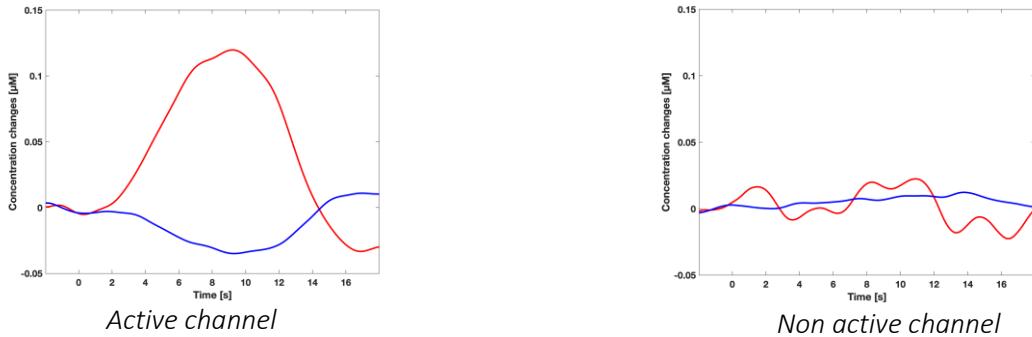
The final step to conclude the pre processing and find the different HRF is performed based on two distinct methods. The first paradigm is the *block averaging* while the second is known as the *GLM*. Usually the block averaging is used when we do not have any short separation channels signal inside our net. Instead in the GLM is used commonly to regress the signal from the content inside the signal recorded in the short separation channels.

NIRS: Standard processing pipeline – HRF recovery: Block averaging

The block averaging consists in averaging trials associated with the same experimental condition. The block averaging uses the weight average between all the haemodynamic response measured. This depends on the channel that we are considering, the chromophore that we are analysing, and the condition considered. The analytical relation that needed to be used to get information about the overall HRF is the next one:

$$HRF_{AVG}(Ch, HbO/HbR, Cond) = \frac{1}{N} \sum_{i=1}^N HRF_i$$

Where HRF_i represents the haemodynamic response of the $i - th$ iteration. In the previous formula it is assumed that we have N different trials. The results that we could obtain is the next one for the condition of active and non – active channels.



NIRS: Standard processing pipeline – HRF recovery: GLM

Instead, if we choose to use also information coming from the short separation channels, we could use the GLM method. In this situation we need to use the next analytical relation:

$$Y = X\beta + \epsilon$$

Where the term Y represents the measured data in concentration changes, X represents the design matrix and it is made up of the different regressor, β are the parameters that needed to be estimated and finally ϵ represents the noise additive to the measure. The regression matrix is made up of different regressor which depends on the function used to model the HRF and some other information coming from physiological

signals or short separation signals. The previous relation could be also described by developing the matrixial product between X and β . This will lead to the next results.

$$Y = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \epsilon$$

Where X_1, X_2, X_3 and so on are the so – called column of the regression matrix.

The model that could be used to fitting the HRF are exactly two, the *gamma* or *double gamma function* and the *gaussian functions*. In the first one the HRF is described mathematically as the analytical relation reported below:

$$HRF = \frac{b_1^{a_1} t^{(a_1-1)}}{\Gamma(a_1)} e^{-b_1 t} - c \frac{b_2^{a_2} t^{(a_2-1)}}{\Gamma(a_2)} e^{-b_2 t}$$

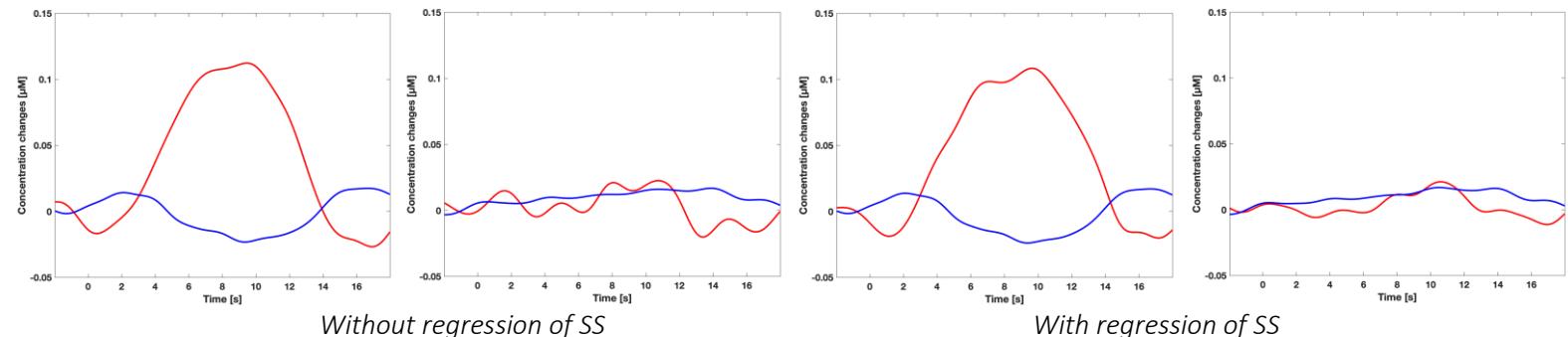
Where if we consider just the first part of the previous function the model used is not more a double gamma function but just a simple gamma one. Each term of the previous written as a specific meaning which are highlighted down below.

b_1 Peak dispersion time constant	a_2 Dispersion time
b_2 Undershoot dispersion time constant	c Ratio peak / undershoot
a_1 Peak time	Γ normalizing factor

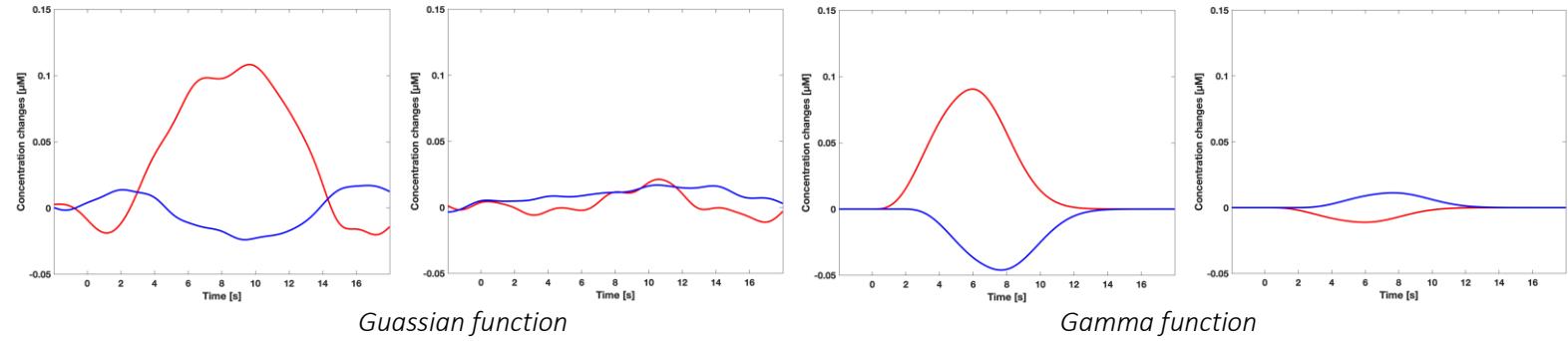
In each column is reported the different HRF that are activated after certain stimulation. In NIRS this is not a common model because we have a higher frequency sampling rate, hence we could detect shapes and not only the peak. Thus it is used a more flexible model known as a set of *Gaussian functions* that could be described with the next mathematical relation:

$$HRF = e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

Where the value of μ represents the mean, while the value of σ represents the standard deviation of the distribution. This pattern is more flexible because allow to estimate shape differences in HRF thanks to the fact that it is based on a more flexible model. By solving the GLM we can estimate a value of β_i for each regressor. The HRF can be reconstructed by multiplying each gaussian function with its estimated weight β . The results differ between the usage or not the short separation channels.



By comparing the set of HRF detected with respectively the Gaussian function and the gamma one we obtain the following results. As could be clearly seen from the figures the two estimation results are quite different.



NIRS data contain serial correlation due to physiology (much more than fMRI) and this is due to the high sampling frequency. In the following lecture we will appreciate the different methodology to avoid this correlation and the method that also in practice it is widely used.

```
1 %
2 %     Title: Laboratory 9
3 %
4 %     Note:
5 %     fNIRS analysis visualization and data manipulation. The theory part in
6 %     this case is delocalized over the script because are reported just the
7 %     main comment to the different operations.
8 %
9 % Author: Matteo Martin
10
11 %% CCC
12 % Clear all, close all, clear the command window
13
14 close all; clear all; clc
15
16 %% DATA - LOAD
17 % Load of the data that are contained inside the folder data.
18
19 SNIRF = SnirfLoad('subj02.snirf');
20
21 %% OBJECT - EXTRACTION
22 % Extraction of the different main object that are needed to correctly
23 % perform the analysis inside our case.
24
25 NIRSDATA = SNIRF.GetData(); % Information of the data
26 STIM      = SNIRF.GetStim(); % Information of the stimulus
27 PROBE     = SNIRF.GetSd();   % Information on the probes configuration
28
29 %% TIME SERIES - EXTRACTION
30 % Extraction of the time series from the Data object load in the previous
31 % step. One additional information that we calculate is the sampling
32 % frequency. This last quantity is simply derivable from the time vector.
33
34 I = NIRSDATA.GetDataTimeSeries();    % Information on the intensity
35 T = NIRSDATA.GetTime();             % Information on the time instant
36
37 %% DATA - VISUALIZATION (FULL)
38 % Visualization of the data inside dedicated graphical window. The
39 % visualization is differentiated between the two distinct wavelength
40 % analysed.
41
42 figure, subplot(1,2,1),
43     plot(T,I(:,1:end/2)),
44     title('Wavelength 1'),
45     xlabel('Time [s]'), ylabel('Intensity [A.U.]')
46     xlim([T(1) T(end)]), grid on, grid minor, box on
47
48     subplot(1,2,2),
49     plot(T,I(:,end/2+1:end)),
50     title('Wavelength 2')
51     xlabel('Time [s]'), ylabel('Intensity [A.U.]')
52     xlim([T(1) T(end)]), grid on, grid minor, box on
53     sgttitle(sprintf('Intensity signals (TOT = %i)',size(I,2)))
54
```

```
55 %% DATA - VISUALIZATION (SINGLE)
56 % Visualization of a single channel signal at the two distinct wavelengths.
57 % This is needed to compare the behaviour of the tissue at correspondence
58 % to the two wavelength in time.
59
60 chSel = 65; % Channel to visualize
61
62 figure, plot(T,I(:,[chSel, chSel + 84])), %
63 title(sprintf('Channel %i',chSel))
64 xlabel('Time [s]'), ylabel('Intensity [A.U.]')
65 xlim([T(1) T(end)]), grid on, grid minor, box on,
66 legend('Wavelength 1','Wavelength 2')
67
68 %% DATA - SELECTION
69 % Selection of those channel that presents a mean value inside the range
70 % between 0.001 and 2.5 and with a minimum signal to noise ratio equal to
71 % 2. These information are passed through a function where are computed all
72 % the values.
73
74 meanTH = [0.001 2.5]; SNRTH = 2; % Options
75
76 selCh = noiseFreeChannel(I, meanTH, SNRTH); % Signal validation
77 nCHS = sum(selCh)/2; % Number of selected
78 nCH = length(selCh)/2; % Number of selected
79 selChW1 = find(selCh,nCHS,'first'); % Index wavelength 1
80 selChW2 = find(selCh,nCHS,'last'); % Index wavelength 2
81 mlAct{1} = selCh; % Homer notation
82
83 %% DATA - VISUALIZATION (FULL)
84 % Visualization of the data filterd by channel inspecintg the properties
85 % previously highlighted.
86
87 figure, subplot(1,2,1),
88 plot(T,I(:,selChW1)),
89 title('Wavelength 1'),
90 xlabel('Time [s]'), ylabel('Intensity [A.U.]')
91 xlim([T(1) T(end)]), grid on, grid minor, box on
92
93 subplot(1,2,2),
94 plot(T,I(:,selChW2)),
95 title('Wavelength 2')
96 xlabel('Time [s]'), ylabel('Intensity [A.U.]')
97 xlim([T(1) T(end)]), grid on, grid minor, box on
98 sgttitle(sprintf('Intensity signals (TOT = %i)',sum(selCh)))
99
100 %% OPTICAL DENSITY - COMPUTATION
101 % Computation of the optical density by first taking the absolute value of
102 % the whole measurements and than by taking the logarithm of the ratio
103 % between the absolute value of the record over its mean value.
104
105 OD = -log(abs(I)./mean(I,1)); % Optical density computation
106
107 NIRSOD = DataClass(); % Object generation
108 NIRSOD.SetTime(T); % Object generation: Time vector
```

```
109 NIRSDOD.SetDataTimeSeries(OD);           % Object generation: Time series matrix
110 NIRSDOD.SetM1(NIRSDATA.GetM1());         % Object generation: M1
111 NIRSDOD.SetDataTypeDOD();                % Object generation: Optical Density
112
113 %% DATA - VISUALIZATION (FULL)
114 % Full data visualization of the optical density calculated at the step
115 % before by open up a dedicated graphic window.
116
117 figure, subplot(1,2,1),
118     plot(T,OD(:,selChW1)),
119     title('Wavelength 1'),
120     xlabel('Time [s]'), ylabel('Intensity [A.U.]')
121     xlim([T(1) T(end)]), grid on, grid minor, box on
122
123     subplot(1,2,2),
124     plot(T,OD(:,selChW2)),
125     title('Wavelength 2')
126     xlabel('Time [s]'), ylabel('Intensity [A.U.]')
127     xlim([T(1) T(end)]), grid on, grid minor, box on
128     sgttitle(sprintf('Optical Density Signals (TOT = %i)',sum(selCh)))
129
130 %% DATA - SELECTION
131 % In this case we need to select those time points that are affected from
132 % motion artifacts and those time points which are not. In this situation
133 % we first evaluate the possible time points that are affected from motion
134 % noise, we then register them in a whole matrix and inside a summary
135 % vector. The paramters that are used to compute correctly the motion
136 % artifacts extraction and detection and the tMotion, the tMask, the
137 % SDThreshold and the AMPThresh. By changing them we could see the changing
138 % in the output of the algorithm.
139
140 tMotion = 0.5; tMask = 1; SDThreshold = 12; AMPThresh = 0.5;    % Options
141
142 [tInc, tIncCh] = hmrR_MotionArtifactByChannel(NIRSDOD,...          PROBE,[],mlAct,[],...
143                                                 tMotion, tMask,...          SDThreshold, AMPThresh);
144
145
146
147 %% DATA - VISUALIZATION (FULL)
148 % Visualization of the full data that we have computed the optical density
149 % and that we have selected inside the analysis. Additional information are
150 % proposed because over these signals are represented some additional
151 % information to see those interval of time neglected.
152
153 MOTION = (tInc{1} == 0)*min(OD(:)) + (tInc{1} == 1)*max(OD(:));
154
155 figure, subplot(1,2,1),
156     hold on,
157     plot(T, OD(:,selChW1)),
158     h = plot(T,MOTION,'r','LineWidth',2);
159     hold off,
160     h.Color = [1 0 0 0.05];
161     title('WAVELENGHT 1')
162     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
```

```
163         xlim([T(1) T(end)]), grid on, grid minor, box on
164         ylim([min(OD(:)) max(OD(:))])
165
166         subplot(1,2,2),
167         hold on,
168         plot(T, OD(:,selChW2)),
169         h = plot(T,MOTION,'r','LineWidth',2);
170         h.Color = [1 0 0 0.05];
171         title('WAVELENGHT 2')
172         xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
173         xlim([T(1) T(end)]), grid on, grid minor, box on
174         ylim([min(OD(:)) max(OD(:))])
175         sgttitle(sprintf('Motion Artifcat Corrected (TOT = %i)',sum(selCh)))
176
177 %% DATA - VISUALIZATION (SINGLE)
178 % Visualization of the single data that we have computed before in term of
179 % optical density. Visualization of the different instant of time where it
180 % is activated a motion artifact correction.
181
182 chSel = 65;
183
184 m      = min(OD(:,chSel));
185 M      = max(OD(:,chSel));
186 MOTION = (tIncCh{1}(:,[chSel, chSel + nCH]) == 0)*m + ...
187           (tIncCh{1}(:,[chSel, chSel + nCH]) == 1)*M;
188
189 figure, hold on,
190         plot(T,OD(:,[chSel, chSel + nCH])), 
191         h = plot(T,MOTION,'r','LineWidth',2);
192         h(1).Color = [1 0 0 0.05];
193         h(2).Color = [1 0 0 0.05];
194         hold off
195         title(sprintf('Channel %i',chSel))
196         xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
197         xlim([T(1) T(end)]), ylim([m M]),
198         grid on, grid minor, box on,
199         legend('Wavelength 1','Wavelength 2')
200
201 %% DATA - SPLINE INTERPOLATION
202 % Motion correction through the spline interpolation technique.
203
204 NIRSODMCS = NIRSOD.copy(); p = 0.99;
205
206 NIRSODMCS = hmR_MotionCorrectSpline(NIRSODMCS, mlAct, tIncCh, p, 1);
207 ODMCS     = NIRSODMCS.GetDataTimeSeries();
208
209 %% DATA - VISUALIZATION (SINGLE)
210 % Visualization of the effect of the spline interpolation on the data that
211 % are available. These data are contained inside the NIRSODMC structure.
212 % This is a snirf object that contains the usual information. To compare
213 % the results of the spline interpolation we put together two distinct
214 % figures.
215
216 m      = min([OD(:,chSel); ODMCS(:,chSel)]);
```

```

217 M      = max([OD(:,chSel); ODMCS(:,chSel)]);
218 MOTION = (tIncCh{1}(:,[chSel, chSel + nCH]) == 0)*m + ...
219           (tIncCh{1}(:,[chSel, chSel + nCH]) == 1)*M;
220
221 figure, subplot(1,2,1)
222     hold on,
223     plot(T,OD(:,chSel)),
224     plot(T,ODMCS(:,chSel)),
225     h = plot(T,MOTION(:,1),'r','LineWidth',2);
226     h.Color = [1 0 0 0.05];
227     hold off
228     title('WAVELENGHT 1')
229     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
230     xlim([T(1) T(end)]), ylim([m M]),
231     grid on, grid minor, box on,
232     legend('Non Corrected','Corrected','Location','SouthWest')
233
234     subplot(1,2,2)
235     hold on,
236     plot(T,OD(:,chSel + nCH)),
237     plot(T,ODMCS(:,chSel + nCH)),
238     h = plot(T,MOTION(:,2),'r','LineWidth',2);
239     h.Color = [1 0 0 0.05];
240     hold off
241     title('WAVELENGHT 2')
242     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
243     xlim([T(1) T(end)]), ylim([m M]),
244     grid on, grid minor, box on,
245     legend('Non Corrected','Corrected','Location','SouthWest')
246     sgttitle(sprintf('Channel %i',chSel))
247
248 % SPLINE
249 % With the spline technique we could clearly see the presence of a given
250 % high alteration of the baseline. The alteration of the baseline is due to
251 % the non complete identification of the artifact inside the optical
252 % density signal. This produce, beaucse the spline technique compute the
253 % new baseline on the samples suddenly before the identification of
254 % the artifact, an alteration of the baseline it self. In particular with
255 % the application of the spline technique the signal is not more affected
256 % from motion artifacts but some of those are not completely identified.
257 % Thus due to this high shift in the baseline the technique is not very
258 % reliable but it would be if the identification of the different motion
259 % artifacts was pefectly correct.
260
261 %% DATA - PCA INTERPOLATION
262 % Motion correction through the PCA interpolation technique.
263
264 NIRSDMCP = NIRSDOD.copy();
265 tMotion = 1; tMask = 2;
266 SDThresh = 11; AMThresh = 0.3;
267
268 NIRSDMCP = hmrR_MotionCorrectPCArecuse(NIRSDMCP, PROBE, [], mlAct, ...
269             [], tMotion, tMask, SDThresh, AMThresh, 0.97, 5, 1);
270

```

```
271 ODMCP      = NIRSDMCP.GetDataTimeSeries();  
272  
273 %% DATA - VISUALIZATION (SINGLE)  
274 % Visualization of the effect of the PCA interpolation on the data that  
275 % are available. These data are contained inside the NIRSDMCP structure.  
276 % This is a snirf object that contains the usual information. To compare  
277 % the results of the spline interpolation we put together two distinct  
278 % figures.  
279  
280 m      = min([OD(:,chSel); ODMCP(:,chSel)]);  
281 M      = max([OD(:,chSel); ODMCP(:,chSel)]);  
282 MOTION = (tIncCh{1}(:,[chSel, chSel + nCH]) == 0)*m + ...  
283           (tIncCh{1}(:,[chSel, chSel + nCH]) == 1)*M;  
284  
285 figure, subplot(1,2,1)  
286     hold on,  
287     plot(T,OD(:,chSel)),  
288     plot(T,ODMCP(:,chSel)),  
289     h = plot(T,MOTION(:,1),'r','LineWidth',2);  
290     h.Color = [1 0 0 0.05];  
291     hold off  
292     title('WAVELENGHT 1')  
293     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')  
294     xlim([T(1) T(end)]), ylim([m M]),  
295     grid on, grid minor, box on,  
296     legend('Non Corrected','Corrected','Location','SouthWest')  
297  
298     subplot(1,2,2)  
299     hold on,  
300     plot(T,OD(:,chSel + nCH)),  
301     plot(T,ODMCP(:,chSel + nCH)),  
302     h = plot(T,MOTION(:,2),'r','LineWidth',2);  
303     h.Color = [1 0 0 0.05];  
304     hold off  
305     title('WAVELENGHT 2')  
306     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')  
307     xlim([T(1) T(end)]), ylim([m M]),  
308     grid on, grid minor, box on,  
309     legend('Non Corrected','Corrected','Location','SouthWest')  
310     sgttitle(sprintf('Channel %i',chSel))  
311  
312 % PCA  
313 % The PCA technique used to remove different artifacts from all the signals  
314 % identified works well and not. In this case there is a very strong  
315 % alteration of the baseline but are keeps as the previous case all the  
316 % physiological compontents inside the signal. Moreover, the artifacts are  
317 % not removed at all. This beacuse the PCA technique works on the whole  
318 % matrix that contains information about all the signals. Thus, the motion  
319 % correction works very well if the artifact is presented inside all the  
320 % signals instead the motion correction will be not performed very well.  
321 % This could be seen as instance inside thos channels that are not  
322 % currupted from motion artifacts.  
323  
324 %% DATA - WAVELET CORRECTION
```

```

325 % Motion correction through the wavelet motion correction technique
326
327 NIRSODMCW = NIRSOD.copy();
328 a      = 1;
329
330 NIRSODMCW = hmrR_MotionCorrectWavelet(NIRSODMCW, mlAct, a, 1);
331
332 ODMCW      = NIRSODMCW.GetDataTimeSeries();
333
334 %% DATA - VISUALIZATION (SINGLE)
335 % Visualization of the effect of the wavelet correction on the data that
336 % are available. These data are contained inside the NIRSODMCW structure.
337 % This is a snirf object that contains the usual information. To compare
338 % the results of the spline interpolation we put together two distinct
339 % figures.
340
341 m      = min([OD(:,chSel); ODMCW(:,chSel)]);
342 M      = max([OD(:,chSel); ODMCW(:,chSel)]);
343 MOTION = (tIncCh{1}(:,[chSel, chSel + nCH]) == 0)*m + ...
344           (tIncCh{1}(:,[chSel, chSel + nCH]) == 1)*M;
345
346 figure, subplot(1,2,1)
347     hold on,
348     plot(T,OD(:,chSel)),
349     plot(T,ODMCW(:,chSel)),
350     h = plot(T,MOTION(:,1),'r','LineWidth',2);
351     h.Color = [1 0 0 0.05];
352     hold off
353     title('WAVELENGHT 1')
354     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
355     xlim([T(1) T(end)]), ylim([m M]),
356     grid on, grid minor, box on,
357     legend('Non Corrected','Corrected','Location','SouthWest')
358
359     subplot(1,2,2)
360     hold on,
361     plot(T,OD(:,chSel + nCH)),
362     plot(T,ODMCW(:,chSel + nCH)),
363     h = plot(T,MOTION(:,2),'r','LineWidth',2);
364     h.Color = [1 0 0 0.05];
365     hold off
366     title('WAVELENGHT 2')
367     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
368     xlim([T(1) T(end)]), ylim([m M]),
369     grid on, grid minor, box on,
370     legend('Non Corrected','Corrected','Location','SouthWest')
371     sgttitle(sprintf('Channel %i',chSel))
372
373 % WAVELET
374 % Inside the wavelet motion correction technique we could use different
375 % types of parameter alpha in the function. Lower it is lower is the
376 % variability inside the final signal that it is maintained. In the
377 % analysis is keep just the 0.8 alpha paramter and relative results. A
378 % value of alpha higher will not be able to detect all the artifact as the

```

379 % 0.8 and 0.1 alpha paramters value could. Moreover we do not choose the
380 % value of alpha equal to 0.1 beacuse it will remove the major percentage
381 % of the variability of the signal. In particular it will remove the
382 % physiological compontent and also some HRF response. If we choose to
383 % filter the signal with a value of alpha equal to 1.5 the variance of the
384 % signal kept is much higher and thus also some motion artifacs could not
385 % be corrected. This is the reason why that the value of alpha equal to 0.8
386 % is the best configuration that we could to choose to manipulate and
387 % correct all the motion artifact.

388

389 %% MOTION CORRECTION - COMPARISON

390 % Comparison of the different motion correction techniques applied in the
391 % previous cases in particular by enhancing the components that made up the
392 % 65th signals measured at the two distinct wavelength.

393

394 m = min([OD(:,chSel);ODMCW(:,chSel);ODMCP(:,chSel);ODMCS(:,chSel)]);
395 M = max([OD(:,chSel);ODMCW(:,chSel);ODMCP(:,chSel);ODMCS(:,chSel)]);
396 MOTION = (tIncCh{1}(:,[chSel, chSel + nCH]) == 0)*m + ...
397 (tIncCh{1}(:,[chSel, chSel + nCH]) == 1)*M;

398

399 figure, subplot(1,2,1)
400 hold on,
401 plot(T, OD(:,chSel), 'k')
402 plot(T, ODMCS(:,chSel), 'b')
403 plot(T, ODMCP(:,chSel), 'm')
404 plot(T, ODMCW(:,chSel), 'r')
405 h = plot(T,MOTION(:,1),'r','LineWidth',2);
406 h.Color = [1 0 0 0.05];
407 hold off
408 title('WAVELENGHT 1'),
409 xlabel('Time [ms]'), ylabel('Amplitude [A.U.]')
410 grid on, grid minor, box on,
411 legend('Original','Spline','PCA','Wavelet','Location','SouthWest')
412 xlim([T(1) T(end)]), ylim([m M])

413

414 subplot(1,2,2)
415 hold on,
416 plot(T, OD(:,chSel + nCH), 'k')
417 plot(T, ODMCS(:,chSel + nCH), 'b')
418 plot(T, ODMCP(:,chSel + nCH), 'm')
419 plot(T, ODMCW(:,chSel + nCH), 'r')
420 h = plot(T,MOTION(:,1),'r','LineWidth',2);
421 h.Color = [1 0 0 0.05];
422 hold off
423 title('WAVELENGHT 2'),
424 xlabel('Time [ms]'), ylabel('Amplitude [A.U.]')
425 grid on, grid minor, box on,
426 legend('Original','Spline','PCA','Wavelet','Location','SouthWest')
427 xlim([T(1) T(end)]), ylim([m M])

428

429 % COMPARISON

430 % Between the techniques that we have seen so far the best one between the
431 % three proposed is the wavelet in this case. The spline shows a good
432 % behaviour too but due to the non correct identification of all the

```
433 % duration of all the artifacts the baseline of the signal sometimes could
434 % be affected of some changes. The PCA gives not suitable result for this
435 % dataset because in the dataset the motion artifact does not corrupt all
436 % the channels but just someone and in certain interval of time. Thus, the
437 % signals that are not affected from motion artifact could suffer from this
438 % condition since the PCA is applied on all the channels.
439
440 %% DATA - NOISE REMOVAL
441 % Remove of the noise with the band pass filter. The frequency that are
442 % chosen to be removed are the one higher than 0.01 and lower than 3. This
443 % will keep the information about the HRF activated but will cut off the
444 % baseline (lowest frequency component) and the highest frequency component
445 % that could be noises.
446
447 NIRSODNR = NIRSODMCW.copy();
448 lowerCutOff = 0.01; higherCutOff = 3;
449
450 NIRSODNR = hmrR_BandpassFilt(NIRSODNR,lowerCutOff,higherCutOff);
451 ODNR      = NIRSODNR.GetDataTimeSeries();
452
453 %% DATA - VISUALIZATION (SINGLE)
454 % Visualization of the effect of noise removal from the data that we have
455 % loaded.
456
457 m      = min([OD(:,chSel); ODNR(:,chSel)]);
458 M      = max([OD(:,chSel); ODNR(:,chSel)]);
459 MOTION = (tIncCh{1}(:,[chSel, chSel + nCH]) == 0)*m + ...
460           (tIncCh{1}(:,[chSel, chSel + nCH]) == 1)*M;
461
462 figure, subplot(1,2,1)
463     hold on,
464     plot(T,OD(:,chSel)),
465     plot(T,ODNR(:,chSel)),
466     h = plot(T,MOTION(:,1),'r','LineWidth',2);
467     h.Color = [1 0 0 0.05];
468     hold off
469     title('WAVELENGHT 1')
470     xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
471     xlim([T(1) T(end)]), ylim([m M]),
472     grid on, grid minor, box on,
473     legend('Non Corrected','Corrected','Location','SouthWest')
474
475 subplot(1,2,2)
476 hold on,
477 plot(T,OD(:,chSel + nCH)),
478 plot(T,ODNR(:,chSel + nCH)),
479 h = plot(T,MOTION(:,2),'r','LineWidth',2);
480 h.Color = [1 0 0 0.05];
481 hold off
482 title('WAVELENGHT 2')
483 xlabel('Time [s]'), ylabel('Amplitude [A.U.]')
484 xlim([T(1) T(end)]), ylim([m M]),
485 grid on, grid minor, box on,
486 legend('Non Corrected','Corrected','Location','SouthWest')
```

```

487         sgttitle(sprintf('Channel %i',chSel))
488
489 % FITLERING
490 % From the filtering process that we could appreciate in the figure
491 % reproduced before we could notice the removal of the slowest component of
492 % the optical density signal. The physiological information stored inside
493 % the data structure after the filtering phase is still presents and this
494 % will be removed inside the GLM method by using some additional
495 % physiological information and measurements. Moreover, with the GLM we
496 % will also perform some data decorrelation processing that let all the
497 % datapoints be uncorrelated with each other. This because the NIRS data
498 % due to the high sample frequency have a very high temporal correlation.
499 % This means that, given a general instant of time t, the sample before and
500 % the sample after are highly correlated with the in line sample.
501
502 % FNIRS TECHNOLOGY:
503 % In the fnirs technology the light is not continuous but it is pulsed. This
504 % let the hardware be able to record information between detector and
505 % sources in pair in different instant of time. The detection method of the
506 % different stimuli delivered from the sources could be measured with two
507 % main techniques. The first one is the so called time multiplexing. In
508 % this case the stimuli are delivered at different instant of time which
509 % are not overlapped between each other. This let the hardware be able to
510 % detect separately all the stimuli delivered. The second method is the so
511 % called frequency multiplexing. In this case the stimuli are characterized
512 % with a different frequency and the detector once calculate the fourier
513 % transform of the signal could reconstruct the main component related
514 % with a given detector.
515 % Another important things inside the fnirs technology is that the
516 % amplitude of the stimulus could not reach a certain maximum value because
517 % need to satisfy all the condition and prerequisites in terms of amplitude
518 % of the internation law in materia of electromagnetic waves.
519
520 %% CONCENTRATION CHANGES - COMPUTATION
521 % To solve this problem we use the function hmrR_OD2Conc. The only thing
522 % that we need to do is the calculation of the DPF with the given value
523 % for the different parameters.
524 % We need to implement the formula of DPF and apply it for a subject
525 % involved in the acquisition phase of 28 years old.
526 % After having compute this quantity we could use the previous mentioned
527 % function to compute again the concentration changes in time.
528
529 a = 223.3;      b = 0.05624;    g = 0.8493;
530 d = -5.723e-7;  e = 0.001245;   x = -0.9025;
531 l = [760, 850]; age = 28;
532
533 DPF = a + b * age ^ g + d * l.^3 + e * l.^2 + x * l;
534
535 NIRSCC = hmrR_OD2Conc(NIRSDNR, PROBE, DPF);
536 CC      = NIRSCC.GetDataTimeSeries()*1e9;
537
538 %% DATA - VISUALIZATION (SINGLE)
539 % Visualization of the concentration changes in time inside two dedicated
540 % command window.

```

```

541
542 figure, hold on,
543 plot(T, CC(:,(chSel-1)*3+1), 'r-')
544 plot(T, CC(:,(chSel-1)*3+2), 'b-')
545 hold off,
546 title('CONCENTRATION CHANGES - HbR, HbO'),
547 xlabel('Time [ms]'), ylabel('Amplitude [nM]')
548 grid on, grid minor, box on
549 xlim([T(1) T(end)]),
550 legend('HbR','HbO')
551
552 % CONCENTRATION CHANGES
553 % The two signals seems very well correlated between each other. The output
554 % of the function hmrR_OD2Conc gives a SNIRF object that contains at its
555 % own time the different time series. Once extracted if we look to the size
556 % of the matrix we may appreciate a higher dimension respect to the one
557 % that we have seen for the initial time series. This because is increased
558 % the number of columns. In fact, in the time series data structure finally
559 % get is characterized by a different information organization.
560 % Specifically, the number of columns reflect the number of channels times
561 % the number of wavelength investigated plus one. In this case the number
562 % of wavelength is exactly equal to 2 and by adding 1 we got 3 and thus,
563 % because the channels are 84, the final number of column is equal to 252.
564 % The time series are organized in triplets. These are grouped per channel,
565 % and the first represents the amount of haemoglobin (HbR) deoxygenated, the
566 % second the amount of oxygenated haemoglobin and (HbO) and the third
567 % column represents the sum between the two previous concentration (HbS).
568 % The measure in output is represented in molar [M] but, because of the
569 % low amplitude of the final signal, the measures are then represented in
570 % [nM] by multiplying all the signals for the factor 1e9.
571
572 %% HAEMODYNAMIC RESPONSE - COMPUTATION
573 % To correctly compute the different haemodynamic response (HRF) we need to
574 % use some implemented function in Homer3 package. The information needed
575 % are reported down below and used then in the function:
576 %
577 % - tRange : Identify the time interval that it will be used as
578 % reference to compute the different HRF. In this case
579 % this parameter needed to be set to [-2 18]
580 % - glmSolveMethod : This is used to define the method to solve the GLM.
581 % The solution in this case are two and are distinct.
582 % The first is 1 and select the ordinary least square
583 % method. Instead the second, which is developed
584 % specifically for fNIRS data, is the number 2 and is
585 % represented with the least square with correction
586 % of serially correlated error. This last one has
587 % been developed because the fNIRS data shows a high
588 % temporal correlation between its samples. Thus to
589 % remove this statistical properties we need to apply
590 % some decorrelation steps that are already
591 % implemented in the function.
592 % - idxBasis : With this parameter we are going to set up the type
593 % of model that we wish to use to model the HRF. If
594 % we select 2 we use the gamma functions (that we

```

```

595 %
596 %
597 % - paramBasis : have seen that does not shows a good results) and
598 % if we set 1 we use some gaussian functions.
599 %
600 % - rhoSD_ssThresh : With this paramter we are defining the paramters
601 % that we want to investigate inside the HRF
602 % reconstruction.
603 %
604 %
605 % - flagNuisance : By specifying this value we will set up the
606 % distance between the source and the detector for
607 % which a channel is considered a short separation
608 % channel. The value of this paramter for this case
609 % is set to 15 mm.
610 %
611 % - RMethod
612 %
613 %
614 %
615 %

616 % The function is computationally expensive and thus the output will takes
617 % some time before get out from there.
618 % The most important output is the HRFs detected in all the different
619 % channels correlated with all the possible task. This in the output is the
620 % first term of the output vector. The final structure is a little bit
621 % complex because have another column dimension. In this case the number of
622 % column is equal to the number of stimulus times the number of channels
623 % times the number of detected concentration changes per channel.
624
625 trange = [-2 18]; glmSolveMethod = 2; idxBasis = 1; paramsBasis = [2 2];
626 rhoSD_ssThresh = 15; flagNuisanceRMethod = 1; driftOrder = 0;
627
628 NIRSHRFs = hmrR_GLM(NIRSCC, STIM, PROBE, [], [], [], trange, ...
629                 glmSolveMethod, idxBasis, paramsBasis, rhoSD_ssThresh, ...
630                 flagNuisanceRMethod, driftOrder, []);
631
632 HRFs = NIRSHRFs.GetDataTimeSeries()*1e9;
633 THRF = NIRSHRFs.GetTime();
634
635 %% DATA - VISUALIZATION (MULTIPLE)
636 % Visualization of multiple channels response through the visualization of
637 % the HbO and HbR responses when they are specifically activated over
638 % specific brain position.
639
640 figure, chSel = [39, 57, 72, 76];
641
642 for i = 1:length(chSel)
643     h(i) = subplot(2,2,i);
644     hold on,
645     plot(THR, HRFs(:,(chSel(i) - 1)*3+1), 'r-'),
646     plot(THR, HRFs(:,(chSel(i) - 1)*3+2), 'b--'),
647     plot(THR, HRFs(:,(chSel(i) - 1)*3+1+nCH*3), 'm-'),
648     plot(THR, HRFs(:,(chSel(i) - 1)*3+2+nCH*3), 'c--'),

```

```
649 plot(THRF,HRFs(:,(chSel(i) - 1)*3+1+nCH*3*2),'g-') ,
650 plot(THRF,HRFs(:,(chSel(i) - 1)*3+2+nCH*3*2),'k--') ,
651 hold off,
652 title(sprintf('CH = %i',chSel(i)))
653 xlabel('Time [ms]'), ylabel('\Delta Hb [nM]')
654 grid on, grid minor, box on,
655 legend('HbO RH','HbR RH','HbO LH','HbR LH','HbO FT','HbR FT')
656 end
657
658
```

LESSON 22: DOT STANDARD PROCESSING PIPELINE

The GLM can solved in different ways and the standard approach consist in the least square fitting of the data. The NIRS data contains a lot of correlation due to the high sampling rate. This is not shown from the fMRI signals and this is why we do not need any correlation correction in this last type of data analysis. The physiological noise and the high sampling rate, in NIRS signals, causes serially correlated errors. These is what affect the *FPR* (*False Positive Rate*) of the model, thus not estimating some HRF even if there is some response. The GLM could be expressed in the next mathematical formulation by explicating the gaussian additive noise:

$$y_t = h_t + \epsilon_t \text{ with } \epsilon_t \sim N(0, \sigma^2)$$

The serial correlation needed to be reduced. In this way the *FPR* of the model will not be affected and could be improved. The serial correlation can be removed or decreased with pre whitening based on an autoregressive model *AR* of the error term. In this case the autoregressive model could be expressed as:

$$AR(1): \epsilon_t = \rho_1 \epsilon_{t-1} + \nu_t$$

Where ν_t not more contains any correlation between the data because it is already contained inside the first term of the AR model. Thus with the autoregressive model we are trying to estimate the parameters to let the gaussian noise be uncorrelated and become a white noise term which will be added to the initial measurement. We would like to apply the next discrete autoregressive filter on both side of our initial signal description used for the GLM model. Hence, we obtain:

$$\begin{aligned} f \otimes y_t &= (y_t - \rho_1 y_{t-1}) = [h_t + \epsilon_t - \rho_1(h_{t-1} + \epsilon_{t-1})] \\ &= [h_t - \rho_1 h_{t-1} + \epsilon_t - \rho_1 \epsilon_{t-1}] = (h_t - \rho_1 h_{t-1}) + \nu_t \end{aligned}$$

Where we could see that there is not more the term ϵ_t , thus, because of the term ν_t is just white noise, there is not more correlation inside the equation thanks to the elimination of ϵ_t . Thus if we apply the filter, we could then we can estimate the HRF without the physiological correlation inside the signal that will reduce the performances in the *FPR*. In practical problems if we would like to extract the physiological correlation inside the data, we could even use a higher order autoregressive model. Thus the relation reported before could be expressed as:

$$\begin{aligned} AR(p): \epsilon_t &= \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + \cdots + \rho_p \epsilon_{t-p} + \nu_t \\ f &= [1 \quad -\rho_1 \quad -\rho_2 \quad -\rho_3 \quad \dots \quad -\rho_p]^T \\ f \otimes y_t &= f \otimes h_t + \nu_t \end{aligned}$$

In practical cases, we have different measurement coming from different channels of NIRS. Thus the system where we need to estimate all the different parameters could be expressed with the next relation:

$$y = X \beta + \epsilon$$

This can be corrected for serially correlated errors with a pre whitening filter f . Thus because of the autoregressive model needed to be applied at every channel inside the equation reported previously, we could summarize the convolutional operation as a matrix one where the convolutional matrix is called with F and is calculated starting from f . Mathematically:

$$Fy = FX\beta + F\epsilon$$

The incognita of the previous relation are the parameters β to associate them to the different regressor inside the X matrix. The estimation of these parameters could be done only with the least square approach.

$$\beta = (X^T F^T F X)^{-1} X^T F^T F y$$

Where the covariance matrix of the estimates could be expressed with the next relation:

$$cov(\beta) = \sigma^2 (X^T F^T F X)^{-1}$$

It has been observed that signals measures present a very high autocorrelation function value. This is not a good thing because it could worsen the data fitting and the FPR. After the application of the whitening processing the FPR and the estimation improves.

NIRS: Standard processing pipeline – Statistical analysis

What we want to do after having performed all the different computations and having followed the standard processing pipeline is the computation of the overall statistics. In particular, the objective of the NIRS data acquisition is that we would like to understand if specific tasks are activated in determine situation and condition in which the subject is surrounded. The analyses that could be done are something like the ones that are performed in fMRI which are also called the *SPM style* known as the *Statistical Parametric Maps* to compare condition. The usual statistical analysis performed with the NIRS technology are the ones cited next. In this case we could compare: the *amplitude*, the *latency*, the *width* of the HRF between cases. This analysis could be done thanks to high shape reconstruction accuracy of the HRF in the NIRS technology. In fact, due to the low reconstruction accuracy in fMRI the same analysis could not be performed for these types of data. Other possible characteristics investigated are the *mean around the peak* (avoid the oscillations in the plateau high phase of the HRF), the *peak*, the *FWHM*, the *AUC* or other possible parameters.

DOT: Introduction

Until now we have dealt with the NIRS analysis and its standard pipeline to follow to reach certain result. In this case the signals coming from the different electrodes are considered separately and they are analysed independently. This may lead to some non-integrative view of what is happening inside certain situation proposed to the subject. For this reason it has been proposed and introduced a new type of analysis called *Diffuse Optical Tomography (DOT)* to see the concentration changes inside the cortical layer in HbR and HbO or another specific chromophore.

The interaction between near infrared light and biological tissues are dominated by scattering. Thus achieving high resolution imaging with optical methods is very challenging. We could not get a high

resolution as the fMRI that presents a spatial resolution more or less of 3 mm, but we could get maximum a 1 cm of spatial resolution. This is due to the high interaction between the near infrared light and the biological tissues. However, each NIRS signal will inherently contain significant spatial information (remember the linear relationship between the source detector distance and depth reached by photons). There is therefore the possibility to produce maps of cortical functions, by displaying on the cortical layer the measured NIRS brain activation.

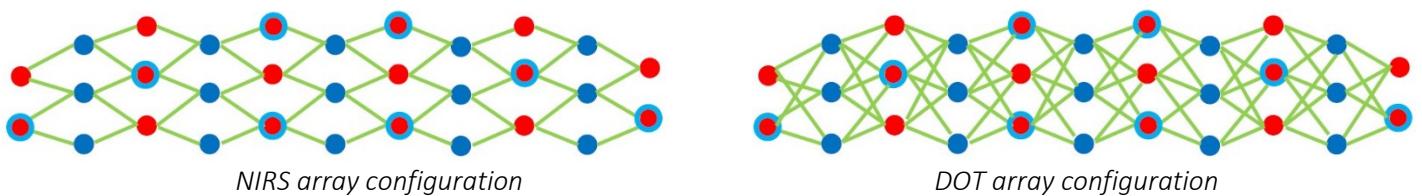
DOT: Differences with NIRS

The type of analysis that we will do is a channel wise analysis by estimating the HRF for each different stimulus proposed inside the designed experiment. With the *Diffusive Optical Tomography*, thanks to the optodes array configuration, we could achieve a good surface reconstruction activity in changes of the HbR and HbO. To reach the results of this type of analysis we need a lot of information and to do so we obviously need a lot of channels were performing measurements. To solve this problem we could use a multiple connection optodes array configuration that will be explained in detail later. The result will be some topographical maps where we could see the different activation of the distinct brain area.

The DOT requires the same data that are needed to correctly perform a NIRS analysis and thus we need to use the same hardware devices. The main difference between the two techniques is how we have to build up our array for acquisition and the results finally obtained.

DOT: Array

To perform the DOT the array needed to satisfy two characteristics. The first is that we need an array that have a multi distance channels which we could use to reconstruct the activation of signals with different depth, and the second is that we also need a lot of overlapping channels. This last characteristic is needed to measure in different channel the same activity and confirm the activation. This is as previously said a difference from the NIRS technique because in this last one the array of channels has always, or maybe in the most of time, the same distance between each source and each detector inside the array configuration. Hence, the brain depth investigated is always the same. In the DOT we are trying to estimate the behaviour inside voxel where we have not placed any optodes over. The array configuration needed and used also in our labs is the one reported before. The image on the left highlights the NIRS array configuration, while the image on the right highlights the DOT array configuration.

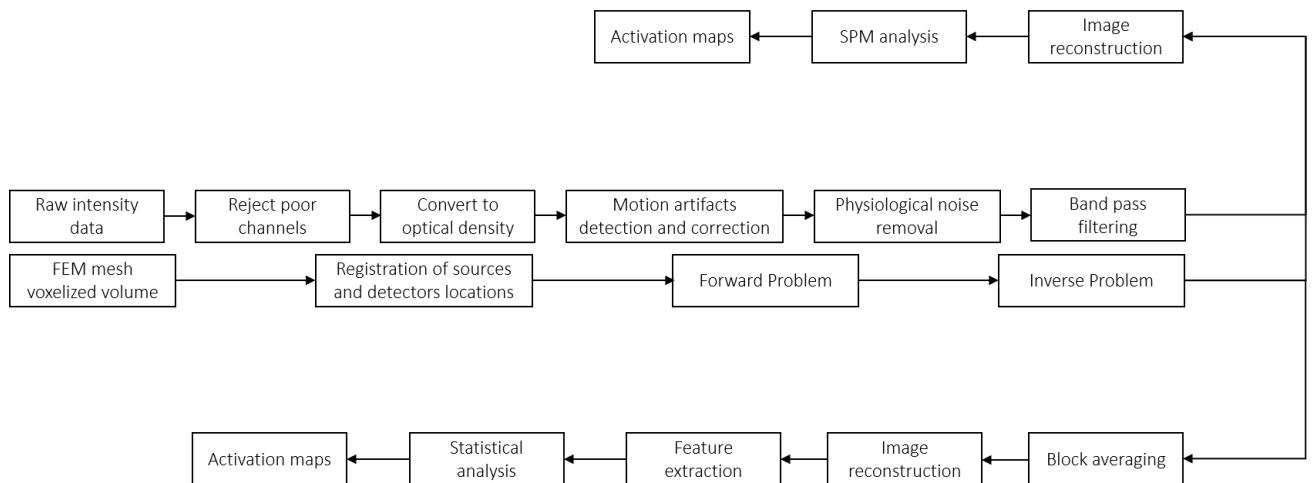


If we consider using the left array configuration for the DOT analysis this will not be worth it because are not satisfied the main characteristic needed for a DOT array. In the array configuration reported on the right instead we could appreciate the presence of multi distances and overlapping channel with which we could map and detect the activity of a certain brain area. To conclude the speech and discussion on the DOT and

NIRS the only things that let us select one given analysis instead of the other from the hardware point of view is array configuration.

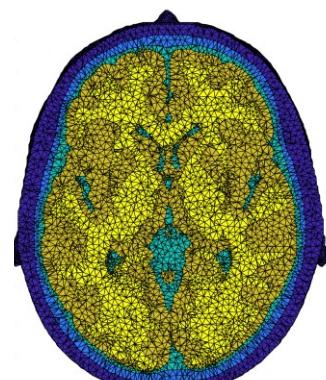
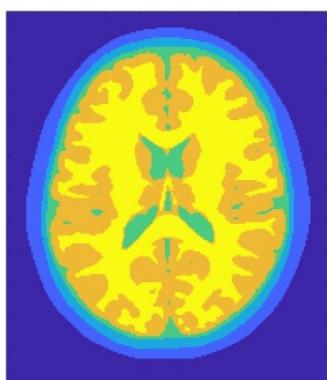
DOT: Standard processing pipeline

The standard processing pipeline used in DOT is very similar in terms of pre – processing steps to the NIRS one. To reach the final results of the pipeline the different processing steps needed are the forward and inverse problem solution that are part of the pre – processing phases. As with the NIRS after the pre – processing phase we have the statistical analysis section where we need to compute really the HRFs and find the different activation maps. In the following paragraphs and lectures we will not see the SPM analysis inside the DOT, but we will go deeper inside the black average method. Down below is reported the scheme used to describe the DOT processing steps.



DOT: Standard processing pipeline – Head model

To reach the end of the analysis, to perform the registration of sources and detector's location, the forwards and inverse problem solution we need a *head model*. This last one is a discretized model of the object under investigation and can be represented as a *voxelated volume (VV)* or a *finite element mesh (FEM)*. In the voxelated volume we have different voxel that have a shape of cubes or small parallelepipeds made all from of the same tissue. Instead, the FEM model is made up of different tetrahedron that are part of different tissues. Down below we could see two different head model based on VV (image on the left) and FEM (image on the right) approaches. With different colours indicates different tissues.



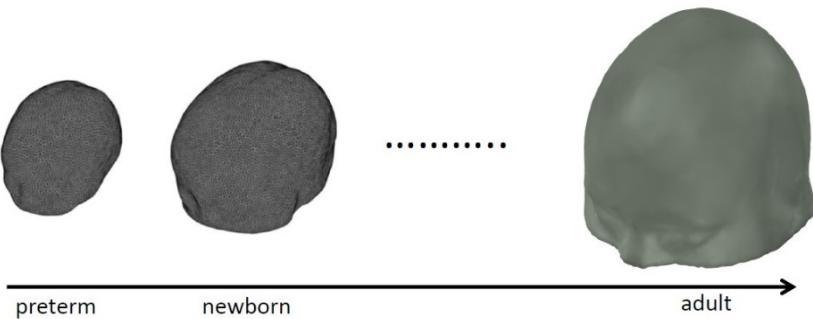
The voxelated model is made up of voxel and each voxel is assigned a number which is associated to a given tissue as instance the skull, the CSF, the grey matter and the white one. The FEM, instead, is made up of different tetrahedron. These three – dimensional elements are made up of four faces and four nodes too. These last ones are defined as the vertices of the triangles. Each of this tetrahedron is made up of different dimension and is associated to unique tissue type. The FEM model is much better and accurate instead the VV one. This because the tetrahedron gives more flexibility in build up the head model. These are the distinct ways that we could use to build up a head model of the subject.

DOT: Standard processing pipeline – Head model: Information

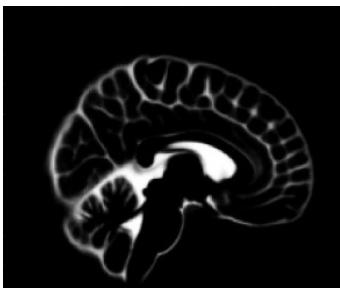
To build up a head model we need to base our analysis and build up process using some specific information. One of the best approaches to build up a head model suitable for the subject involved in the analysis is to perform some MRI images reconstruction to build up again the three – dimensional model of the subject's head. If this information or T1 or T2 weighted images are available, we could follow this way. Usually due to the high cost of the MRI scans and exams this method is not chosen inside clinical and research field. The other methods that could be implemented consists in using an ATLAS. This last one is a cheaper method and thanks to the high quantity of ATLAS available in literature for every subjects ages. This solution is less accurate than using the individual structural MRI, but the error introduced is lower than the spatial resolution of the DOT technique and for this reason it is a good alternative instead of using expensive exams as MRI.

DOT: Standard processing pipeline – Head model: ATLASES

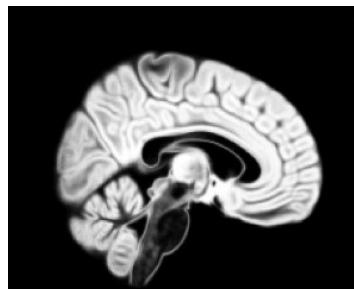
There exist several MRI atlases for different ages, since ideally, image reconstruction should be performed with an age matched head model (above all when neonates, infants, children, and elderly are acquired). This because the anatomical information of the head of the participants changes with time and thus, we need to access to this variability source.



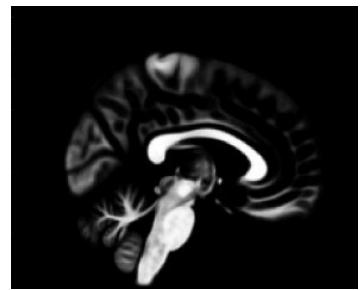
The MRI atlases are built up with the same procedure all the time. First, are acquired a lot of anatomical MRI images on a population of subjects or just a few and then we perform tissue segmentation to highlight just the important structure of the brain and to find the real section of the brain where we would like to perform registration. After this last step we could generate some three – dimensional head model and we could voxelated it or we may generate some surface mesh. After having created the three – dimensional model the ATLASES comes also with some mask to highlight where are located tissues like CSF, GM or WM. These are voxel wise masks that contains 0 or 1 if that voxel contains or not the specific tissue. In the figure reported in the next page we could see some examples of this.



CSF



Grey Matter



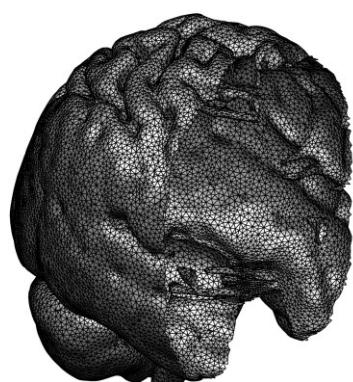
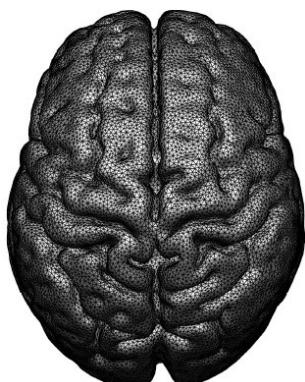
White Matter

The previous masks are made up starting from the tissue segmentation procedure. In this case we consider a given anatomical map and we perform some probability estimation that a certain voxel is associated to a given tissue instead of another. Then, by setting up different thresholds we could determine if one voxel belongs more to a given tissue instead of another.

The community have developed algorithm to determine and recognizing the inner section of the brain as the white and the grey matter (anatomical information about the organ). The main problem is found inside the extracerebral tissues. This because the MR developers thinks to enhance the organ's structure instead the external elements and moreover the ATLASES have been built by co – registering on the grey and white matter the different anatomical information available. Thus the extracerebral structure are more smoothed and should not be considered in our analysis.

Nowadays there exists several approaches to segment extra cerebral tissues. These comes from methods developed with custom code in MATLAB to method implemented in commonly used segmentation software. The manual segmentation procedure is an option as well (although quite time consuming). Some examples of segmentation software are as instance SPM which offers the possibility to segment the MR image into 6 tissue types respectively the *scalp*, the *skull*, the *CSF*, the *GM*, the *WM*, and finally the air. To improve the extra cerebral layer segmentation it is preferable using both T1w and the T2w images, since their different contrast carries useful information for the processes.

Until now the ATLAS build up is a voxelated three – dimensional structure. For each voxel that we have associated to a given tissue type, we could also compute and build up the volumetric mesh. This could be computed quite easily. In fact the tetrahedron is much more accurate on the definition of the brain tissues. In fact, in the FEM ATLASES not only the tetrahedron are associated to a given tissue types but also to all the given nodes and faces. Hence, form the segmentation mask we could also build the surface mesh of the scalp and the cortical layer thanks to the surfaces characteristic of the finite element used. The surface mask is not filled but shows only the surfaces of the different layers. Down below are reported some examples.



The surfaces elements that made up the surfaces mask reported before are triangles. All these last ones needed to be connected to obtain the final mesh. Higher is the dimension of the triangles lower is the accuracy of the reconstruction of the cortical layers.

DOT: Standard processing pipeline – Head model: State of art

The head models available are a lot and could be distinguished in a lot of different classes depending as instance on the age of the subjects involved in the reconstruction. A fact that underlines the variability of the head model that could be used are as instance the one that represents the evolution of the baby inside his or her mother. In last this case, the different tissue types are not well distinguishable because due to the developing processes still active. Specifically, the extracerebral structures could not be discriminate so much well because they are very thin (think also to the neonate's fontanelle).

A new creation has been introduced from the *Developing Human Connectome Project*. This last one is an organization or a scientific community that is trying to let available much more information and data as possible for the neuroscientist communities. The new creation generates and include an ATLAS where we could follow from the preterm to term population. In this last new creation the engineers and clinicians were able to developing and segmenting different anatomical element such as the hippocampus and the brainstem. The segmentation tools that were used in the infants are not the same as the one used with the adult's population. The last update of this ATLAS is date back to 2021 (the current year).

Another example of important ATLAS is the one that has been developed in the John Richard's lab. This covers a life span neurodevelopmental MRI databased from 2 weeks to 89 years old of a population of subjects. The ATLAS proposed a model for each groups of age where we could appreciate the difference between brain structure size in ages. This could be explained by considering the atrophy phenomenon in which the brain during the last years is involved.

In literature we could find also some head model for the life span between the 3 month and the 34 years. The information in this case is provided via FEM meshes that are much more accurate than the voxelated ones. As a complementary information for the life span cover with the previous head model, for the adult population, it is available the ATLAS called MNI152 which is an MRI ATLAS. This was built averaging 152 MRIs of the adult population and building up a FEM head model.

Other available head model is known as the COLIN27 and is made up on just one single subject called Colin. This ATLAS is highly detailed and convoluted because we have averaged 27 different brains images that are acquired over the same subject. Thanks to the high level of detail and consistency in the information presented in this ATLAS clinician were able to distinguish between distinct brain regions. Hence, of this head model are available also the anatomical brain parcellation. If we use this ATLAS instead of the other, we will introduce a lot of bias due to the high singular variability of the parcellation.

DOT: Standard processing pipeline – Registration of sources and detectors

Before moving on the forward and inverse problem we need to register the sources and the detectors by locating them on the head model chosen. Thus we need to register the channels, in particular sources and

detectors on the head model. This process of registration is made up mainly of two distinct approaches. The first one is to create a *template of the array* and register the array on the head based on anchor points. The second instead uses subject specific coordinates of the sources and the detectors and register those on the head model.

The array can be created on a 2D geometry and anchor points can be defined based on where the array should be placed on the head. Anchor points are usually defined based on the 10 – 20 EEG international reference system or the higher density version like the 10 – 10 and the 10 – 5 system. The main points are the nasion, the inion, those in the z lines and eventually the mastoids.

By exploiting these maps we could determine some anchor points where the different electrodes needed to be placed. In the first case we assume to register the array on the head model by not starting from the coordinates of the array on the head of the subject but starting from a template. This means that we have our digital array, and we would place it over the head model by registering it. The first thing to do, if we do not have the array model already, is to reconstruct the array model inside our computer by representing the real distances and position of each electrodes. In this case we could appreciate the actual distances inside the real case and the application of the array inside our head. The problem to solve is the method and the model that should be used to define the registration of the array over the head because some physical distances needed to be kept much more than other. For this reason we need to identify some anchor points and we need to add some springs to constraint the movement of the different electrodes one far from each other. In particular the anchor points usually are some additional sources or detectors that are virtually inserted in the array configuration to keep the array be stable over certain points when it is applied over the head model. Usually, the anchor points are defined as some nasion or inion points. The rigidity of the structure depends on how many springs we add to the array. In general the registration process and algorithm find the minimum energy solution to place the array over the scalp.

LESSON 23: DOT STANDARD PROCESSING PIPELINE

In the last lecture we have talked and discussed about the DOT. The three main steps that we need to perform are the *acquisition*, then we need to choose the right head model to apply to that specific subject which could be chosen from an ATLAS or by using subject specific MRI image. If we choose to use the ATLAS, we need to select the most correct in function of the age of the subject considered. Once chosen the head model, we need to find a way to register our optodes inside this last one. In the last lecture we have presented the approach that uses the template of the arrays and thanks to the usage of the springs method we could model the methodology on how the optodes array adapt to the scalp of the subject. The springs constraints the distances between the different sources and channels. We will use the same optode array configuration inside all the subjects, and we will register it inside a given ATLAS. The problem is that, by using the same ATLAS for all the subjects we lose the inter subject variability contribution.

DOT: Standard processing pipeline – Registration of sources and detectors

Another way to register the array on the head is by accessing to this last variability introduced as last concept of the previous paragraph neglected in the array template method. In this case we need to measure the real distances between the optodes by identifying its physical world coordinates from real world model and brings this information inside a virtual environment. To do this are recognized in literature two main methods. The first one is called the *digitizer approach*, while the second is known as *photogrammetry approach*.

DOT: Standard processing pipeline – Registration of sources and detectors: Digitizer

The *digitizer approach* uses a stylus that needed to be placed orthogonal to the surface where the optodes is placed. The stylus is characterized from a button that if it is clicked it acquires the three – dimensional position inside a real word reference set up at the beginning of the analysis. In this way we could get different information about the real position of the sources and the detectors thanks to some accelerometers and gyroscopes that are implemented inside the stylus. The points that needed to be acquired are firstly the so – called cranial landmarks. These are the *Nz*, *Cz* positions, the *Left* and the *Right Pre – Auricular* points and finally *Iz*. Where *Iz* and *Nz* are respectively the inion and the nasion of the subject.

By mapping these points we could get all the information about the head shape. The idea is to use these five points to build up a reference system to map every optode position inside the ATLAS considered. In fact, with these five reference points we could estimate an affine transformation between the cranial landmarks of the subject and the ATLAS ones. This transformation is very simple and could be represented in a matrixial form as reported below:

$$y = Tx \Leftrightarrow y = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} x$$

Where the first three columns are the terms responsible for the rotation, scales and shears, instead the fourth column terms refer to the different translations. Moreover As we could see from the matrix, all the coordinates needed to be expressed in homogeneous coordinate reference system. During the

measurement process of the position of these five points we could record the other position of the different sources and detectors that thanks to the estimation of the previous affine transformation could be directly mapped inside the head model considered.

The transformation T can then be applied to all other optode positions to register the optodes to the ATLAS or to all ATLAS mesh nodes to register the atlas to the subject's head. The head model is always the ATLAS head model, but it has been registered to the subject's head based on his or her cranial landmarks. This is a model specific ATLAS that could be used to get information on the HRF activation localization.

DOT: Standard processing pipeline – Registration of sources and detectors: Photogrammetry

The other method is known as the *photogrammetry approach*. The idea is to record a video of the subject head, as instance by rotating the smartphone around the him or her and we need to extract from this video a lot of pictures that represents the situation from different perspectives. By using this information we could create a virtual model for the head of the subject. Thanks to the rapid method of taking nowadays a video of the subject head this is largely used to reconstruct child head and optodes location.

The idea inside the photogrammetry approach is that we could extract the information about the optodes by identifying their colours placed on the top of the head (center of mass). What could be determine with precision from the video is a cloud of points that will be later on used. By using this method a more accurate registration can therefore be computed between the subject's head and the ATLAS, exploiting the entire external mesh of the subject's head. This mesh can be registered to the ALTAS one using a point cloud registration approach (machine learning technique).

To perform the registration of the cloud of point it is used the *Coherent Point Drift* algorithm (CPD). It can be used to register the ATLAS mesh to the subject's one. This method, which applied an affine transformation, aligns two surfaces solving a probability density estimation problem where the nodes of the two surfaces represent, respectively, the Gaussian mixture model centroids and the data to be fitted. What we could get is a head model that is always an ATLAS head model. But it has been registered to the subject's head based on his or her entire head features. Thus, it is the ATLAS that it is registered on the real data presented inside the subject head.

DOT: Standard processing pipeline – Forward model

Now that we have the head model and all the mapped sources and detectors inside it, we need to solve the forward and inverse problems. These two are also called the forward or inverse models.

The forward model refers to the use of the physical representation of our head represented with a model voxelated or with a FEM mesh as the basic space where to run and find a solution space for models of photon transport. These models predict how near infrared light will travel from each source location on the scalp through the tissue of the head, producing a *fluence distribution*. This photo transport model will estimate how the light will distribute among the head and the intensity of the electromagnetic radiation inside it.

The way in which light travels through a medium is governed by the *radiative transfer equation* known as *RTE*. This depends on a lot of properties as instance the absorption and scattering coefficients of the tissue,

the frequency of the electromagnetic light (thus near infrared and red light have different characteristic and fluence distribution) and then the refractive index of the tissue. The radiance is indicated mathematically as the term ϕ and represents the quantity of electromagnetic waves reflected from a unitary surface. The equation could be written as:

$$\frac{1}{C} \frac{\partial \phi}{\partial t} = -\mu_a \phi - \mu_s \phi + \mu_s \int_{S^2} \Theta(\hat{s}, \hat{s}') \phi(\hat{s}') d\hat{s}' - \hat{s} \nabla \phi + q$$

The unit of measurement of the radiance is the next one [$Wm^{-2}sr^{-1}$]. Where the sr^{-1} is the steradian unit of measure of the solid angle. From the previous equation we could see that the radiance of an electromagnetic wave loss energy due to absorption, scattering and divergence. These are respectively the first, the second and the fourth term in the equation. Then we could see that always related with the scattering the radiance could gain energy. This term is account inside the integral while that last term is a gain contributor because represents the source of the electromagnetic wave. This equation is impractical and too complicated to be solved in large volumes.

DOT: Standard processing pipeline – Forward model: FEM

The first approach is known as the *FEM approach*. In this case we use an approximation of the previous equation called the *diffusion approximation* indicated with the acronym *DA*. This is a good approximation of the *RTE* for almost isotropic fluences, but it is not good enough near boundaries and surfaces.

Using the *Finite Element Method* known as *FEM* the *DA* can be discretized and solved within complex geometries within a volumetric FEM mesh. The *diffusion approximation* is much easier than the *RTE* equation but still very complex. Its analytical representation could be summarized as:

$$\frac{1}{C} \frac{\partial \Phi}{\partial t} = -\mu_a \Phi + \nabla \cdot (\kappa \nabla \Phi) + q_0$$

Where here the radiance is assumed to be dependent on just the absorption of the electromagnetic waves from the tissue which interact with and depends on the divergence calculated over the product between the diffusion coefficient κ and the gradient of the radiance. The diffusion coefficient depends on the absorption and scattering coefficient indicated respectively with μ_a and μ'_s . This last one is called the reduced scattering coefficient and could be expressed as function of the anisotropy parameter g of the tissue.

$$\kappa = \frac{1}{3(\mu_a + \mu'_s)} \quad \mu'_s = (1 - g)\mu_s$$

Finally the term q_0 account for the energy emitted by the electromagnetic wave. The difference between the *RTE* and the *DA* equation is the second is a contracted from of the first. In this case all the losses of the previous equation *RTE* are summarized inside the second term of the *DA* equation. This mathematical relation could be solved within our head model very easily. For each optode on the scalp we could estimate the fluence by solving the equation inside the head model.

DOT: Standard processing pipeline – Forward model: Monte Carlo

The second approach uses the *Monte Carlo modelling approach*. These models are numerical, probabilistic way to find light distribution that satisfy the *Radiative Transfer Equation*. The light path that we could get from these simulations need to satisfy the previous equation. In this case, instead of solving the equation, we simulate the possible light travel according to the RTE. This way uses to model the path of the single photons by only randomly sample the next step length. Therefore we need to simulate the path and randomly sampling the step length, as previously said, the scattering angles and the depositing energy as a function of the local absorption coefficient. By doing this procedure a billion of times it is possible to reach a stable solution with building up a photon density distribution (fluence).

With the *Monte Carlo approaches (MC)*, we could get the real distribution of the photon's distribution approaches. This is not possible with the previous approach because with the FEM we could not solve the RTE but only minimizing the error from the real value. This method could be used both in voxelated volume and FEM meshes.

Both the FEM and the Monte Carlo simulation approaches requires an important information about the head model composition. The RTE and the DA depend on μ_s and μ_a as instance. Specifically they need to know the optical properties of the tissues composing the head model. Each node or voxel will have specific optical properties assigned depending on the tissue type it belongs to. Another important characteristic and parameter is the refractive index. This last one represents how much the light will travel fast inside our head. This has unit adimensional unit of measurement.

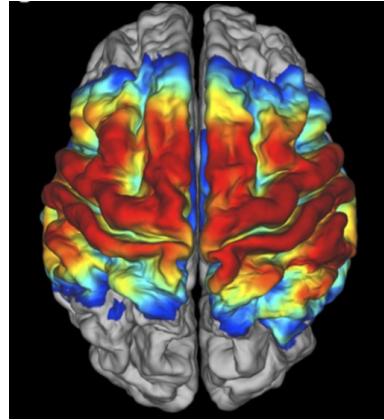
	absorption coefficient μ_a (mm^{-1})	reduced scattering coefficient μ'_s (mm^{-1})	refractive index
Scalp	0.018	0.69	1.3
Skull	0.017	0.92	1.3
CSF	0.003	0.12	1.3
GM	0.018	0.75	1.3
WM	0.019	1.10	1.3

With both the two methods we could reach the definition of a fluence distribution and estimation for each optode inside our array. Using the adjoint method and by exploiting the available simulations for each channel we could determine the sensitivity distribution. Thus we could characterize each PMDF of each channel and then we could compute the overall PMDF by only applying a principle of overlapping effects to describing the final sensitivity map also known as overall PMDF.

This allows us to determine how our measurements of intensity would be altered by a given change in the optical properties within our solution space. This sensitivity distribution on the margin does not have a huge impact on what we could measure inside the scalp. The most important thing is that the measurement needed to be performed with the highest deliver of photons inside the scalp to detect the best variation of the optical properties. This is known as solving the forward problem and yields the Jacobian matrix J .

This last structure present in every row the sensitivity (PMDF) of a given channel. While the columns are referring to different nodes or voxels. Thus, in this way we have a value of the PMDF function for each channel distributed all over our solution space. The sensitivity map of the array can be computed by

summing the PMDFs of all the channels. An instance of sensitivity map of our array is the next cortical representation of the head model.



DOT: Standard processing pipeline – Inverse problem: First approach

The last step before the different statistical analysis and result interpretation coming from the DOT approach is known as the inverse problem solution. This challenge stand on reconstructs the activity of HbR and HbO over the cortex of the brain where the array is positioned. The problem of image reconstruction is the one formalized below:

$$y = A(x)$$

Where the value of x represents the optical properties of the object under investigation which are some parameters of the images and are function of the images position considered. Then we have our measured data or channels indicated with the symbol y and finally there is a so – called *forwards operator* indicated with A that solve the forward problem, thus from the activity of the source reconstruct the different signals measured. The operation involved inside the forward problem are of non – linear type. The aim is to determine the image x starting from our channel data measured on the scalp. This is somethings that we have seen with the NIRS. We have our intensity data, and we would like to estimate the value of the absorption and the scattering coefficient inside the solution space.

This is an ill posed and under determined non – linear inverse problem. Thus this means that we have an infinite solution space, and we use the term *ill posed* because the solution is not unique, and it is not stable. To find the best solution for the problem, we need to determine the error term that could be defined as:

$$\text{Error} = \sum_{i=1}^{N_{\text{meas}}} (y_{\text{meas}} - y_{\text{model}})^2 \quad \text{Error} = \sum_{i=1}^{N_{\text{meas}}} (y_{\text{meas}} - A'(x_{\text{model}}))^2$$

The previous function is, in our case, the cost functions that could be used. Specifically, the solution of the inverse problem could be done with an iterative approach by first using some initial guess for the scattering and the absorption coefficients and then calculate, before testing the cycle condition, the model data using as instance the diffusion approximation equation reported before as solution of the forward problem. Now, we need to estimate the error and test if this last one is enough small to skip out the cycle or remain inside it. Thus, the error is compared with a threshold. If the error however is still higher than the threshold the

algorithm keeps working and update the absorption and scattering coefficient thanks to the Jacobian calculated matrix. Because of this algorithm is such expensive in computational term we need to solve the problem in another way. The idea is to use the most common approach already seen to perform image reconstruction inside the NIRS technique by recovering the HbR and HbO by solving the inverse problem stated inside the modified Lamber Beer Law. The only assumption that is done is that the changes in haemodynamic due to functional stimulation are small enough that could be neglected and do not corrupt our results.

Let assume there is a change in optical properties from x_0 to x_1 . The forwards operator $A(x_1)$ can be expanded over x_0 using a Taylor series as reported below:

$$A(x_1) = A(x_0) + \frac{\partial A(x_0)}{\partial x} (x_1 - x_0) + \frac{1}{2} \frac{\partial^2 A(x_0)}{\partial x^2} (x_1 - x_0)^2 + \dots$$

We could see the presence of different order of the derivative. Because of the expansion could reach infinite order but the highest order terms do not carry a lot of information we could neglect them and consider only the term until the first order. By using this last approximation we could write the equation reported before as a linear one.

$$A(x_1) \approx A(x_0) + \frac{\partial A(x_0)}{\partial x} (x_1 - x_0) \Rightarrow A(x_1) - A(x_0) \approx \frac{\partial A(x_0)}{\partial x} (x_1 - x_0)$$

Where the first difference represents the measurements changes that are detected from the optodes (same measure as the NIRS technique). Then the derivative contribution could be replaced with the Jacobian matrix, thanks to its definition, and finally the last finite difference shows the changes in the optical properties of the discretized space.

$$\Delta y = J \Delta x$$

To solve the previous problem we need to take the inverse of the matrix J . For each wavelength we have a given equation that needed to be solved. The Jacobian tells us how much the measurements y will change for a given change in the object's background optical properties x . Thus the value of J depends only on x_0 and thus finally the problem is a linear one.

The value of Δy represents a vector of measurements, of dimension $1 \times N_{meas}$, where the value of N_{meas} represents the number of channels. The value of Δx represents the vector of optical properties and is characterized of a dimension $N_{node} \times 1$ that corresponds to the number of nodes inside the FEM mesh. Finally, the Jacobian (J) has a dimension $N_{meas} \times N_{node}$. By inverting the initial relation between these quantities we obtain:

$$\Delta x = J^{-1} \Delta y$$

The problem is that J is not a square matrix and thus there is an underdetermined problem because we have many more unknowns than measurements. There is any exact solution, and the problem is therefore ill posed. This means that it is badly conditioned and thus the noise is amplified.

The inversion therefore requires the regularization. To perform this operation it is required the assumption that neighbour nodes are similar, the standard approach is to use the *Tikhonov regularization*:

$$J^{-1} \approx J^T(JJ^T + \lambda I)^{-1}$$

Where I is the identity matrix and the value of λ is our tuneable regularization parameter. This is the easiest way to calculate the inverse of a non – square matrix. The previous regularization (Tikhonov regularization) could also account for the variability presents inside the data and the model. Hence the previous relation could be written as:

$$J^{-1} \approx \Sigma_u J^T(J\Sigma_u J^T + \Sigma_v \lambda)^{-1}$$

Where the value of Σ_u is the covariance matrix of the model and Σ_v is the covariance matrix of the measurements. The value of λ could be expanded with the next relation, where the value of λ_1 is very low.

$$\lambda = \lambda_1 \frac{\text{trace}(J\Sigma_u J^T)}{\text{trace}(\Sigma_v)}$$

The estimated value of Δx that representes the changes in the absorption coefficient. The modified Beer Lamber law can be applied node wise to estimate the concentration changes of HbO and HbR in the image space from the node wise changes in absorption coefficient recovered from both wavelengths data. The image reconstruction approach can be applied to all time points of the time series of the haemodynamic response function that we want to reconstruct. The different images can be displayed one after the other as in a movie.

DOT: Standard processing pipeline – Inverse problem: Second approach

There is another method that could be used to solve the problem and reconstruct the concentration changes. We could estimate these last quantities by using one single step approach represented and named *multippectral image reconstruction*. In this case we directly apply the Beer Lamber law inside the forward problem equation.

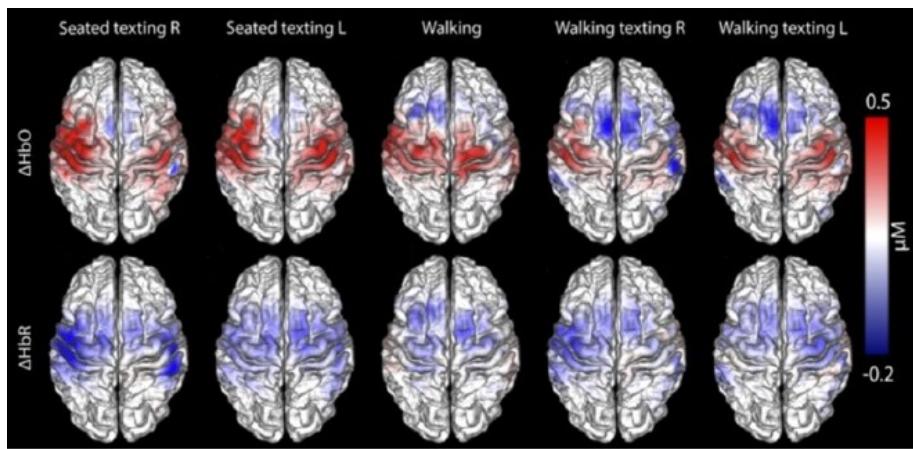
$$\Delta y = J \epsilon \Delta c$$

Where the value of ϵ is also known as the specific absorption coefficient. Thus, the relation is the next one:

$$\begin{bmatrix} \Delta OD_{\lambda_1}^1 \\ \Delta OD_{\lambda_1}^2 \\ \vdots \\ \Delta OD_{\lambda_1}^n \\ \Delta OD_{\lambda_2}^1 \\ \Delta OD_{\lambda_2}^2 \\ \vdots \\ \Delta OD_{\lambda_2}^n \end{bmatrix} = \begin{bmatrix} J_{\lambda_1} \epsilon_{\lambda_1}^{HbO} & J_{\lambda_1} \epsilon_{\lambda_1}^{HbR} \\ J_{\lambda_2} \epsilon_{\lambda_2}^{HbO} & J_{\lambda_2} \epsilon_{\lambda_2}^{HbR} \end{bmatrix} \begin{bmatrix} \Delta C_{HbO} \\ \Delta C_{HbR} \end{bmatrix}$$

For each channel is estimated the optical density data and while before we have just the first part of the equation now, we have the same column vector made up of two wavelengths. Then we have the whole Jacobian matrix that is made up of the smaller Jacobian calculated in both wavelength that are multiplied for the different absorption rate that depends on the wavelength and the chromophores considered. Thus in the final matrix we have doubled the dimension of the Jacobian matrix. In the first column we reported the specific absorption coefficient for the HbO while for the second we have the one related to HbR. The optical density changes are represented as changes between the active and the rest state from the baseline.

We have a maximum depth of resolution for the HRF estimated. In the image we could see the different activation of the brain region depending on the different task that it is activated during a specific situation. Instead of just visualizing the three – dimensional representation we project the activate that we see inside the cortical surface and what we could see is a cortical surface activation map. Some possible results that come from the analysis could be the one reported below.



The results are projected on the cortical layer to improve the visualization of the activation maps that are obviously obtained. We could see that by moving the hand and the walking condition reflect the ERD and ERS activation maps. To perform even better reconstruction of the experimental settings and results we could create some movies where we could represent how the concentration of the two main chromophore evolves in time.

DOT: Standard processing pipeline – Statistical analysis

The last reconstruction step is to perform some statistical analysis to determine which region is active once obtained measured from given channels. In this case there is not standard pipeline. A thing that we need to account is that we are performing a lot of multiple statistical tests, hence, we need to implement some multiple comparison correction too. One possible solution is by using permutation.

```
1 %
2 % Title: Laboratory 10
3 %
4 % Note:
5 % DOT analysis of some optical intensity data
6 %
7 % Author: Matteo Martin
8 %
9 % THEORY:
10 %
11 % In this laboratory we will use the AtlasViewer Toolbox. This is a
12 % software developped mainly for Matlab that help to create array
13 % configuration and solve eventual forward and inverse problem of the
14 % analysis loaded. When it needed to be used we need to add first
15 % completely to the Matlab path the folder where it is contained and then
16 % we can call it directly from the comand window and it will pop up a GUI.
17 %
18 % In this analysis the idea is to use the Homer3 package to perform all the
19 % possible analysis on the NIRS data and then uses AtlasViewer to visualize
20 % the different results, in terms of array configuration and solution of
21 % the possible inverse or forward problem. Because of these last two
22 % operations needed to be formalized carefully, one of the main ingredients
23 % needed is the so called solution space or the head model.
24 % This last component could be load inside the opened GUI. This help us to
25 % define a given ATLAS and anatomical dimesion subject specific.
26 %
27 % To generate an head model for the DOT analysis, we need to produce:
28 % - Head Volume Mesh
29 % - Scalp Surface Mesh
30 % - Tissue Mask
31 % - Tissue Surface Mask
32 % - International standard system used
33 %
34 % ATLAS VIEWER TOOLBOX
35 % Inside the Atlas Viewer Toolbox are available some head model. The one
36 % that we will use is the Colin27 ATLAS. It comes with a parcellation. This
37 % information could be exploited by looking on the labels inside the brain.
38 % Other possible ATLASES that could be used depends on the subject's age
39 % and usually they comes with a given anatomical parcellations.
40 %
41 % ATLAS VIEWER TOOLBOX - ROTATE VIEW
42 % The graphic user interface is made up of different boxes. The main one is
43 % called "Rotate/Zoom Axes". Here we could rotate in wherever direction the
44 % visualized head model in output of the graphical window. To set up
45 % correctly the visualization we could use the azimuth and the elevation
46 % levels or by using the buttons that are already provided inside this box.
47 %
48 % ATLAS VIEWER TOOLBOX - REFERENCE SYSTEM
49 % Another box opened with the graphical user interface is the reference
50 % points. Here we could select and choose between to visualize the Labels,
51 % the Circles and finally the Head Dimension. The first visualize over the
52 % head model considered the different position of the standard system. The
53 % circles highlight the position of the standard system and finally, the
54 % head dimension instead visualize the anatomical dimension of the patient.
```

```
55 %
56 % ATLAS VIEWER TOOLBOX - HEAD MODEL VISUALIZATION
57 % There are some important information that could be visualized graphically
58 % inside the head model. This is made by selecting the Lables setting in
59 % the Brain box. To do not visualize the head and the scalp we can click
60 % over the Head button place down below the brain box.
61 %
62 % ATLAS VIWER TOOLBOX - ARRAY REGISTRATION
63 % The first operation to register the array inside the ATLAS considered is
64 % to import the array configuration contained inside an .SD file. This
65 % step could be easily done by clicking under Tools and then selecintg
66 % Probe and finally Import. This will pop up a window where we need to
67 % specify the array configuration that we would like to load.
68 % Then to register the probe configuration on the surface of the head model
69 % we need to click inside the Probe box the Register Probe to Surface
70 % button and after few seconds we will see graphically that the array will
71 % arrange over the head model.
72 %
73 % ATLAS VIWER TOOLBOX - ARRAY EDIT
74 % To edit some properties of the array loaded inside the workspace of
75 % Matlab we need to click under Tools, then selecting Probe and finally
76 % Edit. This will pop up a window where we could see different array
77 % information. These last one are, the optodes and their positions inside
78 % the optodes box. Generally, when the array configuration is open it is
79 % not applied the springs inside the visualization of the array model. Thus
80 % we could not find the table anchor points under the probe box. This last
81 % one defines the number of optodes that needed to be anchor to certain
82 % points inside the standard reference system. Just to remark an important
83 % theoretical result, when we add springs to our model we are putting some
84 % constraints such as the electrodes in the anchor position will move less
85 % as possible and so also the other probes.
86 %
87 % ATLAS VIWER TOOLBOX - FORWARD MODEL
88 % To obtain a certain solution for the forward model we need to set first
89 % the Monte Carlo paramters and then we need to click under the Forward
90 % Model button and select Generate MC input. The paramters are set instead
91 % inside the Set MC Paramters window where we need to specify some tissue
92 % paramters like the absorption and scattering coefficient and the number
93 % of photons to simulate. When this all is set up we could easily compute
94 % the Generate MC button. This procedure will calculate the Jacobian
95 % matrix to solve the forward and inverse problem. The Jacobian and the
96 % sensitivity matrix could be estimated by only looking on the projection
97 % of these inside the brain surface. The procedure of the forward model
98 % solution is quite computationally expensive.
99 %
100 % All this result coming from the last step of the analysis will be later
101 % used to compute the fowards problem and the fluence distribution.
102
103 %% CCC
104 % Clear all, close all, clear the comand window
105
106 close all; clear all;clc
107
108 %% DATA - LOAD (MAT)
```

```

109 % Load of the data that are contained inside the folder data.
110
111 load('HeadVolumeMesh')
112 load('ScalpSurfaceMesh')
113 load('GMSurfaceMesh')
114 load('TissueMask')
115
116 %% DATA - LOAD (FILE)
117 % Load of the data inside the workspace of Matlab from text files
118
119 fid = fopen('10-5_Points.txt','r');
120 tmp = textscan(fid, '%s %f %f %f', 'CollectOutput',1);
121     fclose(fid);
122
123 tenFive      = tmp{2};
124 tenFiveLabel = tmp{1};
125
126 fid = fopen('LandmarkPoints.txt','r');
127 tmp = textscan(fid, '%s %f %f %f','CollectOutput',1);
128     fclose(fid);
129
130 cranialL    = tmp{2};
131 cranialLabel = tmp{1};
132
133 clear fid tmp ans;
134
135 %% DATA - VISUALIZATION
136 % Visualizaion of the data inside some dedicated graphic window open.
137
138 nCOL = size(parula(),1);
139 nLVL = 5;
140 COL  = parula();
141 COL  = COL(1:round(nCOL/nLVL):nCOL,:);
142
143 figure, imagesc(TissueMask(:,:,:,50)),
144     title('TISSUE MASK'),
145     colormap(COL);
146     CBAR = colorbar('Ticks',[0:5]....
147     'TickLabels',{'EXT','SCALP','SKULL','CSF','GM','WH'});
148
149 % As we could see from the image reported in output we have manage the
150 % colorbar to identify the different tisuses inside the one analysed. This
151 % is important to visualize the brain composition inside our mask.
152
153 %% DATA - VISUALIZATION
154 % Visualization of the different mesh inside a graphical window.
155
156 % TOTAL VIEW -----
157 figure,
158 plotmesh(HeadVolumeMesh.node(:,1:3), HeadVolumeMesh.elem);
159 title('HEAD - TOTAL VIEW'), grid on, grid minor, box on
160
161 % SAGGITAL VIEW -----
162 figure,

```

```
163 plotmesh(HeadVolumeMesh.node(:,1:3), HeadVolumeMesh.elem, 'x<0');
164 title('HEAD - SAGGITAL VIEW'), grid on, grid minor, box on
165
166 % AXIAL VIEW -----
167 figure,
168 plotmesh(HeadVolumeMesh.node(:,1:3), HeadVolumeMesh.elem, 'z<0');
169 title('HEAD - AXIAL VIEW'), grid on, grid minor, box on
170
171 % CORONAL VIEW -----
172 figure,
173 plotmesh(HeadVolumeMesh.node(:,1:3), HeadVolumeMesh.elem, 'y<0');
174 title('HEAD - CORONAL VIEW'), grid on, grid minor, box on
175
176 % GRAY MATTER - VOLUME -----
177 figure,
178 plotmesh(HeadVolumeMesh.node(:,1:3), ...
179     HeadVolumeMesh.elem(HeadVolumeMesh.elem(:,5) == 4,1:3));
180 title('HEAD - GREY MATTER - TOTAL VIEW'), grid on, grid minor, box on
181
182 % GRAY MATTER - VOLUME -----
183 figure,
184 plotmesh(HeadVolumeMesh.node(:,1:3), ...
185     HeadVolumeMesh.elem(HeadVolumeMesh.elem(:,5) == 4,1:3),'x<0')
186 title('HEAD - GREY MATTER - SAGGITAL VIEW'), grid on, grid minor, box on
187
188 % GRAY MATTER - SURFACE -----
189 figure,
190 plotmesh(GMSurfaceMesh.node(:,1:3), GMSurfaceMesh.face, 'x<0')
191 title('HEAD - GREY MATTER SURFACE'), grid on, grid minor, box on
192
193 % SCALP - SURFACE -----
194 figure,
195 plotmesh(ScalpSurfaceMesh.node(:,1:3), ScalpSurfaceMesh.face)
196 title('HEAD - SCALP SURFACE'), grid on, grid minor, box on
197
198 % ARRAY CONFIGURATION -----
199 figure,
200 hold on,
201 plotmesh(ScalpSurfaceMesh.node(:,1:3), ScalpSurfaceMesh.face)
202
203 plot3(tenFive(:,1),tenFive(:,2),tenFive(:,3), ...
204     'sb','MarkerFaceColor','b')
205
206 plot3(tenFive(1,1),tenFive(1,2),tenFive(1,3), ...
207     'og','MarkerFaceColor','g','MarkerSize',10)
208
209 plot3(cranialL(:,1),cranialL(:,2),cranialL(:,3),...
210     'om','MarkerFaceColor','m','MarkerSize',10)
211 hold off
212 title('HEAD - ARRAY CONFIGURATION')
213
214 %% ARRAY REGISTRATION
215 % Registration of the array inside the head model loaded and available
216 % inside the AtlasViewer Toolbox and available also offline.
```

```
217
218 AtlasViewerGUI
219
220 %% DATA - LOAD
221 % Load of the data inside the Matlab's workspace of the results coming from
222 % the AtlasViewer Toolbox analysis.
223
224 load('fwmodel.mat')
225 load('registeredProbe.mat')
226 load('Adot_18.mat')
227 load('Adot.mat')
228
229 %% OPTODE ARRAY - VISUALIZATION
230 % Visualizaiton of the optode array inside a three dimensional space
231
232 figure,
233 hold on,
234 plot3(optpos_reg(1:16,1),optpos_reg(1:16,2),optpos_reg(1:16,3),...
235     'ro','MarkerFaceColor','r','MarkerSize',10)
236 text(optpos_reg(1:16,1),optpos_reg(1:16,2),optpos_reg(1:16,3), ...
237     num2str([1:16]),'Color','k')
238 plot3(optpos_reg(17:39,1),optpos_reg(17:39,2),optpos_reg(17:39,3),...
239     'bo','MarkerFaceColor','b','MarkerSize',10)
240 text(optpos_reg(17:39,1),optpos_reg(17:39,2),optpos_reg(17:39,3), ...
241     num2str([17:39]),'Color','k')
242 plot3(optpos_reg(40:43,1),optpos_reg(40:43,2),optpos_reg(40:43,3),...
243     'go','MarkerFaceColor','g','MarkerSize',10)
244 text(optpos_reg(40:43,1),optpos_reg(40:43,2),optpos_reg(40:43,3),...
245     {'Inion';'Nasion';'MastdL';'MostdR'},'Color','k')
246 hold off,
247
248 title('ARRAY CONFIGURATION - REGISTERED'), grid on, grid minor, box on
249
250 %% SEGMENTATION MASK - VISUALIZATION
251 % Visualizaiton of the tissue segmentation mask
252
253 nCOL = size(parula(),1);
254 nLVL = 5;
255 COL = parula();
256 COL = COL(1:round(nCOL/nLVL):nCOL,:);
257
258 figure,
259 imagesc(imrotate(squeeze(fwmodel.headvol.img(:,:,128,:)),90));
260 colormap(COL);
261 CBAR = colorbar('Ticks',[0:5],...
262     'TickLabels',{'EXT','SCALP','CSF','GM','WM'});
263 title('SEGMENTATION MASK')
264
265 %% FLUENCE DISTRIBUTION - VISUALIZATION
266 % Visualization of the fluence distribution
267
268 figure(),
269 imagesc(imrotate(squeeze(log(A(:,:,133))),-90)), colorbar,
270 caxis([-20, -5]), title('FLUECE DISTRIBUTION')
```

```
271
272 %% SENSTIVITY MAP - VISUALZATION
273 % Visualization of the sensitivity map
274
275 fwmodel.mesh.vertices(:,4) = log(sum(Adot(:,:,1)));
276 figure(), plotmesh(fwmodel.mesh.vertices, fwmodel.mesh.faces),
277         colorbar, caxis([-1, 0])
278
279 fwmodel.mesh.vertices(:,4) = log(Adot(24,:,:));
280 figure(), plotmesh(fwmodel.mesh.vertices, fwmodel.mesh.faces),
281         colorbar, caxis([-8, 0])
282
283
284
```

LESSON 24: EEG TOPOGRAPHICAL MAP

In these last lectures we will look something about the *EEG Sources Reconstruction*. We will see how we could exploit the data that comes from the EEG to obtain different images.

EEG: Introduction

EEG is the recording of the electrical activity of the brain. The EEG is highly employed both in clinical and research fields. Some of its possible application is to be used to ascertain the brain death. It is used too in psychological and cognitive science to directly measure the neuronal activity. Instead of the NIRS and the DOT this technique is a direct measure of the neuronal electrical activity. The EEG temporal resolution is much better than other possible measurement system. Both in NIRS and in EEG we could measure and detect the stimulus and the different phases during the elaboration of it.

EEG: Instrumentation

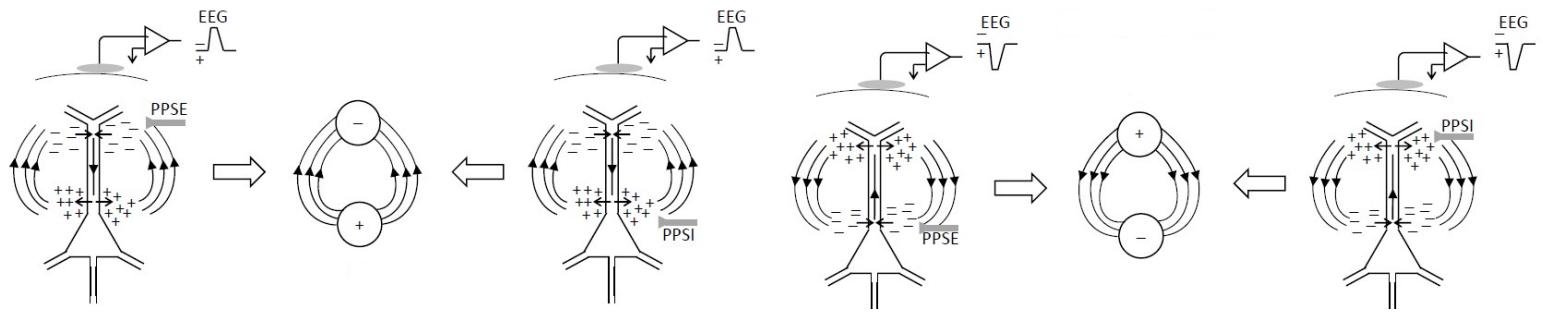
In terms of components that are required to correctly perform the measurement, the NIRS technology and the EEG are very similar. In fact, we have a cap where are placed different electrodes that measure electrical activity inside the brain. The typology of electrodes that could be used and mounted inside the cap are a lot, as instance the active one which detect if the electrical coupling between the electrode and the scalp is correctly set up or not. Some other examples of electrodes are the needles that are implanted subcutaneously. To put in contact the electrode and the scalp we could use some conductive paste. Other example of electrodes is as instance the sponge which are small electrodes that to be correctly set up needed to be surround of saline solution. Rather than talking about the electrodes variability we could also find in literature, that are recognized worldwide, some standardized montages. These are the 10 – 20 system, the 10 – 10 and finally the 10 – 5 system. The number indicates the percentage of distance that separate two given electrodes depending on the anatomical distances of measured in the subject. Another important difference between the standardize system is the number of electrodes involved in the analysis. These are respectively 64, 75 and 300 electrodes This is important to know because affect the spatial resolution of the measurements and the different results obtainable after the processing phases.

EEG: Signal generation

What we measure is a macroscopic phenomenon which reflects the synchronized activity of extended neuronal population. The scalp EEG mainly reflect the activity of cortical neurons because it is more difficult to measure the deepest neurons inside the brain due to the low signal power generated from these cells.

In the image reported below we could see the different neurons that are presents inside the cortical cortex. We could distinguish the neurons in mainly two categories, the one that affect the measurements and the one which are not. The first types are also called *pyramidal neuron*. These are organized to be parallel one each other and they have a dendrite, called *apical dendrite*, which is orthogonal to the cortex's surface. The other neurons instead are not organized inside the cortical surface with a predominant direction and thus they do not affect too much the electrical activity measured due to the stochastic contribution of each one.

What it is measured with the EEG is the post synaptic potential. Specifically, the overall process that leads to the signal measured is the next one. From a given neuron starts an action potential which will arrive to the target neuron. Here it will be generated a post synaptic stimulus which is the signals that causes the extra cellular depolarization or hyperpolarization, depending on the types of the synapse, that will be measured and detect inside the EEG signals. Obviously, the signal measured does not come from just a single neuron but from a high amount of them that starts firing synchronously. In the next image is reported the main configuration on what happens and what we could measure with the EEG instrumentation.



In the image above is reported the summary representation of what happens and what we could measure with the EEG instrumentation. First, the four different configuration that we are going to analyse different between the types of the synapse and the position of this last one. Let consider now an excitatory synapse connected with the apical dendrite on the top. When an action potential reaches the terminal button, we have a flow of positive charge towards the inside of the dendrite. Thus outside the neuron there is a negative extracellular space. The positive charges move and comes out from the further part, from the synapse, of the dendrite. This will create a dipolar configuration where we have a negative pole towards the scalp and the positive pole towards the inner section of the brain. Thus, the potential measured is a negative deflection. The same way of thinking could be applied for the other configurations. One additional information that it is quite important to cite to describe completely the situation is that the post synaptic potential are graduate responses and thus reflect the strength of the stimulation received.

The only contributor to the EEG signals are the pyramidal cells. This because these last ones are the only thing that present a high organization inside the cortical layers. To measure the pyramidal neuron activity these types of cell needed to present a high temporal synchronicity and spatial summation that allow to generate an enough strong field potential that can be measured from the scalp. These neuronal activities generate current that flows in the head volume conductor. These are reflected as potential difference over the scalp.

EEG: Analysis – Time and Frequency domain

The types of analysis in which we could involve the EEG signal are a temporal and a frequency domain – based ones. We could perform these types of examination because the EEG signal reflects eventual stimuli proposed to the subject and because inside the history were identified four main frequencies band that are predominant inside given task performed from the subject.

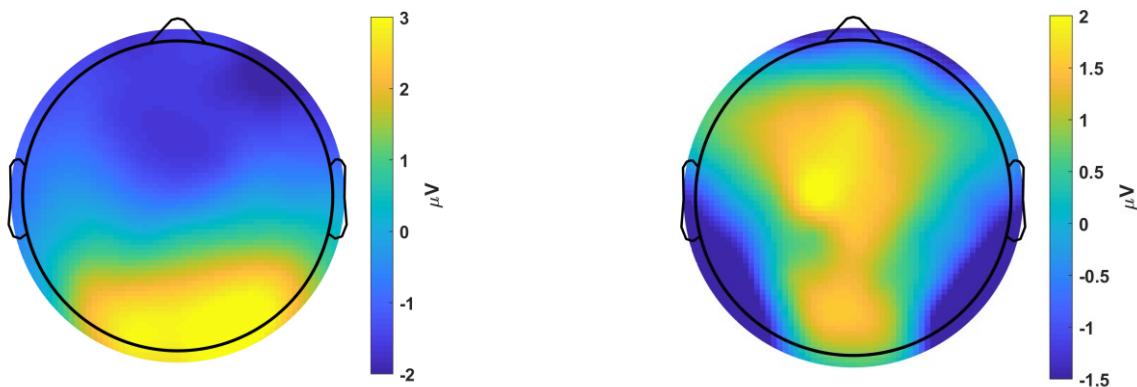
The standard temporal domain EEG technique is aiming in estimating the evoked potential. These are cognitive response to stimulus proposed. In an evoked potential we could have negative and positive deflection. In general in the EEG the positive and the negative amplitude is reverse from the standard

visualization. In fact, the ordinate values (voltage axis) are flipped. The important thing used to describe in literature an evoked potential is by the usage of two letters the P and the N. The first indicates a positive deflection of that wave while the second a negative deflection. Another important feature to characterize the different waves in the EEG signals inside an evoked potential is the number which is placed after the letters that represents the cardinality of the waves detected.

EEG: Analysis – Topographical domain

Rather than frequencies and temporal analysis, a way to represent and analyse data is through the usage of topographical maps. This way to visualize the results are the precursor method of source reconstruction. These are just projection and interpolation of the scalp measured potential. The spatial information can be derived from these maps, but their incorrect interpretation might lead to misleading conclusions regarding the source generators. We could see down below that by looking on the scalp surface there is a different topographical map even if the task performed from the subject is equal. In fact, in the image reported on the left we represent a topographical map where the subject is looking to an animal. Instead, in the image reported on the right the subject is looking another subject.

From these maps we could derive some spatial information that could lead to some misinterpretation of the real meaning of what found, and they could lead even to a bad reconstruction of the source generators. The only thing that we could say about the next topographical map is that the generator reported in the left image are just different from the one reported on the right. To understand better where the different generators are positioned, we need to introduce the source reconstruction method.



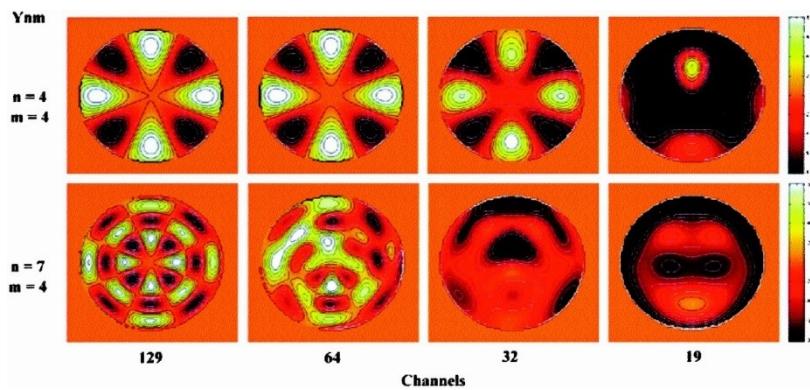
EEG: Analysis – Sources reconstruction: Introduction

The topographic maps are very useful for sources reconstruction procedure. EEG can be seen as an imaging modality when topography or source localization reconstruction is performed. The aim of source reconstruction is to localize the sources generating the potential measured at the scalp surface. So, we would like to evaluate where the potential that we have measured in the scalp are generated. Since there is not a linear propagation the reconstruction is not easy. Compared to MRI and NIRS based methods, these images show direct neural signalling information and not indirect ones as the metabolic changes. The high temporal resolution of EEG, furthermore, allows to visualize the temporal dynamics of neuronal signalling in large scale neural networks. Source reconstruction requires a forward model and an inverse problem to be solved similar to DOT image reconstruction seen for the NIR signals. From the forward model we could solve the inverse problem to find the sources that have generated that specific topographical maps.

Electrical potentials are recorded from a multitude of electrodes distributed over the scalp. From these, potential maps can be estimated at every time point, depicting the configuration of the potential field over the scalp in that given instant. This could lead to define a lot of EEG topographical maps. In this way we create a collection of maps that are associated to different instant or interval of time. When we obtain these, we know that these maps are for sure produced from different sources inside the brain. This is a preliminary step for source reconstruction. We could not use and perform a source reconstruction instant of time per instant of time with just one electrode or measurement. In order to estimate the topographic maps a proper sampling of the electromagnetic field over the whole scalp is required. To achieve a high spatial resolution we need a high number of electrodes. Usually, placing a high number of electrodes on the scalp could be time consuming. Technical advances in the field have led to EEG high density system that are commercially available and easily and quickly applicable. One example of this system is the *HydroCel Geodesic Sensor Net*. This uses sponge electrodes that are soaked with saline solution that increases the conductivity between the scalp and the electrode. This of course makes the set up and the acquisition much faster because we do not need any electrode pasta or skin abrasion to improve the adherence to the scalp.

EEG: Analysis – Sources reconstruction: Number of electrodes

One main problem, if we do not have available, some sponge electrodes EEG system, is to determine how really many electrodes we need to place over the patient scalp to reach however an enough high reliability of the result. For this question there is still not a correct answer. This depends on the spatial frequency of the scalp potential field, which is limited by the blurring caused by volume conductor effects (the skull has low conductivity). As for spatial frequency the maximal spatial frequency has to be correctly sampled to avoid aliasing hence spatial sampling should respect this rule too. This because the Nyquist rate that should be satisfied need to have twice the highest frequency of the signal to overcome aliasing. This is required because we are sampling in the temporal domain, and we are doing a discretization from the spatial point of view too. Having a spatial frequency of the potential field higher than the spatial sampling frequency distance between electrodes can distort the topographical map leading to its misinterpretation and potential mis localization of the sources. Down below there is an example of the spatial frequencies that could be resolved with a different number of electrodes. The sources that needed to be reconstruct should reflect some spherical harmonics.

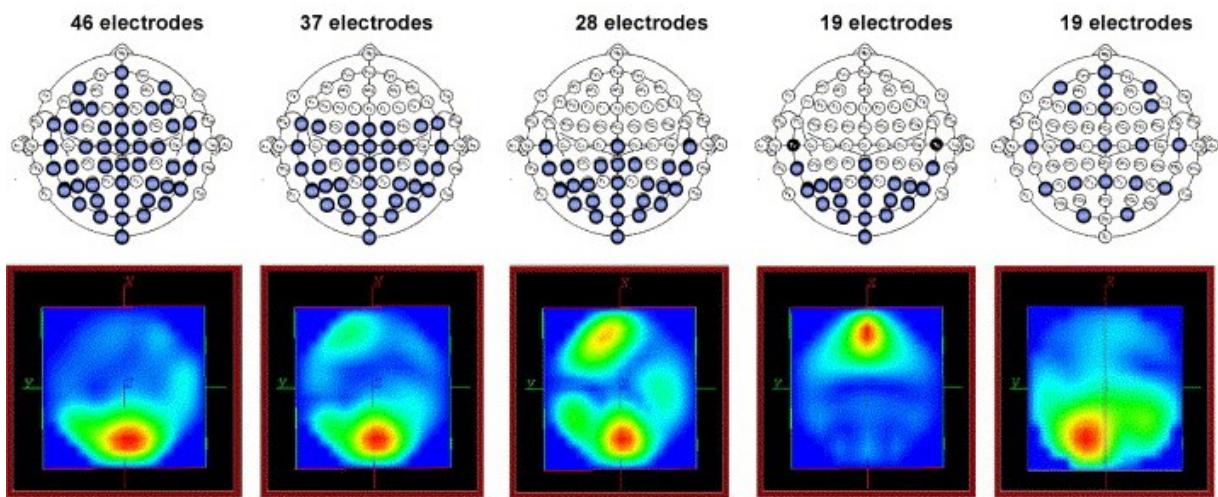


In the first case the value of n represents the azimuth which is the rotation inside the axial plane counterclockwise while m indicate the altitude. From the first to the second row the spatial frequency associated to the spherical harmonics are increased. In fact, from this visualization we could appreciate that

inside the first row until a number of 64 electrodes the measurements do not distort the pattern of the spherical harmonics. Instead, in the second row due to the half number of channels from the 129 condition the spatial distribution is not more well reconstructed anymore. Thus, from this study carried on from Luu and his colleagues, it is possible to understand that not using an appropriate number of channels for the source reconstruction we may incur in some filtering of the highest frequencies of the real signal that needed to be detected.

The EEG measurement system come with a predefine number of electrodes that could varies between 32, 64, 128 or 256 electrodes for the standard configuration. The problem that many times inside the literature has been proposed is how it is possible to choose the best number of electrodes. It has been estimated, through a modelling procedure, that the maximal spatial frequency of scalp electric field. These studies, that has carried on this analysis, suggested an inter electrode distance between the range of 2 to 3 cm at least. This means having only 100 electrodes over the scalp.

Simulation studies where they were evaluating the dipole localization error comparing different source localization algorithms and different number of electrodes, has demonstrated that localization accuracy increases as the number of electrodes increases up to 100 electrodes. Furthermore, higher is the number of electrodes, lower the localization error of deep sources is. Some other experimental studies have evaluated the source localization error adopting a down sampling procedure of the most performant array configuration and comparing the reconstructed sources with the different set up.



In the studies result reported above we could appreciate that by varying the number of electrodes considered we change also the position of the sources reconstructed. In the experimental setting the task in which the subject is involved is a visual one. Hence, the right position where we should except to reconstruct the sources are mainly over the occipital area, as seen in the image on the left. We could notice from the image reported above that as instance when we down sample from the initial to the second situation considering only the electrodes reported in the posterior part of the head there is an activation of the same occipital region but also a frontal one, which is such a misleading result. By maintaining the precedent configuration but down sample this again we reconstruct a result which is very misleading. A good final result is obtained by down sampling the electrode grid but not just placing around the head in a certain position where we should except the activation. Thus, this is a quite enormous difference from the

NIRS and DOT analysis. This because in these last two approaches we place the array configuration just over the cortical area for which we could measure the metabolic activation.

Also in some clinical application to localize the epileptic foci, we need to use an elevated number of electrodes because in this way the sources reconstruction is much more accurate. Specifically one example of scientific article that review the seizure onset in paediatric patients shows that between 128, 96 and 64 electrodes used for the EEG reconstruction does not change to much the foci area localized. Instead with a lower number of electrodes yes. In the image down below we could notice how much more reliable is the source reconstruction that implements a higher number of electrodes.



The results of studies seen so far estimated that 64 to 128 electrodes are desirable for accurate spatial sampling. However, it has been demonstrated that the spatial frequency highly depends on the conductivity values used for the different tissue types comparing the head model, above all their ratio. The previous studies simulated a conductivity ratio between skull and brain of 1 to 80 which is probably misestimate. The conductivity of the skull is likely higher, therefore the blurring smaller and the spatial frequency higher. It is likely that the brain to skull conductivity ratio is 1 to 25 or 1 to 20. With these values it has been demonstrated that spatial resolution further increases with 256 electrodes compared to 128. With the 128 electrode we could perform source reconstruction, but we will estimate them with less accuracy instead of using a 256 electrodes configuration.

Higher density arrays however are more influenced by measurement noises which is the limiting critical factor affecting spatial resolution. Thus when we need to perform source reconstruction, we need to hold in mind the number of electrodes as the number of elements that are used to reconstruct the different sources.

To summarize the higher the number of electrodes, the more accurate should be the source reconstruction. Ideally, the 256 electrodes configuration should be used when one wants to perform source reconstruction, above all when small regions of interest are the target. 128 and 64 electrodes can still be used for source reconstruction purposes, above all when larger regions of interest are the target of the analysis.

EEG: Analysis – Topographical domain: Practical concept

We have decided that we need a lot of electrodes in the scalp to have a good accuracy in the source reconstruction. The topographic analysis allows to overcome one of the limitations of standard EEG ERP analysis methods. The standard EEG and ERP analysis relies on metrics extracted from the signal. These measures are ambiguous because the EEG is a bipolar signal. If the reference electrode is changed, the pattern of the estimated evoked potential for example will change. Thus if we use a refence electrode in the center of the head or in the left hear we could detect different pattern and shape of the ERP.

Therefore in the ERP studies we need to use reference system very close to the previous performed studies to compare the results obtained. The topographic analysis is less affected by the choice of the reference and the source localization and reconstruction too. In the topographical maps the reference is changing just the zero level of the channel, thus changing for example the amplitude of some ERP components does not change the topographic feature of the map. The zero – level is than shifted and can be removed by using the common average reference at each time point. When EEG data are analysed in the frequency domain, though, the choice of the EEG reference is critical also for topographic analysis since it is not possible to have reference free potential measures.

Even in the topographical analysis we could determine and localize which electrodes that has been activated. This because, instead of the source localization, they do not localize the sources that has produced the electric field but just the different electrodes activated. Spatial details of the scalp potential maps can be sharpened calculating the scalp current source density or *surface Laplacian* of the potential. This last one is the second derivative of the potential field in the local curvature.

EEG: Analysis – Topographical domain: Laplacian surface

The surface Laplacian is useful when sensitivity to local activity has to be enhanced. It can be interpreted as an estimate of the current density entering exiting the surface. It is independent of the choice of the reference electrode because the potential common to all electrodes is automatically removed.

The scalp potential maps could be characterized by two measures the *strength* known also as the *global field power*, and the *topography* known also as *map dissimilarity measure*.

EEG: Analysis – Topographical domain: GFP

The first one follows the next analytical definition:

$$GFP = \sqrt{\sum_{i=1}^N \frac{(u_i - \bar{u})^2}{N}}$$

Where the u_i represents the voltage of the map u at the electrode i . Then we have the number of map which is represented with the letter N , and finally the value of \bar{u} which is the average voltage of all electrodes of the map u considered. The GFP maps shows a better signal to noise ratio. This measure corresponds to the spatial standard deviation.

EEG: Analysis – Topographical domain: GMD

The second one follows the next analytical representation:

$$GMD = \sqrt{\frac{1}{N} \sum_{i=1}^N \left[\frac{u_i - \bar{u}}{GFP_u} - \frac{v_i - \bar{v}}{GFP_v} \right]^2}$$

All the terms of this equation have a particular meaning. Specifically, we have that the u_i represents as before the voltage of the map u of the electrode i . Then we have the average voltage of all the electrodes indicated with \bar{u} . The same information is contained inside the v_i and \bar{v} term. Thus with this metric we are estimating the dissimilarity between two given parametric maps considered.

Usually, this metric could assume value between 0 and 2. When the value is contained inside this interval and it is much lower than the extremis this means that the two considered maps are different. Instead, if the values is approximative equal to 0 or 2 the meaning are respectively, the fact that the two maps are equal and the two maps presents the same topography but with reversed polarity.

If two maps differ in topography independent on their strength: the two maps are generated by a different configuration of sources in the brain. The opposite is not always true: an infinite number of source configurations can produce the same scalp potential topography. GMD is the first step in order to evaluate whether the processes begin compared could be generated by different sources.

LESSON 25: EEG SOURCES RECONSTRUCTION

In the previous lecture we have seen a brief introduction on what the EEG is. We have learned that the number of electrode and how the electrode are distributed over the scalp could modify what we could finally get in terms of accuracy of the spatial and the temporal information. Higher is the number of electrode higher is the accuracy in reconstruction of the different sources. Low density electrodes configuration could not be able to reconstruct completely the sources time courses. One of the last concepts analyse is that at different topographical maps are associated different sources reconstruction.

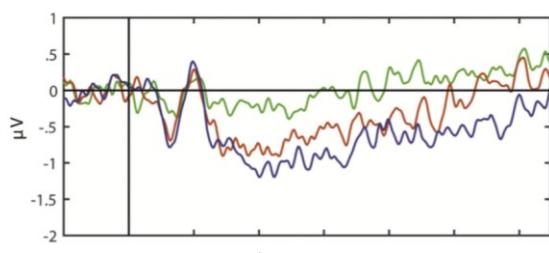
EEG: Analysis – Spatial temporal decomposition

The first step that we could perform inside our analysis to reach the sources reconstruction final result, is what is called the *spatial temporal decomposition*. The source reconstruction can be applied at any time point of the considered EEG signal.

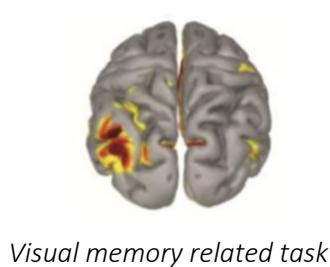
EEG: Analysis – Spatial temporal decomposition: ERP computation

The first approach is the simplest and it relays on the computation of the ERP in the time domain. These types of signals show specific features inside this domain such as waves and characteristic amplitude that made up the signal. The objective with this approach consists of in to be able to reconstruct the source of the P300 or N1 waves. The technique in details require to start from an ERP in the time domain. Hence, we should take the EEG records and isolate the temporal window where the ERP should be activated and calculating the average to enhance the ERP and to improve SNR. Once did this task we should compute the sources reconstruction algorithm.

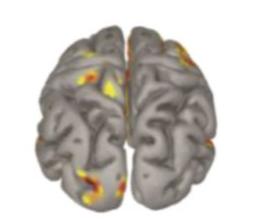
Down below is reported an example of task related activity where we have a subject that it is performing a memory visual related task. The different type of ERP that we could reconstruct are highlighted in different colour. The question is where are located these sources that made up this patter. We could average the time interval that present this answer and then we could perform some sources identification. As we could see in the image reported below, we could appreciate and notice the activation of the *intra parietal circus* correspondently to the task performed by the subject.



ERP Time domain



Visual memory related task



Control related task

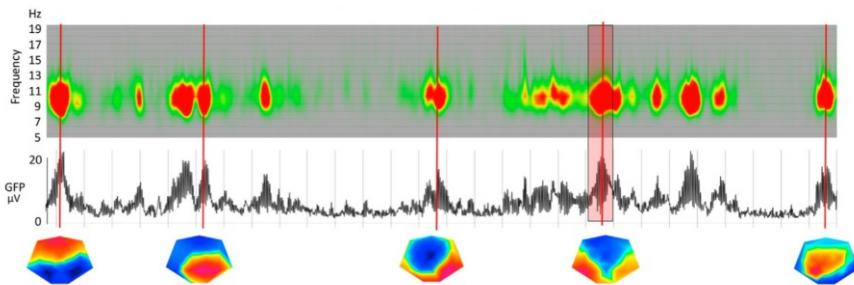
However, even if it is one of the simplest approaches to enhance the spatial localization of certain ERP this is not one of the best because we could detect only the ERP are activated in a specific position. When we have high density recording that are available, we have two different problems even with the EEG itself. The first is that the scalp sites have different peak and latencies. Moreover the waveform at certain electrodes changes if the position of the reference electrode is moved.

EEG: Analysis – Spatial temporal decomposition: Frequency domain

An alternative approach consists in defining the components based on the topography of the potential field. The methods that could perform this reconstruction are called the *spatial temporal decomposition methods*. These last family of approaches find a series of distinct components or modes that describe the EEG recording and some components that can differ between conditions. We need to isolate some of these and to solve the sources reconstruction problem.

The idea is to find a limited number of optimal factors that best represent a given dataset. Their load will vary with time. With this analysis we have a lot of other components that will be so neglected. Thus, as final results, we could have different components at distinct time points. Each factor represents a certain potential topography, and we can apply source localization to these maps so as to estimate a limited number of sources in the brain that explain a full time series of multi – channels EEG signals.

Down below we could see the frequency analysis and the GFP and the sources distribution. This figure holds the conclusion that it is not a good assumption that the source generator does not change their activation in time. In fact in this figure we could see the α frequency band where are mainly located the major quantity of frequency component that are activated at different instant of time. If we isolate the topography associated to the different α burst these maps are not the same because the sources that has product the stimulation is changed. This is a good way to choose the time points where to perform the sources reconstruction. Thus, the recap we need first to choose the time points associated to variation of the frequency component inside a given frequency band and then we compute the different topographical maps associated to these time points.



EEG: Analysis – Spatial temporal decomposition: PCA and ICA

The most commonly used spatial decomposition are based on *principal component analysis (PCA)* and the *independent component analysis (ICA)*.

The PCA imposes orthogonality between factors, the first component accounts for the maximally possible amount of data variance and each following components accounts for the maximum possible residual variance. The factors that contribute little to the explained variance can be neglected. Thus the results will be a reduction of the multi – channel data in space and in time. This is a different application instead of the one seen for NIRS signal. In this last one in fact we have used this technique as motion correction approach while here we use it to perform some sources reconstruction. Usually, the components that are mostly considered are the one with the highest value of variance explained associated to them, instead, the remaining are neglected because the sources associated are not explaining a high variance of the EEG signal.

The other method exploits the *ICA* which imposes statistical independence between factors. Each factor is supposed to represent a temporally independent component and to each factor ICA assigning a weight coefficient. ICA can be used to decompose the brain activity into a number of maps that can be used for source reconstruction. The ICA is used to solve the so called the cocktail party problem. The assumption of statistical independence inside the sources that generates the EEG signals means that we have not cross talk between different brain area. Thus the ICA could not uncover the regions of the brain that are actually speaking with each other.

EEG: Analysis – Spatial temporal decomposition: k – means clustering

Another spatial temporal decomposition approach is based on the spatial *k – means clustering*. This does not assume any constraints instead of the previous technique. The idea is that we can use GMD to cluster maps with high spatial correlations and determine the centroid of these clusters as representative template map for each cluster. We could have different topography that could represents our signals and thus thanks to the different topography we could use them to cluster the data. In particular we iteratively apply some *k – means* algorithm with the best number of clusters to optimally explain the data with a minimal number of maps. Once find the different centroids, when we fit the cluster map to the data, usually one specific maps dominates for a certain amount of time. This maps that dominate for a time interval is called *EEG microstates*.

EEG: Analysis – Forward problem – Model: Equivalent current source

To solve the forward problem, first a *source model* should be defined. To solve the forward problem with the DOT, we need the *head model* (solution space) and how the light was propagated inside the scalp once a light beam was delivered. In the EEG we should have the same information. Here, instead of considering how light travel inside our head, we need to assign electrical properties to each source. Neuronal currents can be modelled as an electrical current dipole composed of a pair of elements: the *source* and *sink* with an infinitely small inter distance. This *equivalent current dipole source model* is the simplified version of the source that produces the potential measurable from the scalp. This model works well when the brain activity is concentrated in a given brain region. Some examples are related with pathology like *epilepsy*.

EEG: Analysis – Forward problem – Model: Distributed current source

In literature if we do not want to use the equivalent current dipole source model are available some models which are more complex but more exhaustive. There are are the *distributed current dipole source models*. Here the neuronal activities over a small region are modelled by a current dipole located at each region. Therefore the brain activity with any distribution of neuronal currents can be represented by a source model consisting of a distribution of current dipoles evenly placed within the entire brain volume. At each location, there are three orthogonal dipoles known as the *volume current density model*. This is a model where we have distributed sources.

EEG: Analysis – Forward problem – Model: Cortical current density

When we are solving the forward model which means find the topographical maps or voltages that the different electrode would measure given a determine sources activities. One way to improve the model is

that we could constraint the current source on just the grey matter. This could be useful to find a certain solution for the application. Thus we need some anatomical information from MR images, and we could also impose that the current flow is perpendicular to the cortical surface. Thus we could create a more constraint model known as the *cortical current density model*. Thus we need to use some a priori information such as anatomical ones, to estimate the electrical activity inside the brain. In this way we could apply a certain dipole in just a certain position thanks to some anatomical knowledge. The dipole will represent the overall post synaptic activity of the neurons that are located in that region of the cortex where we have placed the dipole.

EEG: Analysis – Forward problem – Head volume conductor

Solving the forward problem means to obtain the distribution of the electrical potential Φ on the scalp surface given any known distribution of current density inside the brain and the conductivity values of the head model. Specifically, we could solve the forward problem by knowing the current at each source and then project this on the scalp surface. To complete this operation we need at least one another information. This is the head model. In EEG analysis is called the *head volume conductor model*. Several volumes conductor's models have been proposed through the years. The first and the simplest is known as *the infinite homogeneous model*. The one that shows the most realistic geometry, and it is the most complex is called the *inhomogeneous head model*. The difference between these two are the conductive properties of the mediums.

EEG: Analysis – Forward problem – Head volume conductor: Infinite homogeneous model

In the *infinite homogenous model* in terms of conductive medium we could describe the electrical potential over scalp electrodes with the next relation:

$$\Phi = \frac{1}{4\pi\sigma} \int \nabla \left(\frac{1}{r} \right) \vec{J} d\nu$$

In this equation we have first with Φ the electric potential over scalp electrodes. Then we have the electrical conductivity σ , the gradient operator ∇ applied over the inverse of the radius. Finally we have the source element behaving like a dipole and representative of a neuronal currents indicated with \vec{J} . This equation is described as the driving force of the electrical activity that could be measured and finally it is a very simple but unrealistic model.

EEG: Analysis – Forward problem – Head volume conductor: Spherical models

In fact, this is not more used because there are a lot of solution discovered through the years that improve the results and better explains what happens. The spherical models represent and model a head with a sphere. These models are recognized as these three separated classes. The *single sphere*, the *three concentric sphere* and finally the *four concentric sphere models*. The number of spheres represents the number of tissues that are considered and modelled. In this type of model, we assume that electrical properties inside a sphere are constant inside all its points. Thus, we could compute the electrical potentials on the scalp surface due to a current dipole by using some analytical expression or with some computational analysis. In particular, with just the single sphere model is possible to find the analytical expression, while

for the other this is too complex, and we neglect to do this. The three spheres model has been widely used in the past since incorporating the skull surface. This is a good approximation for the head volume conductor. The head shape inside this model is ignored.

The spherical models do not distinguish between areas of the brain that contains grey matter and what it is white matter. The accuracy of the EEG forward solution can be improved by constraining the source space to the grey matter. The simplest approach is to map the individual tissue segmentation mask derived from the MRI to a sphere and use a multi spheres model with the solution space constrained to the grey matter.

EEG: Analysis – Forward problem – Head volume conductor: BEM

More advance and accurate technique uses head volume conductor models by keeping into consideration of the shapes of the head and the conductivity inhomogeneity of the tissues. The most popular forward solution is based on the *Boundary Element Method* known as *BEM*. This last one provides forward solutions imposing homogeneous conductivity profiles within each compartment since only surfaces are modelled. We do not have the same model obtained with FEM modelling approaches, but we have just the external surface hence the name of this technique. Thus, to model the surface, are not needed any tetrahedron but just triangles. Thus, by modelling the different compartment in this way we could assume homogeneity inside the electrical properties of the brain in a given compartment bounded with the surface mesh.

EEG: Analysis – Forward problem – Head volume conductor: FEM

To model inhomogeneous conductivity distribution within the head we could use even the *Finite Element Method* known as *FEM*. This could be useful to model the inhomogeneous conductivity distribution of the head inside a given boundaries. The challenge of FEM approach is the need to build the FEM mesh from the MRI and the TC images of the subject. The fact that the FEM mesh is not simple to retrieve is because it is not very simple to perform segmentation to the MR images coming from scanner. The disadvantages to use this method instead of the other is the requirement for the subject to undergo MR scanning other than registering only EEG signals. However, even if it requires a lot of information, this is the best model and the most accurate that we could use to describe our situation. It is still costly because the measures that needed to be performed are a lot and are very complicated.

EEG: Analysis – Forward problem – Lead Field Matrix

Once head model, the electrode positions and source model have been decided, the forward model can be solved. The conductivity values should be assigned to each tissue type. There are several studies that have tried to estimate the conductivity value. Down below is reporting a summarizing table of the value that were proposed during time.

compartments	Geddes & Baker (1967)	Oostendorp (2000)	Gonçalves (2003)	Gutierrez (2004)	Lai (2005)
scalp	0.43	0.22	0.33	0.749	0.33
skull	0.006 – 0.015	0.015	0.0081	0.012	0.0132
cerebro-spinal fluid	-	-	-	1.79	-
brain	0.12 – 0.48	0.22	0.33	0.313	0.33
$\sigma_{\text{scalp}}/\sigma_{\text{skull}}$	80	15	20–50	26	25

To solve mathematically the forwards problem we need to use the *Poisson equation* that links the potentials at any position in a volume conductor model with the applied current sources. The equation is the one reported below:

$$\nabla \cdot J = I\delta(r - r_1) - I\delta(r - r_2) = I_m$$

Where the first term represents the divergence of the vector field *current density*. This represents how much a current is exploding from a given point. Then we have a contribution of the current source position defined as the first term, where the δ refers to the current coming out the source or falling into a sink. The minus term instead represents the amount of current that ends up in a sink.

Then from the Poisson equation we could use the *ohm's law* and estimating the current density. Analytically:

$$J = \sigma E$$

By exploiting the previous relation we could substitute inside the Poisson equation the density current term with the electrical field obtaining a divergence applied over the electrical filed. Knowing that the electrical field could be described as the inverse of the gradient applied over the potential field we could obtain:

$$E = -\nabla V \Rightarrow \nabla \cdot [\sigma \nabla V] = -I_m \Rightarrow \nabla \cdot [\sigma \nabla V] = -I\delta(r - r_2) + I\delta(r - r_1)$$

The gradient operator of the potential field represents the direction where the potential changes. If we put together all the equation, as previously done, we could write the final Poisson equation that would be used to solve the forwards problem.

The potential V are computed for a given current source density I_m in a volume conductor model. The forward solution is usually expressed as the *Lead Field Matrix*. This represents the potential measured form the electrodes due to the different dipoles. Entries of this matrix represent what the electrode would be measuring when the dipole have a certain orientation. We estimate what we could see on the surface depending on the dipole that we have in our model. Now, that we have exhaustively defined the forward problem we could define the inverse one.

EEG: Analysis – Inverse problem

In the EEG inverse problem, the brain electrical sources are reconstructed from scalp recorded EEG signals. This method is referred to as a *sources localization* when the neuronal activity is assumed to be localized in a few local regions. The source localization problem is just a special case of inverse problem, and this happens only when we decide to use a distributed models for the head volume conductor.

EEG: Analysis – Inverse problem – Dipole Source Localization

For *dipole source localization (DSL)* is most classic solution to EEG inverse problem. This approach works when the primary generators of scalp recorded EEG are localized in one or few smaller regions. Given a specific dipole source model, DSL solves the inverse problem using a nonlinear multidimensional optimization procedure estimating the dipole parameters that best explain the observed scalp potential measurements in a least square sense. All the dipole localization algorithms need a priori knowledge of the

number and classes of the underlying dipole sources. If we have too many dipoles will give some problematic solutions. One way to solve this problem is to have some a priori information regarding the possible solution space in term that it is available. We could constraint the source space solution to have better accuracy. The dipole source localization method has been widely employed to try to localize and map the epileptic foci or the sensory motor cortex before surgery.

EEG: Analysis – Inverse problem – Distributed Brain Source Imaging

This type of approach has been improved with the more complex approach which is called the *distributed brain source imaging*. This offer a more exciting options for imaging and localizing brain functions since no a priori assumptions on the number of dipoles and dipolar nature of sources are required.

The distributed source imaging methods are no ad hoc assumption on the number or configuration of brain electrical sources. The equivalent sources are distributed over the sources space. All the dipoles have a fixed dataset location, and we are considering a thousand of dipoles. These current dipoles are fixed at present locations. The inverse problem aims to estimate the dipole moments by minimizing the difference between measurements and source model predicted scalp potentials. We need to minimize and find the dipole distribution that mimic best the data measured. Thus we could use the relation reported below:

$$\phi = AJ + n$$

In the previous equation the value of ϕ represents the scalp measurements, A the transfer matrix or lead field, J the source distribution and finally n the measurement noise. The estimate of J is equivalent to design an inverse filter B , which can project the measured data into the solution space. This could be written as:

$$J = B \Phi$$

The distributed source imaging can therefore be a linear inverse problem. However it is undetermined due to the number of unknown distributed sources which are greater than the number of scalp electrodes. Moreover, some unique and well posed solution could be obtained if some constraints are imposed. This linear inverse problem could be solved with the general inverse solution or minimum norm least square inverse. This minimizes the least square error of the estimated inverse solution J under the constraint.

$$\Phi = AJ$$

Since the problem is underdetermined when inverting the *Lead Field*, we need to apply regularization to enforce a level of smoothness in the inverted results. Usually, the Tikhonov regularization is used.

$$J = WA^T(AWA^T + \alpha_R I)^{-1}\phi$$

The value of α_R represents the Tikhonov regularization term. The value of this term and parameter could be obtained with the next set of relations:

$$\alpha_R = R$$

$$\alpha = \frac{\max(\text{Eigenvalues}(AWA^T))}{20000}$$

By changing α we could change even the value of the solution found.

EEG: Analysis – Inverse problem – WMN, LORETA, LAURA

The famous variation of this approach are the *lead filed normalized weighted minimum norm (WMN)*, the *low-resolution brain electromagnetic tomography (LORETA)* and the *local autoregressive average (LAURA)*.

LORETA estimates a WMN solution where the weighting is a discrete Laplacian operator thus the estimate is smooth. It normalizes the sources estimates with the corresponding noise sensitivity, the statistical significance of the inverse solution can be assessed and therefore maps of source estimate statistics (*sLORETA standardize LORETA*) could estimate the statistical maps of the generator maps.

```
1 %
2 % Title: Laboratory 11
3 %
4 % Note:
5 % Source reconstruction
6 %
7 % Author: Matteo Martin
8
9 % THEORY:
10 %
11 % The Toolbox that are used for this Laboratory is the FieldTrip toolbox.
12 % The objective of this lab is to start from a topographical map and solve
13 % the inverse problem to obtain the source distribution and their activity
14 % inside the scalp. To do this we need first to visualize the differences
15 % between the topographical map obtained with 64 and 256 electrodes
16 % respectively and then solve the inverse problem.
17 %
18 % DATA:
19 %
20 % The data that are available for this laboratory are some 64 and 256
21 % electrodes measurements. Inside the Data folder we could find the
22 % previous mentioned foleders. Then inside each one we have some MR data
23 % and some simulation data. Inside the MR folder we could find the
24 % electrodes position (.spf) which is a text file where are reported all
25 % the information about the electodes position registered over the head
26 % volume that will be used as solution space for the foward problem. Thus,
27 % inside the same foler is contained also the anatomy pre processed head
28 % model which represents the head model or our solution space.
29 % Then we have the simulated data, which are subdivided into the cerebellum
30 % file and the two dipoles file. The first one is a measurements of the
31 % topographical potential distribution over the scalp analysed and measure
32 % during some cerebellum activities. Then we find the two dipoles file
33 % which are just some simulation of the activity placed somewhere inside
34 % the brain which has to be reconstructed.
35 %
36 % PROCEDURE:
37 %
38 % The procedure holds first to load all the required file and add all the
39 % path of the different Toolbox inside the path folder of Matlab. After
40 % having done this we need to first inspect the simulated and the measured
41 % data via the funciton topoplot fornished from the EEGLab toolbox.
42 % Then we need to compute and visualize the topograhalical map of the lead
43 % filed loaded from the anatomy pre pro head model .mat file available
44 % inside the MR folders. After having done this we perform the sources
45 % reconstruction problem by first estimating the filtering matrix and then
46 % multiplying this matrix via a matrixial product for the measurement
47 % matrix. To conclude the analysis we need to visualize and compare the
48 % different results by changing the value of the regularization paramter.
49
50 %% CCC
51 % Clear all, close all, clear the comand window
52
53 close all; clear all; clc
54
```

```

55 %% DATA - LOAD
56 % Load of the data that are contained inside the folder data.
57
58 path64 = ['C:\Users\teo98\Documents\Uni\Magistrale\I_IMGNEURO\',...
59     'Laboratorio\Laboratory11\Data\64ch'];
60 path256 = ['C:\Users\teo98\Documents\Uni\Magistrale\I_IMGNEURO\',...
61     'Laboratorio\Laboratory11\Data\256ch'];
62 SPM0641D = spm_eeg_load([path64, '\dataset1\sim_data\cerebellum']);
63 SPM02561D = spm_eeg_load([path256, '\dataset1\sim_data\cerebellum']);
64 SPM0642D = spm_eeg_load([path64, '\dataset1\sim_data\two_dipoles']);
65 SPM02562D = spm_eeg_load([path256, '\dataset1\sim_data\two_dipoles']);
66
67 %% DATA - VISUALIZATION
68 % Visualization of the sources stimulated data inside the cerebellum
69
70 figure,
71
72 subplot(1,2,1),topoplot(SPM0641D(:,:,1,1),[path64,...
73     '\dataset1\mr_data\electrode_positions.sfp'],'electrodes','off'),
74 title('64 CHANNEL CONFIGURATION')
75
76 subplot(1,2,2),topoplot(SPM02561D(:,:,1,1),[path256,...
77     '\dataset1\mr_data\electrode_positions.sfp'],'electrodes','off'),
78 title('256 CHANNEL CONFIGURATION')
79 sgttitle('1 DIPOLE')
80
81 figure,
82
83 subplot(1,2,1),topoplot(SPM0642D(:,:,1,1),[path64,...
84     '\dataset1\mr_data\electrode_positions.sfp'],'electrodes','off'),
85 title('64 CHANNEL CONFIGURATION')
86
87 subplot(1,2,2),topoplot(SPM02562D(:,:,1,1),[path256,...
88     '\dataset1\mr_data\electrode_positions.sfp'],'electrodes','off'),
89 title('256 CHANNEL CONFIGURATION')
90 sgttitle('2 Dipoles')
91
92 %% DATA - LOAD
93 % Load of the data that are contained inside the folder data.
94
95 path64 = ['C:\Users\teo98\Documents\Uni\Magistrale\I_IMGNEURO\',...
96     'Laboratorio\Laboratory11\Data\64ch'];
97 path256 = ['C:\Users\teo98\Documents\Uni\Magistrale\I_IMGNEURO\',...
98     'Laboratorio\Laboratory11\Data\256ch'];
99 HM64 = load([path64, '\dataset1\mr_data\anatomy_prep_headmodel.mat']);
100 HM256 = load([path256, '\dataset1\mr_data\anatomy_prep_headmodel.mat']);
101
102 %% DATA - VISUALIZATION
103 % Visualization of the different lead field measured via the 64 and the 256
104 % electrodes EEG configuration.
105
106 I = HM64.leadfield.inside; P = find(I == 1);
107 P5 = P(5); D = zeros(64,3,length(P));
108 for idP = 1:1:length(P), D(:,:,:idP) = HM64.leadfield.leadfield{P(idP)};end

```

```

109
110 figure,
111 subplot(1,4,1), topoplot(sum(D(:,1),3), [path64, ...
112     '\dataset1\mr_data\electrode_positions.sfp'], ...
113     'electrodes','off','electrodes','off'); title('X')
114 subplot(1,4,2), topoplot(sum(D(:,2),3), [path64, ...
115     '\dataset1\mr_data\electrode_positions.sfp'], ...
116     'electrodes','off','electrodes','off'); title('Y')
117 subplot(1,4,3), topoplot(sum(D(:,3),3), [path64, ...
118     '\dataset1\mr_data\electrode_positions.sfp'], ...
119     'electrodes','off','electrodes','off'); title('Z')
120 subplot(1,4,4), topoplot(sum(D(:,:,2, 3)), [path64, ...
121     '\dataset1\mr_data\electrode_positions.sfp'], ...
122     'electrodes','off','electrodes','off'); title('SUM')
123 sgtitle('5^{th} SOURCE DIPOLE')
124
125 I = HM256.leadfield.inside; P = find(I == 1);
126 P5 = P(5); D = zeros(256,3,length(P));
127 for idP = 1:1:length(P), D(:,:,idP) = HM256.leadfield.leadfield{P(idP)};end
128
129 figure,
130 subplot(1,4,1), topoplot(sum(D(:,1),3), [path256, ...
131     '\dataset1\mr_data\electrode_positions.sfp'], ...
132     'electrodes','off','electrodes','off'); title('X')
133 subplot(1,4,2), topoplot(sum(D(:,2),3), [path256, ...
134     '\dataset1\mr_data\electrode_positions.sfp'], ...
135     'electrodes','off','electrodes','off'); title('Y')
136 subplot(1,4,3), topoplot(sum(D(:,3),3), [path256, ...
137     '\dataset1\mr_data\electrode_positions.sfp'], ...
138     'electrodes','off','electrodes','off'); title('Z')
139 subplot(1,4,4), topoplot(sum(D(:,:,2, 3)), [path256, ...
140     '\dataset1\mr_data\electrode_positions.sfp'], ...
141     'electrodes','off','electrodes','off'); title('SUM')
142 sgtitle('5^{th} SOURCE DIPOLE')
143
144 %% SOURCE ANALYSIS - COMPUTATION (CEREBELLUM)
145 % Computation of the sources reconstruction analysis via the ELORETA
146 % algorithm for the cerebellum data measured over the scalp.
147
148 options.method      = 'eloreta';
149 options.eloreta.lambda = 0.05;
150
151 S64   = sourceanalysis([path64, '\dataset1\sim_data\cerebellum'], ...
152   [path64, '\dataset1\mr_data\anatomy_prep_headmodel.mat'], ...
153   'sourceWaveform64', options);
154 S256 = sourceanalysis([path256, '\dataset1\sim_data\cerebellum'], ...
155   [path256, '\dataset1\mr_data\anatomy_prep_headmodel.mat'], ...
156   'sourceWaveform256', options);
157
158 source64 = S64.imagingkernel * SPMO641D(:,1,1);
159 source64 = reshape(source64, [3, 4406]);
160 source256 = S256.imagingkernel * SPMO2561D(:,1,1);
161 source256 = reshape(source256, [3, 4406]);
162

```

```
163 source64S = mat2gray(squeeze(sum(source64, 1)))';
164 source256S = mat2gray(squeeze(sum(source256, 1)))';
165
166 visualize_slices([path64, ...
167     '\dataset1\mr_data\anatomy_prepro_headmodel.mat'], source64S, 0.9),
168 sgttitle('64 CH - \lambda = 0.05 - Cerebellum')
169 visualize_slices([path256, ...
170     '\dataset1\mr_data\anatomy_prepro_headmodel.mat'], source256S, 0.9),
171 sgttitle('256 CH - \lambda = 0.05 - Cerebellum')
172
173 %% SOURCE ANALYSIS - COMPUTATION (TWO DIPOLES)
174 % Computation of the sources reconstruction analysis via the ELORETA
175 % algorithm for the two dipoles data measured over the scalp.
176
177 options.method      = 'eloreta';
178 options.eloreta.lambda = 0.05;
179
180 S64TD = sourceanalysis([path64, '\dataset1\sim_data\two_dipoles'], ...
181     [path64, '\dataset1\mr_data\anatomy_prepro_headmodel.mat'], ...
182     'sourceWaveform64TD', options);
183 S256TD = sourceanalysis([path256, '\dataset1\sim_data\two_dipoles'], ...
184     [path256, '\dataset1\mr_data\anatomy_prepro_headmodel.mat'], ...
185     'sourceWaveform256TD', options);
186
187 source64TD = S64TD.imagingkernel * SPM0642D(:,1,1);
188 source64TD = reshape(source64TD, [3, 4406]);
189 source256TD = S256TD.imagingkernel * SPM02562D(:,1,1);
190 source256TD = reshape(source256TD, [3, 4406]);
191
192 source64STD = mat2gray(squeeze(sum(source64TD, 1)))';
193 source256STD = mat2gray(squeeze(sum(source256TD, 1)))';
194
195 visualize_slices([path64, ...
196     '\dataset1\mr_data\anatomy_prepro_headmodel.mat'], source64STD, 0.9),
197 sgttitle('64 CH - \lambda = 0.05 - Two dipoles')
198 visualize_slices([path256, ...
199     '\dataset1\mr_data\anatomy_prepro_headmodel.mat'], source256STD, 0.9),
200 sgttitle('256 CH - \lambda = 0.05 - Two dipoles')
201
202 %% SOURCE ANALYSIS - COMPUTATION (TWO DIPOLES - LOW REGULARIZATION)
203 % Computation of the sources reconstruction analysis via the ELORETA
204 % algorithm for the two dipoles data measured over the scalp.
205
206 options.method      = 'eloreta';
207 options.eloreta.lambda = 10^-5;
208
209 S64TDL = sourceanalysis([path64, '\dataset1\sim_data\two_dipoles'], ...
210     [path64, '\dataset1\mr_data\anatomy_prepro_headmodel.mat'], ...
211     'sourceWaveform64TDLLOW', options);
212 S256TDL = sourceanalysis([path256, '\dataset1\sim_data\two_dipoles'], ...
213     [path256, '\dataset1\mr_data\anatomy_prepro_headmodel.mat'], ...
214     'sourceWaveform256TDLLOW', options);
215
216 source64TDL = S64TDL.imagingkernel * SPM0642D(:,1,1);
```

```
217 source64TDL = reshape(source64TDL, [3, 4406]);
218 source256TDL = S256TD.imagingkernel * SPMO2562D(:,1,1);
219 source256TDL = reshape(source256TDL, [3, 4406]);
220
221 source64STDL = mat2gray(squeeze(sum(source64TDL, 1)))';
222 source256STDL = mat2gray(squeeze(sum(source256TDL, 1)))';
223
224 visualize_slices([path64,...
225     '\dataset1\mr_data\anatomy_repro_headmodel.mat'], source64STDL, 0.8),
226 sgttitle('64 CH - \lambda = 10^{-5} - Two dipoles')
227 visualize_slices([path256,...
228     '\dataset1\mr_data\anatomy_repro_headmodel.mat'], source256STDL, 0.8),
229 sgttitle('256 CH - \lambda = 10^{-5} - Two dipoles')
230
231 %% SOURCE ANALYSIS - COMPUTATION (TWO DIPOLES - HIGH REGULARIZATION)
232 % Computation of the sources reconstruction analysis via the ELORETA
233 % algorithm for the two dipoles data measured over the scalp.
234
235 options.method      = 'eloreta';
236 options.eloreta.lambda = 100;
237
238 S64TDH = sourceanalysis([path64, '\dataset1\sim_data\two_dipoles'], ...
239     [path64, '\dataset1\mr_data\anatomy_repro_headmodel.mat'], ...
240     'sourceWaveform64TDHIGH', options);
241 S256TDH = sourceanalysis([path256, '\dataset1\sim_data\two_dipoles'], ...
242     [path256, '\dataset1\mr_data\anatomy_repro_headmodel.mat'], ...
243     'sourceWaveform256TDHIGH', options);
244
245 source64TDH = S64TDH.imagingkernel * SPMO642D(:,1,1);
246 source64TDH = reshape(source64TDH, [3, 4406]);
247 source256TDH = S256TD.imagingkernel * SPMO2562D(:,1,1);
248 source256TDH = reshape(source256TDH, [3, 4406]);
249
250 source64STDH = mat2gray(squeeze(sum(source64TDH, 1)))';
251 source256STDH = mat2gray(squeeze(sum(source256TDH, 1)))';
252
253 visualize_slices([path64,...
254     '\dataset1\mr_data\anatomy_repro_headmodel.mat'], source64STDH, 0.9),
255 sgttitle('64 CH - \lambda = 100 - Two dipoles')
256 visualize_slices([path256,...
257     '\dataset1\mr_data\anatomy_repro_headmodel.mat'], source256STDH, 0.9),
258 sgttitle('256 CH - \lambda = 100 - Two dipoles')
```