AALBORG UNIVERSITY
STUDENT REPORT

Evaluation of Basic Preprocessing, aCompCor and FSL FIX for Optimal fMRI Preprocessing Tool in an Individual Differences Study Perspective

3nd semester Masters, Biomedical Engineering & Infomatics - Fall 2018

Project group: 19gr9411

Christian Korfitz Mortensen, Martin Alexander Garenfeld

ii

Preface

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetuer a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetuer. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Contents

WORKSHEETS

1 | Background

The background chapter has the purpose of providing the theoretical overview for some of the essential techniques and methods utilized in this project, thereby laying the foundation for understanding how these methods work and how variations of these can be used for the problems the current project aims to solve. With the main goal of finding the optimal method for denoising functional Magnetic Resonance Images, the background chapter will introduce the following topics: first a brief overview of MRI physics and image reconstruction describing the origin behind the images is made, secondly the theory behind fMRI is introduced explaining the concept of mapping brain activity, afterwards different designs for inducing stimuli is presented, subsequently a pool of common preprocessing methods are introduced followed by sections explaining the principles of independent components analysis and principal component analysis, finally a section describing the commonly used method for testing statistical significance is made.

1.1 MRI physics

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technology, which does not involve potentially damaging ionizing radiation as in other scanners, eg. CT and X-ray. MRI is especially suited for representing soft tissue portions of the body, which makes it a widely used technology in brain imaging, both for clinical and research purposes, as it can depict the anatomical structure at a millimeter resolution. This section will provide information on the physics behind MRI and which physiological properties that can be exploited to create an image of the body.

Magnetic resonance imaging (MRI) is founded on the principle of nuclear magnetic resonance (NMR), which exploits the magnetic properties of the hydrogen nucleus that contains a single proton. The proton is not static, but rotates around its own axis. As the proton is positively charged it creates a magnetic moment in the direction described by the thumb rule, and can interact with an external magnetic field. The human body consists of approximately 10% hydrogen atoms, but as the hydrogen nuclei spins are randomly orientated, the net magnetic moment equals zero, as the nuclei cancel each other out. Placing the body in a strong magnetic field will align the nuclei. A property of the hydrogen nucleus is its quantum spin rate, which can either be $\frac{1}{2}$ or $-\frac{1}{2}$ either in the direction or the opposite direction of the main magnetic field. Most will align in the direction of the magnetic field, while the rest align in the opposite direction, possibly as a result of heat radiation absorbed by the nuclei. The direction of the nucleus is determined by its energy level, leaving the former in a low energy state and the latter in a high energy state. The nuclei do not simply point in the direction or opposite the direction of the magnetic field, but precess. [1] The rate of precession can be calculated by the Lamour frequency:

$$f = y * B_0 \tag{1.1}$$

f is precession frequency, y is gyroscopic ratio and B_0 is magnetic field strength. The equation states that the precession frequency is proportional to the strength of the magnetic field.

After canceling out all opposing precessing nuclei, the net magnetization, or longitudinal magnetization, will point in the direction of the external magnetic field. However, the longitudinal magnetization can not be detected directly as it points in the direction of the strong external magnetic field. Additional techniques are therefore used in NMR, to facilitate a detectable signal. [1] A depiction of how the nucleus precessing can either align along or opposite to the magnetic field depending on its energy state, can be found in figure 1.1.

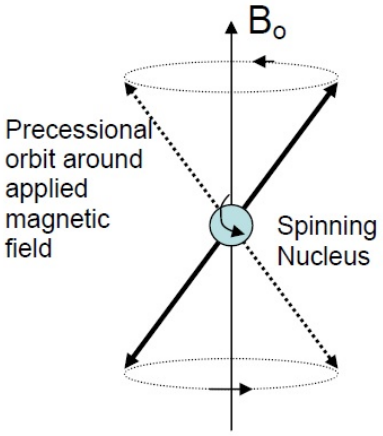


Figure 1.1: The figure illustrates how the nucleus precess and spin in relation to the applied magnetic field B_0 surrounding it. The vectors, indicating the precessing, can go opposite or along the magnetic field depending on the nucleus energy state. [2]

A radio frequency pulse (RF pulse) tuned to the precession of the nuclei is transmitted in the vicinity of the nuclei. The RF pulse is absorbed by the nuclei and more, favorably half of the targeted nuclei population, will enter the high energy state, leaving the longitudinal magnetization to equal zero. The number of nuclei that flip is determined by the amount of energy the RF pulse injects, and the nuclei only exchange energy efficiently if the frequency of the energy from the RF pulse matches the precession rate. The RF pulse furthermore shifts the precession of the nuclei into same phase angle, which creates resonance, and a net magnetization pointing 90° to the longitudinal magnetization. This magnetization is called the transverse magnetization. The coherent nuclei produce a radio signal, or free induction decay signal (FID signal), that can be detected by a radio antenna. After the RF pulse is removed, the nuclei will relax into baseline state. Firstly, the spins of the nuclei will repel each other, as they are positively charged, and thus shift phase. The net magnetization will return to zero. This relaxation is called T_2 or "spin-spin" relaxation, as the energy exchange between the nucleus spins is causing the relaxation. An illustration of the T_2 relaxation can be seen in figure 1.2.

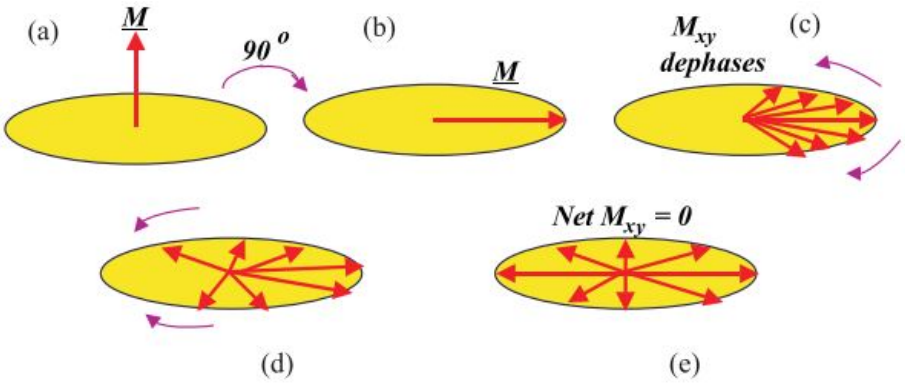


Figure 1.2: An illustration of the T_2 relaxation. Before the RF pulse is emitted the majority of the nuclei are precessing out of phase in the low energy state and the net magnetization is pointing in the direction of the longitudinal magnetization, (a). The nuclei absorb the RF pulse, the magnetization shifts 90° and the nuclei precess in phase, which leaves a high transverse magnetization and FID-signal, (b). Due to interaction between the nuclei spins, the nuclei gradually dephase, until the net magnetization is 0, which is referred to as T_2 relaxation. [1]

A second relaxation appears as the high energy nuclei returns to the low energy state. The energy that was previously absorbed by the nuclei is dissipated in to the surrounding lattice in the form of heat. During this relaxation the longitudinal magnetization is regrown. This relaxation is called T_1 or "spin-lattice" relaxation, as the spins transfer energy to the surrounding lattice. [1] The hydrogen nuclei are located in different local environments in the body. Some are for instance associated with free-floating water molecules, while others are associated with structural and storage molecules such as proteins and lipids, and thus more fixed in position. The nuclei have different T_1 and T_2 relaxation characteristics, depending on the local environment or tissue they are associated with. This can be accentuated and measured in NMR. [1]
The chosen pulse sequence is key to how the tissue will be portrayed in an image, and is described by the T_{echo}, time before the FID signal is measured, and T_{rep}, time before a new RF pulse is applied. In a case of nuclei associated with lipids and water molecules, the nuclei in lipids are fixed and will have a fast T_1 relaxation after exposure to a RF pulse. Meanwhile the nuclei in the water molecules will maintain being in a synchronized phase. At T_{echo}, the nuclei associated with the lipids will have a low amplitude FID signal, as the transverse magnetization is weak, and the nuclei associated with the water molecules will have a high amplitude FID signal, as the transverse magnetization is strong. The water molecules will be assigned a white color on a greyscale image and the lipids as dark grey/black. In this case there is a long T_{echo} and a long T_{rep}, and is referred to as T_2-weighted MRI A commonly used T_2 pulse sequence is the spin echo (SE). Due to magnetic field inhomogenities the nuclei dephase more rapidly. Applying a second RF pulse of 180°after the nuclei have dephased, the nuclei will rephase. Again the constrast in the image is expressed as a result of the relaxation time of the nuclei, which is depends on the surrounding environment. . [1]
In case of T_1-weighted MRI, the T_{echo} and T_{rep} are short. As in T_2-weighted MRI a RF pulse is applied and the nuclei associated with lipids will quickly return to baseline state and the water molecule nuclei will remain a strong transverse magnetization. At this time point a second RF pulse will be induced, referring to the short T_{rep}. Now the lipid nuclei will return to a strong

transverse magnetization state and excite a high FID signal. More low energy state nuclei of the water molecules will absorb the RF pulse and shift to a high energy state, leaving a majority of nuclei in a high energy state. The water molecule nuclei now has a weak transverse magnetization and 180 degrees longitudinal magnetization, thus producing a low-amplitude FID signal. A short T_{echo} after the second RF pulse then shows lipids as white and the water molecules as dark grey/black in a greyscale image. [1]

T_2 relaxation was defined as dephasing due to energy exchange between nuclei. In practice the T_2 relaxation time happens much faster than would be predicted by these natural atomic interactions. Another factor in the dephasing of nuclei is inhomogeneities in the magnetic field. This observed T_2 relaxation is referred to as T_2^* relaxation. In a T_2^* -weighted MRI, a gradient echo (GRE) is used instead of a SE. In a GRE pulse sequence only one RF pulse is emitted with a low flip angle, and the echo time is therefore usually shorter. A gradient is applied after the initiating RF pulse, which enhances the dephasing. When the net magnetization is zero a rephasing gradient with opposite polarity of the dephasing gradient is turned on, which reverses the phase shift. A FID-signal is produced as a GRE. The gradient only reverses the phase shifts that have been affected by the gradient itself, and not those affected by magnetic field inhomogeneities, Contrast in tissues is therefore not decided through natural T_2 relaxation by T_2^*. Thus, T_2^* -weighted MRI only work well in scanners that do not lack magnetic field homogeneity. Due to the fast acquisition time T_2^* -weighted MRI is widely used in functional MRI to image brain activity. [3]

1.2 Functional MRI

Different MRI techniques exist to accentuate particular tissue types, physiological phenomena, etc. Among these the Blood Oxygen Level Dependent (BOLD) will be described in this section. The reason for solely focusing on BOLD in this project, is due to it being the most widely used and accepted method for functional Magnetic Resonance Imaging (fMRI). BOLD is favored in mapping brain activity, because it offers a high contrast to noise ratio and it is relatively simple to implement. [5]

fMRI measures the metabolic changes associated with different neurological tasks in multiple brain areas. fMRI offers advantages which predominantly are high temporal and spatial resolution, low cost, and most importantly being non-invasive, which has made it a exceedingly popular method for imaging brain activity. The versatility of fMRI has made it a very important tool by being a biomarker for diseases and test of of pharmaceuticals. [6]

The physiologic background facilitating the possibility of measuring brain activity through the BOLD contrast is found in the local metabolism during neural activation. Multiple steps in forming and transmitting a neurological signal requires adenosine triphosphate (ATP) consumption e.g. reception and reformation of an action potential. When activating a brain area in e.g. finger tapping, the ATP starts to be processed, leading to a decrease in oxygen concentration and increase in waste. Thereby the metabolic need for oxygen increases. As the movement is planned and executed, factors, which are present in the local tissue of the corresponding

brain area, activate a vasodilation, increasing the blood flow to that area to reestablish the local homeostasis. During this regulation a special and not fully understood phenomenon occurs as more oxygenated blood than needed to compensate for the offset is delivered. Thereby flooding the local region with oxygenated blood. The overall increase in neural activity in that specific area following the need for metabolic regulation thereby permits blood oxygen level dependent (BOLD) contrast to studied. An example illustrating the measurable hemodynamic response can se found in figure 1.3. [6, 7]

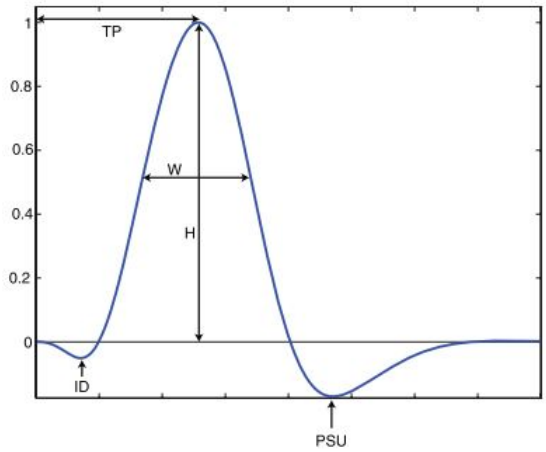


Figure 1.3: A depiction of a single hemodynamic response curve. ID is the initial dip as less oxygen will be present as the metabolic demand increases, TP is time from stimulus until peak, W and H is the width and height of the response and PSU is a post stimulus undershoot. [7]

Figure 1.3 should be considered the perfect noiseless hemodynamic response curve, to a brief stimuli, though the reality is that the response is noise and delayed in time. The peak height of the curve is commonly the most interesting feature of the response, as it portrays the amount of neural activity. The time to peak will occur 4-6 seconds after stimulus onset. The duration of a response is around 20 seconds. There will further be a noticeable initial dip of 1-2 seconds duration and a 20 second poststimulus undershoot. [7]
As established above the BOLD signal is effected by the neural activity producing changes in the local blood flow, blood volume and blood oxygenation. The crucial part to why MRI can detect this natural contrast is that oxygenated hemoglobin (HbO_2) is diamagnetic, and deoxygenated hemoglobin (Hb) has four unpaired electrons thereby making it highly paramagnetic. Thus, the more oxygenated blood in an area, the larger the contrast compared to other brain regions would be visible in an acquired image. [6, 7, 8] Figure 1.4 illustrates how the contrast i dependent on the amount of oxygenated hemoglobin. [6]

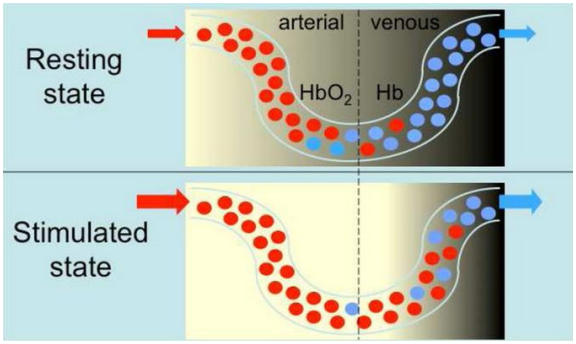


Figure 1.4: Illustration of how the difference in oxygen concentration in the hemoglobin change the magnetic properties, resulting in a higher measurable contrast [6].

This change in local magnetic properties increases the magnetic inhomogeneity, which can be recorded by using a $T2^*$ sequence, which was presented in section 1.1 [5].

1.3 MR image reconstruction

Different pulse sequences and certain physiological properties that can be exploited with certain pulse sequences, have been laid out in the previous sections. This section aims to describe how the corresponding echo signals are reconstructed as a MR image.

Following the Lamour frequency equation (1.1), the main magnetic field causes all hydrogen nuclei to precess with the same frequency. Without any specification of spatial localization a MRI of a human body would consist of a single number. To prevent this, separate coils in the x, y and z directions are introduced. These coils can be adjusted in position, and thus produce gradient magnetic fields with a varying strength depending on position. According to the Lamour frequency the nuclei will precess with different frequencies when in a magnetic field with varying strength. The gradients can be turned on in combination to create any direction in space. These varying frequencies can be exploited to separate parts of the anatomy and ultimately illustrate a desired area. As mentioned, the nuclei only exchange energy efficiently if the frequency of energy, or RF pulse, matches the precession rate. Thus, by altering the magnetic field along the body in one direction, z-direction for the sake of the example, the nuclei will have slightly different precession rates, and the RF pulse will only efficiently affect a desired slice of the nuclei. The nuclei of that slice now precess at the same rate. To get an image with a spatial resolution, the voxels that make out the image needs to be discriminated between. By turning on the gradient of the x-direction the lines in the y-direction are now encoded with a particular frequency. This gradient functions as a frequency encoding gradient, and is illustrated in figure 1.5. [1]

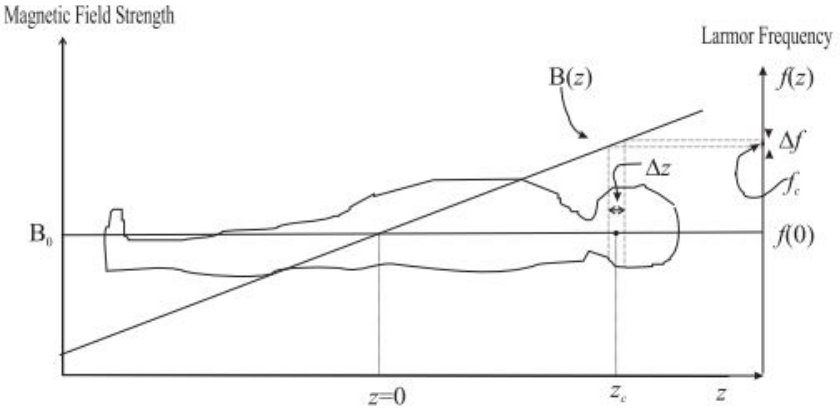


Figure 1.5: The position of the slice is specified through the direction of the frequency encoding gradient ($B(z)$) and though the central frequency of the emitted RF pulse (f_c). The thickness of the slice is dependent on the steepness slope of $B(z)$, and on the bandwith of the emitted RF pulse (δf). [1]

Turning the y-gradient on and quickly off, will de-phase the nuclei while still remaining the same frequency as before. This gradient functions as a phase encoding gradient. When comparing two locations approximately one voxel apart in the x-direction, then based on the amount of gradient strength difference, there will be a certain amount of change in phase between the spins spread across that distance. The farther away from isocenter, where the magnetic field strength is B_0, the higher the change in phase will be. This notion is used to assign the correct spatial location of each voxel, when reconstructing the FID signals into an image. This phase encoding procedure is done in different gradient strengths in iterations to assign unique phases to the nuclei in the both directions. One iteration of a certain strength of the phase encoding gradient followed by a measurement is performed at a time. The only change per iteration is the phase encoding gradient strength. These iterations are then series of measurements acquired at different points in time, where each entry of the slice then represent a certain signal intensity. This time domain measurement is referred to at the raw data. [1]
The next step is to Fourier Transform (FT) the raw data, which will yield frequency information to the acquired signal intensities. This step gives a summation of the signal intensities at the different frequencies produced by the frequency encoding gradient. This is called the k-space, as the k-numbers of a signal describes its relative orientation and frequency. The k-space image contains the contrast in the center and the resolution in the periphery, as there is low or no phase encoding at the center and increasing towards the periphery, giving more brightness in the center and dimmer tones in the periphery. To allocate the voxels in correct spatial localization an inverse 2D discrete FT is performed on the k-space image. This provides the desired image of the anatomy slice. [1] Figure 1.6 shows the acquired signal in k-space and the reconstructed image after the Fourier Transform.

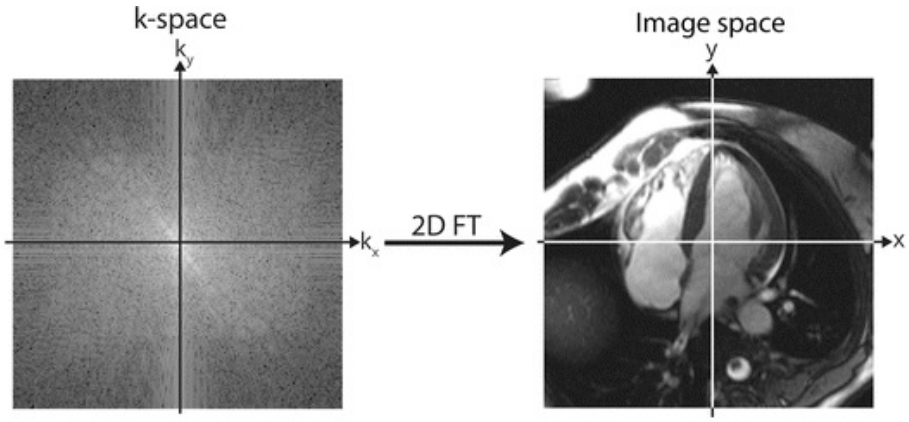


Figure 1.6: A depiction of the acquired signal represented in k-space and the resulting reconstructed image after the inverse 2D Fourier Transform [4].

1.4 Stimuli Design

The following section will describe different standards of designing experiments where the impact of a given stimuli is used to assess the subsequent brain activation. The impact of stimuli on the hemodynamic response and how it is transformed into the hemodynamic response function (HRF) used analysis will be further explained, adding to the knowledge gained in section 1.2.

In order to accomplish a well designed experiment the researcher must consider the multiple types of stimuli delivered, form and duration of these. The researcher must be aware of the timing of events in the scanning session and any responses provoked. Additionally, the researcher should have general knowledge of where in the brain activation is seen and how the hemodynamic response will be presented. [9]

Doing cognitive experiments using fMRI, two main design types are utilized, by either using a block- or event-related design. Event-related design is inducing a series of very short lasting stimuli used to investigate single hemodynamic response. A characteristic of this method is that it permits the possibility of increasing and decreasing the interval between stimuli. Thereby the theoretical likelihood of subject confounds should be reduced as the interval would not become predictable. Event-related stimuli design further allows more temporals characteristics to be inspected, compared to a block design. Characteristics could be hemodynamic response in duration and amplitude. [10]
Block design works by performing a series of less but longer stimuli. Block designs are ideal for experiments involving detection of small differences in BOLD signal across various test conditions where its statistical power is superior. Furthermore artifacts present are more easily detected in the signal time course, because of the signals temporal structure. A block design is easier to design than an event-related, as randomization of intervals of stimuli is not required. The design instead focuses on the total number of stimuli used, block length, inter stimulus interval, block length and TR. An illustration of both design types can be found in figure 1.7.

[10]

To enable use of the hemodynamic response in fMRI analysis, it needs to be transformed in order to represent the ideal physiologic response to the stimulus. Thus, portraying the biological delay from stimuli to response. Therefore stimulus in the design is combined with the hemodynamic response function through convolution. Thus making a function that models how the BOLD signal would be represented if the voxel activity increased in a given area each time stimuli is induced. [9]

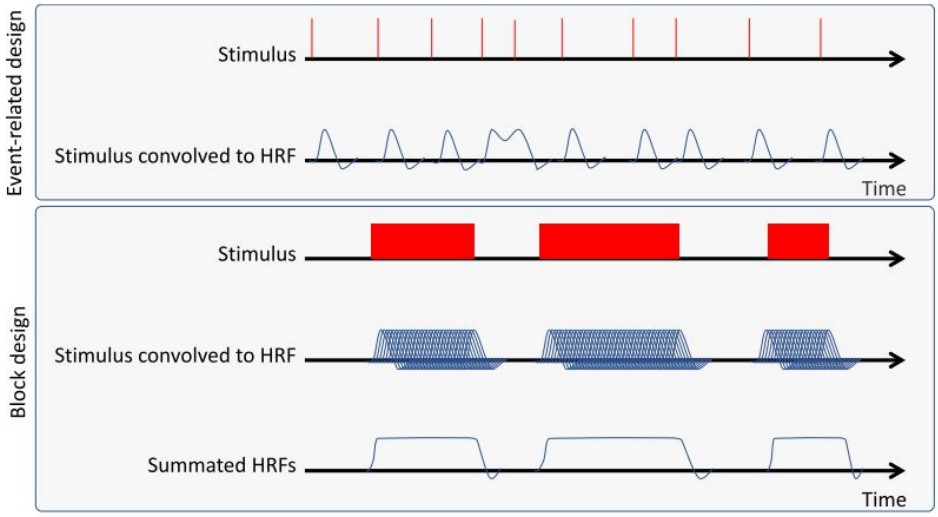


Figure 1.7: The top image show the event-related stimuli design and the stimulus convolved with the hemodynamic response function. The lower image depicts a block design, the stimulus convolved with hemodynamic response function, and the summated HRF response. [9]

1.5 General MRI Preprocessing

The following section will present some preprocessing methods, which generally are applied to remove noise and artifacts. For each method the artifacts/noise problem will be described followed by how the method solves this problem. The section will thereby present some of the common issues to be aware of in fMRI acquisition and what methods to implement in limiting artifact/noise influence.

When MRI data is acquired there are several steps, which need to be taken, before the multi-dimensional images are ready to undergo statistical analysis. These steps involves correction methods, which are often referred to as preprocessing. [9] There are multiple steps in preprocessing fMR images depending on the apparent application and outcome intended. However, there is a standard set of methods that is usually used across all applications. An example of some of the general processing steps for MR imaging can be seen in figure 1.8. [7]

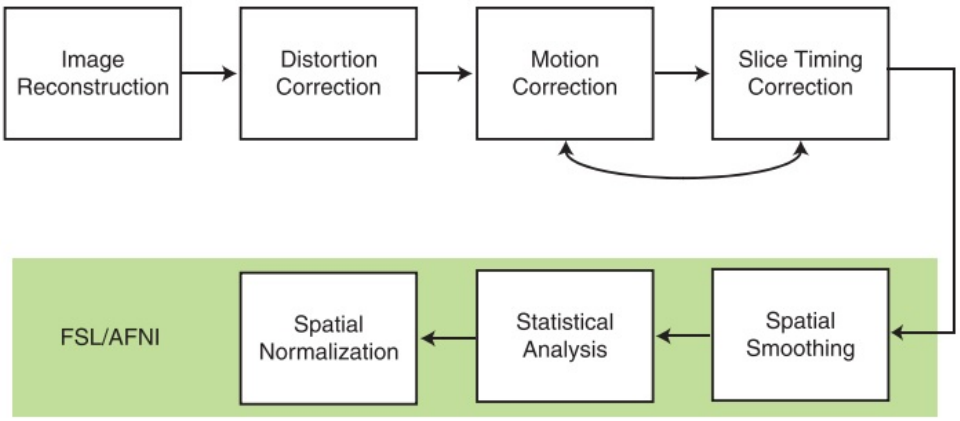


Figure 1.8: The general pipeline for MRI processing done in either FSL or AFNI, showing the different processing steps considered before final statistical analysis. Modified from [7].

1.5.1 Quality control

Conducting a continuous quality control is highly recommended after each performed correction step. Various scanner artifacts can occur while acquiring an MRI series. Before performing any common correction steps, one should consider to look for spike or ghosting artifacts. Spike artifacts are seen as a regular pattern of change in brightness across the entire image. This problem can occur due to instability inside the scanner deriving from e.g. electrical discharges. The artifact called ghosting occur mainly due to two reasons. One being an offset in phase between different lines in K-space and the other due to periodic motion as in heartbeat and respiration. Ghosting can be seen as light copies of the object appearing to either side of the object. Both types of artifacts can corrupt the information contained in the images. However, artifacts of this kind rarely present themselves in newer scanners, nevertheless it is still recommended to perform a quality control of the raw input from the scan as well as the result of the correction steps. [7]

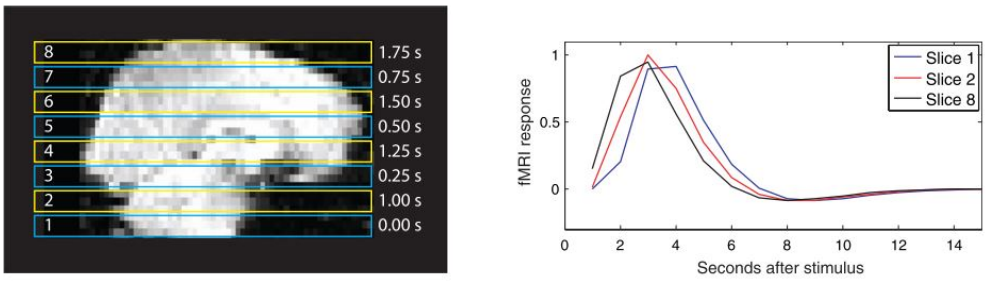
1.5.2 Distortion correction

Some fMRI acquisition methods suffers from artifacts at regions where air and tissue meet. The ear canal and sinuses are areas especially vulnerable. Inhomogeneity in the magnetic field in these areas can cause two types of artifacts: dropout and geometric distortion. A dropout will result in a reduced signal intensity in regions close to the air to tissue transition. When a dropout during an acquisition occurs, the lost signal cannot be restored and the damage is permanent. Therefore it is wise to consider the appropriate acquisition method taking the area of interest into addition. Air to tissue passages can also be subject to spatial distortion due to inhomogeneity created in the magnetic field. This will lead to structures not being located correctly in the captured image. This distortion makes is difficult to align two different scans, as done when aligning fMRI images with structural images. The spatial distortion can partially be corrected by employing field maps. In order to do a field map, the pulse sequence from the scan needs

to be known. The process involves acquiring images at two different echo times. This results in images with two different phases, which can be used to compute the field inhomogeneity. Thereby it becomes possible to calculate the relative distance each voxel has shifted. This makes up a map for the distance shift for each voxel, and by inverting the map the original image can be restored. [7]

1.5.3 Slice timing correction

Acquiring fMRI scans is nearly always done in two-dimensions, where the slices are taken one by one. This can either be in an ascending, descending or interleaved order. Interleaved order [1] is sequentially skipping every either odd or even slice and then afterwards acquiring the skipped slices. Regardless of which order the slices are acquired, a difference in effect in each slice to the same hemodynamic response will be present due to the time difference in the slices. The method and result of interleaved MRI acquisition order can be seen figure 1.9. The difference in time between slices can range up to a couple of seconds depending on the acquisition protocol.



(a) (b)

Figure 1.9: Figure (a) illustrates an example of an MRI acquisition using interleaved order, where initially every odd slice is acquired followed by every even. The information about the hemodynamic response in each slice, and thereby also the difference at each time-point is shown in figure b. Figure modified from [7].

The difference in slice timing constitutes a problem when analyzing the data. The data is formed into statistical model, but since this model assumes that all slices are acquired at the same time point, the actual signal and the statistical model creates a mismatch. To counter this problem slice timing correction has been introduced. The common approach of this method is to choose a reference slice. Usually the slice acquired at T/2, where T is the total scan time, is used to interpolate the others. Linear interpolation can be used for simplicity, but most often sinc interpolation is used as it imposes less smoothing to the signal. [7]

1.5.4 Motion correction

Correcting for motion artifacts when doing fMRI is inevitable, since even the best subjects will not be able to hold still. Even subtle movements as swallowing will be visible in the acquired

[1] FiXme Note: think it is done to avoid crosstalk

image. [7]

Multiple internal and external factors can cause a subject to move. Internal factors are non-avoidable physiologic motion. The heartbeat causes a pulsating movement, which makes the brain move. Additionally, motion created during respiration can cause small changes in the magnetic field around the head. External factors like imposed stimulus might also cause the subject to make sudden movements. Often when doing fMRI the brain activation is measured while the subject is subjected to some kind of stimulus. The stimulus would make the patient move, while some brain regions might also show activation associated with stimulus. Therefore it is easy to mistake brain activation with stimulus correlated movement when analyzing the data, resulting in a weaker or even false statistical analysis. [7]

Motion during image acquisition can result in two primary artifact effects: bulk motion and spin history. Bulk motion refers to the movement of the head as a whole and requires standard correction methods, e.g. the images throughout the series to be realigned to a reference image. The effect of bulk motion can be visual in the entire image of the brain, but the effect will be most predominant at the edges of the brain. Here the artifact will be noticeable as either a drop or increase in intensity as a voxel would switch from containing brain tissue to suddenly not. Spin history is head movement interfering with the MRI signal itself. The interference occur during acquisition when a voxel of excited protons is moved in to a neighboring slice. The scanner will thereby receive a different signal than expected which do not correctly represent the actual local properties. This results in an image where the intensities change in a striped pattern. The standard motion correction methods cannot cope with this type of artifact, but Independent Component Analysis (ICA) might give opportunities to correct for this artifact. [7]

As mentioned earlier motion correction is to realign the series of images to a reference image trying to minimize cost in an introduced cost function. The reference image is usually the one taken midway into the series, justified by it being the closets to the average as well as the scanner at that time would have achieved maximum stability, as the magnetization would have reached steady state. The images are thereafter realigned utilizing an image registration method as it register the brain in each image. The general methods for motion correction treats the brain as a rigid objects, thus only performing rigid body transformations. Subsequently, the brain can either translate or rotate along the three axis, but the shape of the head cannot change. This method is therefore only applicable for bulk motion. [7]

1.5.5 Spatial smoothing

In some cases, introducing spatial smoothing in the preprocessing pipeline proves to be beneficial. Spatial smoothing allows the possibility of gaining a higher signal to noise ratio within the image, though with the consequence of a decrease in spatial resolution as the image gets blurred and smaller areas of activation gets smeared together. The operation can be justified by the closely neighboring voxel being correlated in effect to the hemodynamic response. Spatial smoothing removes the higher-frequency information. This might wash out some of the less significant features in the image, but this is favorable if the signal is increased for the more

significant features. Especially when acquiring small voxels spatial smoothing is favorable, as it reduces the overall noise. Smoothing can also be applied to lessen the anatomical variability in images when doing studies with multiple subjects. [7]

Smoothing is done by applying a kernel, also called filter, to the image. The three dimensional image is convoluted with a three dimensional filter. The most commonly used is a Gaussian filter, where the extent of smoothing is controlled by the width of the distribution. The filter works such that for each voxel a new value is calculated based on a weighted average of the neighboring pixels, where the ones closest contribute the most and those further away contribute the least. In statistical terms this can be put as the width of the distribution is described by the full width at half maximum (FWHM). Using standard deviation this would be:

$$FWHM = 2\sigma\sqrt{2ln(2)} \tag{1.2}$$

, where an increase in FWHM, would result in a greater smoothing. The amount of smoothing needed to be implemented highly depends on the application and purpose. When smoothing fMRI signals for noise, the width of the filter distribution should not be bigger than the area of activation signals of interest. The effect of smoothing is shown on figure 1.10, where it is seen that as width increases smaller activation areas get removed and bigger areas of activation gets smeared together. Using smoothing to reduce the effect of anatomical variability, the optimal distribution width depends on the amount of variability in the subject population. [7]

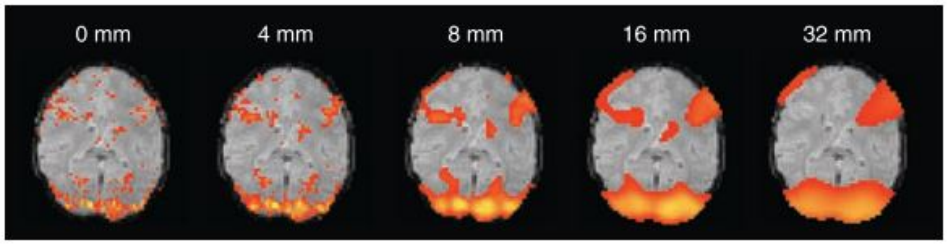


Figure 1.10: Illustration showing the impact of using different distribution width on the activated areas. An increase in width results in greater areas of activation smearing together and the removal of smaller. [7]

1.5.6 Temporal filtering

A characteristic noise which represents itself during fMRI data is the presence of a low-frequency drift. The drift is characterized as a slow increasing trend in the BOLD magnitude, when assessing the signal in the time domain. Doing a Fourier Transform, to analyze the signal in the power spectrum, would reveal low frequency contributers influencing the output. The reason for this type of noise contamination has been heavily investigated, and conclusions state that the noise originates from MRI scanner instability.[2] As this low-frequency drift will always be present, it is very crucial to consider the interval of which tasks or stimuli are performed to avoid the output being present in the noise range of 0 to 0.015 Hz. Therefore stimuli or tasks

[2] FiXme Note: heating of coils and such i believe

should be performed within intervals of 70 s or less. [7]
A two step approach is used to remove the low-frequency contribute from the signal. Firstly a high-pass filter is used to remove the drift trend in the data. Introducing the high-pass filter impose a correlation of the data making the time series correlated over time. This would violate some of the assumptions made in the General Linear Model (GLM), often used for statistical analysis, because the data is not temporally autocorrelated and the variance to be constant over observations, as presented in section 1.8. Not attending to this problem might cause an elevation in false positive rate. Thus, the second step is to estimate the autocorrelation and undo the correlation structure of the data. This is typically done by pre-whitening the data. There are multiple ways to implement a high-pass filter ADD SOMETHING ABOUT LOWESS METHOD [7]

1.6 Independent Component Analysis

Independent component analysis has proven to be a very useful tool in separating noise from wanted signal in many application. The methods has been adapted to work in BOLD-signals, making it a very desirable tool in fMRI. [12] The following section will seek to elucidate the general basic theoretical background behind the analysis exposing it possible use in this project.

The basic idea about Independent Component Analysis (ICA) is to recover m signal sources, which are mixed in n observed signals. The observed signal is given by [13]:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1.3}$$

where $\mathbf{x}$ is a observed signal vector containing the mixed signal elements $x_1, x_2, ..., x_n$, $\mathbf{s}$ is the source signal vector with the elements $s_1, s_2, ..., s_m$ and $\mathbf{A}$ is a mixing matrix with the dimension $n \times m$. Note that the dimensions of the mixing matrix can be equal to each other, meaning that it is required to have at least the number of observed signals as the source signals. The observed signal is assumed to be a linear mix of independent source signals. When performing ICA the goal is to find an inverted mixing matrix A^{-1} that recovers the source signal [13]:

$$\mathbf{s} = \mathbf{A}^{-1}\mathbf{x} \tag{1.4}$$

This can easily be achieved if the mixing matrix is known. However, this is rarely the case, as both the mixing matrix and source signals are unknown. There is no reliable way, to fully determine $\mathbf{s}$, thus a set of assumptions are required, which the ICA method is based on [13]:

- The independent components (IC)'s are assumed to be statistically independent.
- The IC's must be non-Gaussian distributions.

Regarding the assumption of independence, a set of variables $y_1, y_2, ..., y_n$ is not allowed to share mutual information so that $i \neq j$. This can be expressed as the joint probability of the variables is equal the product of each marginal probability of the variables [13]:

$$p(y_1, y_2, \ldots, y_n) = p(y_1) \cdot p(y_2) \cdot ... \cdot p(y_n) \tag{1.5}$$

Satisfying this condition assures independency of the variables. A graphical example of independence in two dimensions is shown in figure 1.11 [13, 14].

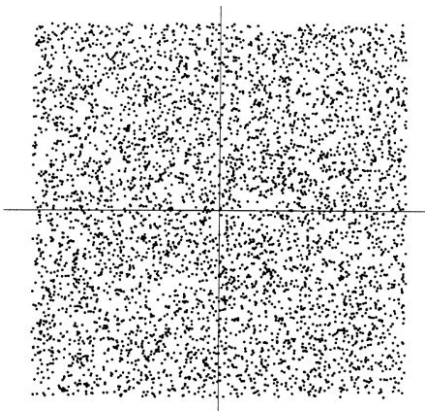


Figure 1.11: Figure illustrating the joint distribution of two independent components having uniform distribution. [14].

It is clear to see that information about a variable on the horizontal axis does not give any information about variables at vertical axis and vice versa. Note that uncorrelatedness does not equal independency. However, whitening of the observed signals and thus ensuring uncorrelatedness is very helpful in solving the ICA problem. Whitening of the data is archived by performing a linear transformation that transforms the components so that the covariance matrix equals the identity matrix, thus having unit-variance. This can be expressed through performing a linear transformation of $\mathbf{x}$ into a random vector $\mathbf{z}$:

$$\mathbf{z} = \mathbf{V}\mathbf{x} = \mathbf{V}\mathbf{A}\mathbf{s} \tag{1.6}$$

A graphical example of whitened data in two dimensions is shown in figure 1.12 [13, 14].

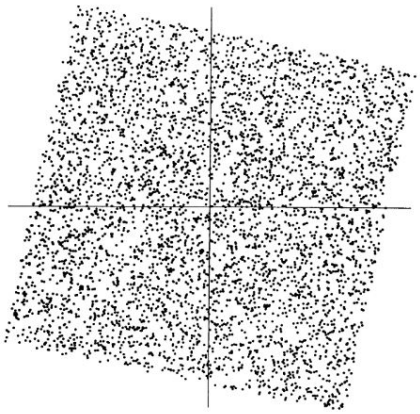


Figure 1.12: Figure illustrating the impact of whitening on the joint distribution of two components [14].

The squared distribution is clearly a rotated form of the independent data. What is left to ensure independence and solving the ICA problem is to estimate an angle that gives the correct rotation. Concerning the second assumption on non-Gaussianity, the joint distribution of uncorrelated Gaussian distributions are not necessarily independent, and the distribution will be symmetrical and no information on the direction of the distribution can be derived. Thus, no transformation that allows independency can be performed, and the mixing matrix $\mathbf{A}$ can not be estimated from the mixtures. This can be shown in a two dimensional example with two sources s_1 and s_2 and the mixing matrix $\mathbf{A}$, the joint probability density function (pdf) of Gaussian distributions is calculated as [13]:

$$\begin{aligned} p(s_1, s_2) &= \frac{1}{2\pi} exp(-\frac{s_1^2 + s_2^2}{2}) \\ &= \frac{1}{2\pi} exp(-\frac{\| \mathbf{s} \|^2}{2}) \end{aligned} \tag{1.7}$$

For an orthogonal mixing matrix $\mathbf{A}$ the inverse can be written as $\mathbf{A^{-1}} = \mathbf{A^T}$, and then $\mathbf{s} = \mathbf{A^T x}$. The joint pdf can then be rewritten as:

$$p(x_1, x_2) = \frac{1}{2\pi} exp(-\frac{\| \mathbf{A^T x} \|^2}{2}) \mid Det(\mathbf{A^T}) \mid \tag{1.8}$$

Because $\| \mathbf{A^T x} \|^2 = \| \mathbf{x} \|^2$ and $\mid Det(\mathbf{A^T}) \mid = 1$, hence the orthogonality of $\mathbf{A}$, the joint pdf is:

$$p(x_1, x_2) = \frac{1}{2\pi} exp(-\frac{\| \mathbf{s} \|^2}{2}) \tag{1.9}$$

The joint pdf is not changed when choosing an orthogonal mixing matrix as well as the property of independence. No further information about the mixing matrix can thus be revealed when the source signals are from a Gaussian distribution. [13]

1.6.1 ICA approaches

As the mixing matrix and source signal most often are unknown, the IC's must be approximated in an iterative process. There are two main ICA iteration approaches: minimizing mutual information and maximizing non-Gaussianity. The former approach seeks to minimize mutual information by maximizing independence between components. In practice this is done by minimizing the difference between the joint density distribution and the product of the marginal density functions, the left-hand side and right-hand side of equation (1.6). This can for instance be done through a Kullback-Leibler divergence between the $p(y_1, y_2)$ and $p(y_1) * p(y_2)$ in a two dimensional case. The second approach seeks to maximize non-Gaussianity of the components. The central limit theorem states that if sources are mixed, the mix tend to get more Gaussian than the individual sources. The strategy here is to find the directions in the data that is as far away from Gaussian as possible through a linear transformation. That direction will most likely be independent components. To find these directions different ICA approaches use fourth order moments and negentropy of the data. [13]

1.7 Principal Component Analysis

The use of principal component analysis (PCA) in denoising algorithms has been exploited in various studies, showing that the properties of the PCA might be useful [11]. In this section the theory and common use of PCA is presented, including a small part on its use in images.

Principal Component Analysis (PCA) is a well renowned and widely used analysis tool, capable of finding the most defining variables in a dataset. This facilitates finding the components that are the most saying for the dataset. This introduces the possibility of dimensionality reduction by lowering the amount of redundant information. The PCA is used to transform a set possibly correlated variables into a set of uncorrelated components, called principle components. Each principal component (PC) is orthogonal on the former and are uncorrelated and have zero covariance. They each define the largest variance in an axis, such that PC 1 describes the direction of the maximum variance of the dataset. Each following PC describes the next highest variance of the dataset, with the constraint that it is orthogonal and has zero covariance with any of the former PCs. PCA is the orthogonal projection of data onto a lower dimension linear space. A PC is found by minimizing the variance by projecting the feature values (blue dots) onto the line (now red dots) describing the highest variance in the data set (black line) as seen on figure 1.13. The PC is found by minimizing the mean square distance between the data points. [15]

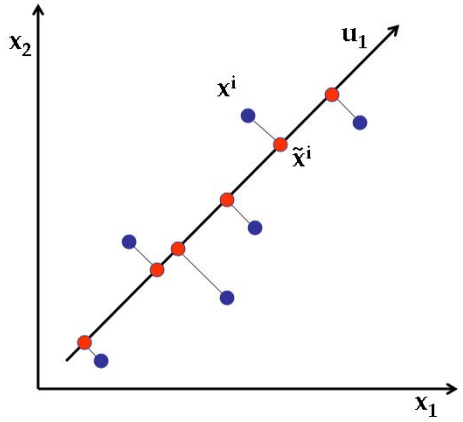


Figure 1.13: Two-dimensional example of projection of data variables (blue dots) onto PC axes (black). (u_1) indicates the direction of the eigenvector. [16]

The algebraic method of calculating the PCs can be done by using Singular Value Decomposition (SVD). The first step is to compute the squared cross product matrix of variances and covariances among every pair of the variables in the data set, where the diagonals are the variances and the off-diagonals are the covariances, as done in the following equation:

$$S = X'X \tag{1.10}$$

Where S is the cross product and X is the dataset matrix. When finding the PCs it includes an eigen-analysis of S. The eigenvalues of are solutions to the following equation:

$$|S - \lambda I| = 0 \tag{1.11}$$

Where λ is the variances of each PC and I is the identity matrix. After solving for λ the eigenvectors can be solved through the following equation:

$$det|S - \lambda I|b_i = 0 \tag{1.12}$$

Where b_i is used to calculate the eigenvectors as in:

$$u_i = \frac{b_i}{\sqrt{b_i' b_i}} \tag{1.13}$$

Where u_i is the i number of eigenvectors that contain a contribution to the principal components. The SVD orders the eigenvalues by size $\lambda_1 > \lambda_2 \ldots > \lambda_i$. The scores for each PC is equal to the corresponding eigenvalue for that exact axis. The eigenvalues describe how much of the variance is accounted for by the associated PC. Summation of all eigenvalues accounts for the total variance of the data set; this is called the trace. To find how much the each PC accounts for, the eigenvalue of that PC is divided by the total variance: $\%\ of\ total\ variance = \frac{\lambda_i}{Trace}$. This can be used for deciding how many components are significant and by how much the dataset can be reduced. [15]

1.7.1 PCA in images

The PCA can also be implemented on images, where the dimensionality reduction principle is mostly used for image compression. Here images can be reconstructed using only very few principle components without much information loss. Similarly PCA can be used to de-noise images, as noise would be presented in some of the less saying components, and by removing these, noise would be removed from the image.
The PCA can be run on the entire image, but methods introducing local PCA in smaller window of the image, has also been introduced for noise removal. [17]

1.8 General Linear Model

In order to determine if a task or stimuli has made a significant contribution to the measured BOLD signal, each voxel in the scan needs to be evaluated. The following section will therefore seek to explain a method of how to determine if any significant change is found. In this project the general linear model (GLM) will be explained and applied as it is the most used and recognized tool for fMRI analysis during the past 20 years [18].

The GLM is a well considered and used analysis tool constituting a simple way of doing standard statistical analysis on fMRI. The overall goal of the GLM model is to determine how well the time course of the scan corresponds to the known used experimental interference. In an fMRI case that means, how well the BOLD signal over time fits the time course of the predicted signal given by the imposed stimuli, causing brain activity. This statistical test is carried out for each voxel independently of its neighbor across the scan resulting in thousands of statistical test in one scan. [9, 19] To explain the implementation of a GLM we can use the example presented

by *Monti et al.* An acquired BOLD signal will throughout a time series, of n images, have a varying output signal. The signal in a voxel Y, at any time point, can be seen as a summation of predictor variables (stimuli regressors) X, additional nuisance regressors modeling noise X, a scaling parameter for each regressor β, and an error term ε. This can be presented in the GLM's basic from as [19]:

$$Y = X\beta + \varepsilon \tag{1.14}$$

Expanding this formulation into a complete scan, including multiple regressors, can be presented on a matrix form. A depiction of a GLM design matrix incorporating the factors making up the BOLD signal can be seen in figure 1.14.

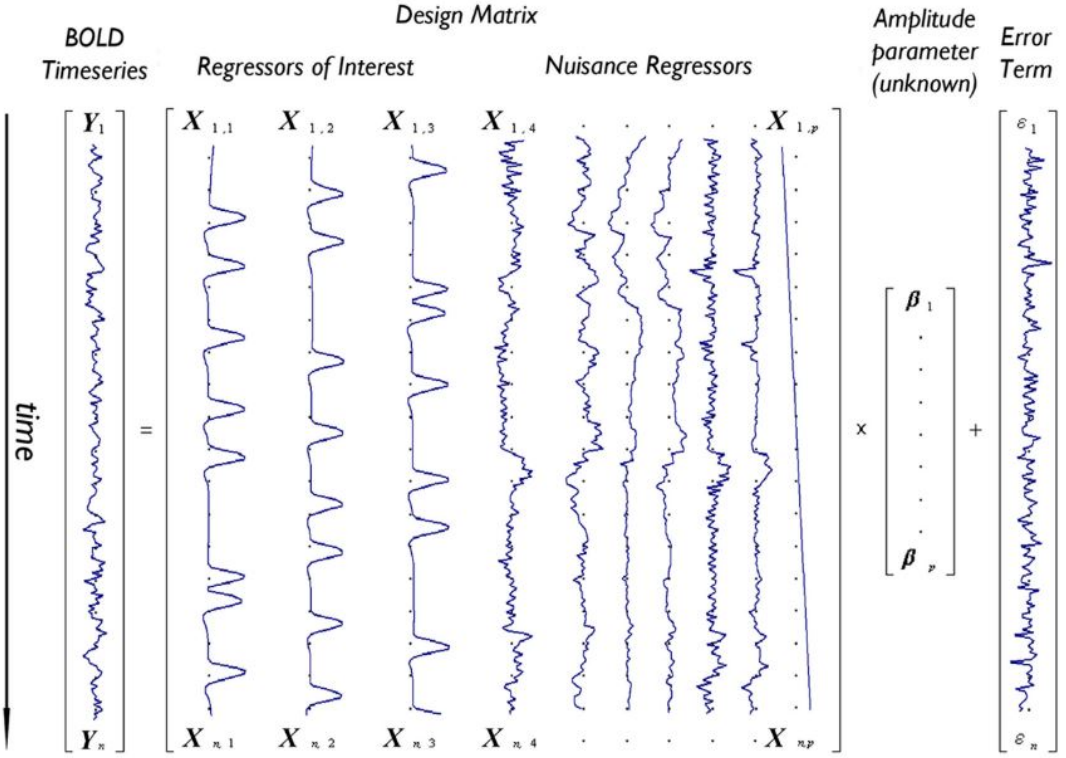


Figure 1.14: Illustration of a design matrix for a given GLM model, describing the BOLD time series Y_n, the regressors of interest $X_{1,1}, X_{1,2}, X_{1,3}$, the additional nuisance regressors $X_{4..p}$, the amplitude parameter $\beta_{1..p}$ and the error terms $\varepsilon_{1..n}$. [19]

The predictor variables are found by modeling what is known or predicted as output. The predictors of interest would come from hemodynamic response curve convoluted with stimuli design as presented in section 1.4. GLM models will often introduce nuisance predictors as well used to model variables such low frequency drift and motion to make a more robust model. These are non interest regressors, as they do not resemble the wanted signal, being the stimuli induced. The amplitude parameter is the unknown weight scaling the magnitude between each predictor value and the data. They describe the strength of the relationship between that regressor and the voxel's BOLD signal course of activation. The error term contains the value for

each observation, which can not be explained by the wighted sum of the amplitude parameter. [9, 19]

The goal is to estimate the value of the scaling parameter β for all regressors and afterwards determine if any regressor significantly account for the variance found in the BOLD signal. A regressor associated with the measured BOLD signal should hypothetically show a greater value for the voxels in the brain area corresponding to a given task. E.g. finger tapping would show greater β values in the motor cortex. A method for estimating β values for the regressors is the ordinary least squares (OLS). The OLS method estimates β by minimizing the sum of squared residuals [19]:

$$\sum_{i=1}^{n}(Y_i - X_i \times \hat{\beta})^2 \tag{1.15}$$

, where Y is the observed signal, X is the predicted signal which is scaled by β. β and the variance of β can be estimated by:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{1.16}$$

$$var(\hat{\beta}) = \sigma^2 (X^T X)^{-1} \tag{1.17}$$

The application of this method rest on the following assumptions being fulfilled: Error terms are independently and Gaussian distributed with zero mean, the regressors in the matrix are independent of error and non stochastic and known and no regressor is a linear transformation of another regressor. [19]

2 | Methods

2.1 FSL FIX

Top saying something about why this methods needs to be tried. The automated de-noising system FSL FIX (FMRIB's ICA-based X-noisefier), made by the Oxford university center for Functional MRI of the brain (FMRIB), provides newly proposed data driven method for de-noising fMRI scans. The method applies the use of Independent Component Analysis (ICA) and numerous classifiers, which will be explained throughout this section. [12]

2.1.1 Calculation of Independent Components

For a single session ICA the MELODIC FSL program utilizes the FastICA algorithm to obtain independent components. This section seeks to lay out how the FastICA algorithm works and which mathematical techniques that are exploited.
As stated in section 1.6, from the central limit theorem we get that observed mixed signals tend to have a more Gaussian distribution than the individual source signals, since the observed signal is a summation of the source signals. As a result of this, an approach is to find a linear transformation that leaves the source signals as non-Gaussian as possible. This principle is what the FastICA algorithm is build up around.
Firstly the vector $\mathbf{b}$ is introduced, which is a row vector in the mixing matrix $\mathbf{A}$, and is used in a linear combination:

$$y = \mathbf{b}^\mathbf{T}\mathbf{x} \tag{2.1}$$

By substituting $\mathbf{x} = \mathbf{As}$ in the previous equation:

$$y = \mathbf{b}^\mathbf{T}\mathbf{As} = \mathbf{q}^\mathbf{T}\mathbf{s} \tag{2.2}$$

Where $\mathbf{q}^\mathbf{T} = \mathbf{b}^\mathbf{T}\mathbf{A}$. From this it can be deduced that y is an IC, when only one of the entries of $\mathbf{q}$ is non-zero and the rest is zero. This means that there is no addition of any random processes and the component will be as non-Gaussian as possible. An IC can then be obtained by calculating a value of $\mathbf{b}$ that maximizes the non-Gaussianity of the distribution of $\mathbf{b}^\mathbf{T}\mathbf{x}$ as $\mathbf{b}^\mathbf{T}\mathbf{x} = \mathbf{q}^\mathbf{T}\mathbf{s}$. This gives an optimization problem with convergence at local maxima. As there exist a local maximum of the non-Gaussianity for both s and s_i, the optimization landscape in a n-dimensional signal gives a total of $2n$ local maxima. To be able to optimize according to non-Gaussianity, a quantitative measure of such is needed. This can be provided by the fourth order cumulant, kurtosis, which is zero when y is Gaussian distributed and non-zero when y is non-Gaussian distributed (negative when sub-Gaussian and positive when super-Gaussian).

Given this property and the central limit theorem, the linear combination $\mathbf{b^T x}$ that yields an IC can be found at the local maxima of the absolute value of the kurtosis of y. The kurtosis of y is given by:

$$kurt(y) = Ey^4 - 3Ey^{2^2} \tag{2.3}$$

As the whitening of the data results in unit-variance, *kurt(y)* can be modified to:

$$kurt(y) = Ey^4 - 3 \tag{2.4}$$

To simplify the theoretical analysis furthermore the linear properties of kurtosis for sums of variables can be utilized. For two random variables, x_1 and x_2, it holds that:

$$\begin{aligned} kurt(x_1 + x_2) &= kurt(x_1) + kurt(x_2) \\ kurt(\alpha x_1) &= \alpha^4 kurt(x_1) \end{aligned} \tag{2.5}$$

For the multiplicative scalar textitα the kurtosis is non-linear, and the optimization problem can thus be written as:

$$kurt(y) = \sum_i q_i^4 kurt(s_i) \tag{2.6}$$

Due to the whitening of the data, where y has unit-variance, a constraint is put on the vector $\mathbf{q}$. Since $Ey^2 = \sum_i^n q_i^4$ the vector $\mathbf{q}$ is constrained to the unit-sphere. For the whitened data $\mathbf{z}$ a linear combination $\mathbf{w^T z}$ that maximizes non-Gaussianity is sought for. From the fact that $\mathbf{q} = \mathbf{VA})^T \mathbf{w}$, the following is given:

$$\| \mathbf{q} \|^2 = (\mathbf{w}^T \mathbf{VA})(\mathbf{A}^T \mathbf{V}^T \mathbf{w}) = \| \mathbf{w} \|^2 \tag{2.7}$$

This expresses that constraining the vector $\mathbf{q}$ to lie on the unit sphere equally constraints $\mathbf{w}$ to the unit sphere. The objective is now to find a value of $\mathbf{w}$ that maximizes the absolute value of the kurtosis of $\mathbf{w^T z}$. Whitening further allows the linear combination $\mathbf{w^T z}$ to be understood as projections on the line in a 1-D subspace spanned by $\mathbf{w}$. What is sought for is the direction of $\mathbf{w}$ where the absolute value is maximized, which then makes out an IC.
To solve this optimization problem a commonly used technique is the gradient algorithm to reach convergence at local maxima, thus calculating the direction where the absolute value of the kurtosis of $\mathbf{w^T z}$ is growing most strongly. The can be computed as the following:

$$\frac{\delta|kurt(\mathbf{w^T x})|}{\delta\Theta w} = 4\ sign\ (kurt(\mathbf{w^T z})[E\mathbf{z}(\mathbf{w^T z})^3 - 3\mathbf{w} \parallel \mathbf{z} \parallel^2] \tag{2.8}$$

However, there is in practice disadvantages associated with using the gradient algorithm, e.g. slow convergence rate and is dependent on a proper choice of learning rate. The fixed point algorithms is suggested as a faster algorithm alternative. At the stable point of the gradient algorithm, the gradient is pointing the direction of $\mathbf{w}$. It is only in this case that $\mathbf{w}$ will not change direction when adding the gradient. This notion about the gradient of the absolute value of the kurtosis can be used to form a fast fixed point algorithm, or the FastICA algorithm:

1. $\mathbf{w} \leftarrow E\{\mathbf{z}(\mathbf{w^T z})^{\mathbf{3}}\} - 3\mathbf{w}$
2. $\mathbf{w} \leftarrow \mathbf{w}/ \parallel \mathbf{w} \parallel$

The right hand side is firstly computed and assigned to $\mathbf{w}$, which afterwards is normalized to follow the constraint of $\parallel \mathbf{w} \parallel^2 = 1$. By iterating over this convergence is reached at an ultimate $\mathbf{w}$, where $\mathbf{w^T z}$ is an IC. In order to obtain several IC's the process is repeated, with the constraint that $\mathbf{w}$ must be orthogonal to all previously computed $\mathbf{w}$.

Bibliography

[1] A. A. Bharath. *Introductory Medical Imaging*. Vol. 3. 1. 2008, pp. 1–186.

[2] John C Edwards. "Principles of NMR". In: ().

[3] Govind B. Chavhan et al. "Principles, Techniques, and Applications of T2*-based MR Imaging and Its Special Applications". In: *RadioGraphics* 29.5 (2009), pp. 1433–1449.

[4] Mushabbar A. Syed, Subha V. Ramen, and Orlando P. Simonetti. *Basic Principles of Cardiovascular MRI*. 2015, pp. 1–338.

[5] Sang Pil Lee, Afonso C. Silva, and Seong Gi Kim. "Comparison of diffusion-weighted high-resolution CBF and spin-echo BOLD fMRI at 9.4 T". In: *Magnetic Resonance in Medicine* 47.4 (2002), pp. 736–741.

[6] Gary H Glover. "Overview of functional magnetic resonance imaging". In: *Neurosurg Clin N Am* 22.2 (2011), pp. 133–139.

[7] Thomas Poldrack A, Russell; Mumford A, Jeanette; Nichols E. *Handbook of functional MRI data analysis*. 2011.

[8] Nishanth Khanna et al. "Functional neuroimaging: fundamental principles and clinical applications". In: *Neuroradiology Journal* 28.2 (2015), pp. 87–96.

[9] Massieh Moayedi, Tim V. Salomons, and Lauren Y. Atlas. "Pain Neuroimaging in Humans: A Primer for Beginners and Non-Imagers". In: *Journal of Pain* 19.9 (2018), 961.e1–961.e21.

[10] Michael W L Chee et al. "Comparison of block and event-related fMRI designs in evaluating the word-frequency effect". In: *Human Brain Mapping* 18.3 (2003), pp. 186–193.

[11] Yashar Behzadi et al. "A Component Based Noise Correction Method (CompCor) for BOLD and Perfusion Based fMRI". In: 6.8 (2013), pp. 90–101.

[12] Gholamreza Salimi-Khorshidi et al. "Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers". In: *NeuroImage* 90 (2014), pp. 449–468.

[13] A Hyvärinen, Juha Karhunen, and Erkki Oja. "Independent Component Analysis". In: (2001).

[14] A Hyvärinen and Erkki Oja. "Independent component analysis: algorithms and applications." In: *Neural networks : the official journal of the International Neural Network Society* 13.4-5 (2000), pp. 411–30.

[15] John L Semmlow. *Biosignal and Biomedical Image Processing*. 2004.

[16] CIS 520. *PCA*. 2018.

[17] Y Murali Mohan Babu. "PCA based image denoising". In: *Signal & Image Processing : An International Journal* 3.2 (2012), pp. 236–244.

[18] Jean Baptiste Poline and Matthew Brett. “The general linear model and fMRI: Does love last forever?” In: *NeuroImage* 62.2 (2012), pp. 871–880.

[19] Martin Monti. “Statistical Analysis of fMRI Time-Series: A Critical Review of the GLM Approach”. In: *Frontiers in Human Neuroscience* 5.March (2011), pp. 1–13.