

TBMI02

Medical Image Analysis

Course Compendium

MRI, fMRI,
Image Registration, Image Segmentation


Anders Eklund, Mats Andersson, Hans Knutsson
{andek, matsa, knutte}@imt.liu.se

Division of Medical Informatics
Department of Biomedical Engineering
Linköping University, Sweden


October, 2010

Table of Contents


1	Course Information	4
2	Magnetic Resonance Imaging	5
2.1	History of MRI	5
2.2	Spatial encoding	6
2.3	How to create an image	7
2.4	Sampling in k-space	12
2.5	Relaxations	14
2.6	Echo types	17
2.7	Sampling patterns	20
2.8	Properties of the Fourier transform	23
2.9	3D scanning methods	24
2.10	Running the MR scanner	25
3	Functional Magnetic Resonance Imaging	28
3.1	Purpose and history of fMRI	28
3.2	fMRI compared to EEG	28
3.3	The BOLD signal and the balloon model	29
3.4	fMRI experiments	31
3.5	fMRI Analysis	32
3.6	Preprocessing of the fMRI data	33
3.7	Motion correction	34
3.8	The general linear model	35
3.9	Canonical correlation analysis	36
3.10	Visualization of brain activity	40
4	Image Registration	41
4.1	Introduction	41
4.2	Transformation models	42
4.3	Intensity based registration using gradient filters	42
4.4	Phase based registration using quadrature filters	49
4.5	Image registration by maximization of mutual information	53
4.6	Motion fields	55
4.7	Interpolation	57
4.8	Using several scales	58
4.9	3D Registration	58
5	Image Segmentation	59
5.1	Introduction	59
5.2	Methods	59
5.3	Thresholding & Classification	60
5.3.1	Thresholding	60

 yellow: missing in this pdf

5.3.2	Classification	61
5.4	Region based methods	62
5.4.1	Region growing	62
5.4.2	Watershed	63
5.4.3	Level sets	64
5.5	Border based methods	65
5.5.1	Live wire	65
5.5.2	Active contours (snakes)	65

yellow: missing in
this pdf

yellow: not valid for
TSBB31



1 Course Information

This course in medical image analysis aims to give relatively deep insights to how medical images are generated and analyzed for some of the most important systems that are used in medicine today.

The course consists of 9 lectures, 6 classes, 3 laborations and a mini project. Lectures and classes will be held in IMT1 on level 13. Laborations will be held in IMT5 on level 12.

Literature

- Course compendium
- Delivered material
- *Signal Processing for Computer Vision*, Gösta Granlund and Hans Knutsson

Examination

To pass the course you need to pass the exam, the mini project and the 3 laborations.

Mini project

The mini project is about filter optimization and image enhancement. To pass the mini project you need to show that your code works and hand in a written report.

Laborations

There are three laborations in the course. The first laboration is about magnetic resonance imaging, the second laboration is about image registration and the third laboration is about image segmentation.

To pass the laborations you have to show the working code for the lab instructor and participate in the lab seminar. The lab seminar is a part of the examination for the laborations (instead of writing a report for each laboration). Before the lab seminar each lab group will prepare a presentation for about 10 minutes for each laboration.

Communication

Examiner:	Hans Knutsson	013 - 28 67 27	knutte@imt.liu.se
Teacher:	Mats Andersson	013 - 28 67 21	matsa@imt.liu.se
Course assistant:	Anders Eklund	013 - 28 67 25	andek@imt.liu.se
Course secretary:	Anna-Carin Stragnefeldt	013 - 28 67 88	annst@imt.liu.se

2 Magnetic Resonance Imaging

2.1 History of MRI

Magnetic resonance (MR) was first used for spectroscopy rather than for imaging. The idea of spectroscopy is to examine what kinds of element an object consists of. The first MR image was published in 1973 and clinical MR imaging started in 1984 by Philips. MRI was in the beginning called nuclear magnetic resonance imaging (NMRI) but the nuclear part of the name was removed since it gave negative associations. The advantage with MRI, compared to computed tomography (CT), is that it is harmless to the patient since no ionizing radiation is used. It is also better if the goal is to examine soft tissue, such as muscle tissue or brain tissue. The disadvantage with MRI is that it can not be used for patients with metal implants, due to the strong magnetic field. The safety aspects of MRI are thus very important.

Today there are many specialized areas of MRI, such as functional MRI (fMRI) to study brain activity, diffusion MRI to measure diffusion of water, and angiography to generate images of the arteries.

Normal values for the big magnet in the MR scanners are 0.5 Tesla (T), 1.5 T, 3 T and 7 T. The strength of the magnetic field of the earth ranges from $30 \mu\text{T}$ to $60 \mu\text{T}$. The magnet in a 1.5T MR scanner is thus about 50 000 times stronger than the earth's magnetic field.

2.2 Spatial encoding

In order to know what location in the object the received signal relates to, spatial encoding is needed. With the help of gradient fields, the strength of the B field varies spatially, such that the frequencies of the spins also vary. The strength of the gradient is normally expressed in millitesla per metre (mT / m) and maximum values for gradients in a real scanner are in the range of $10 - 50 \text{ mT} / \text{m}$. The gradients can not be turned on and off in an instant, but there is a certain rise time and fall time. How fast the gradients are changed is defined by the slew rate and typical values are $20 - 150 \text{ T} / \text{m} / \text{s}$.

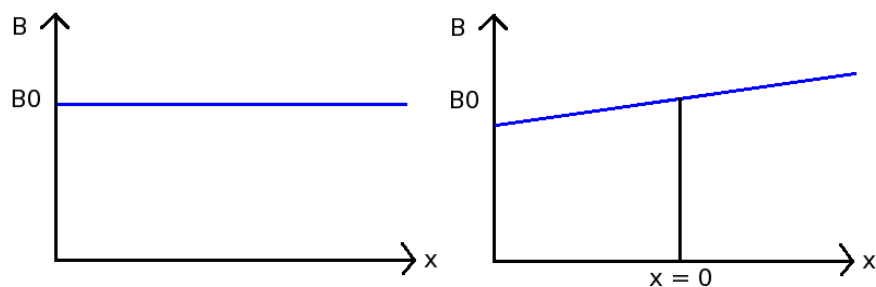


Figure 1: **Left:** Without the gradient field, all the protons experience the same magnetic field and thereby spin with the same frequency, making it impossible to know from what part of the body the signal comes from.

Right: By applying a gradient field, the protons experience a slightly difference magnetic field, making them spin with different frequencies.

The most intuitive way to decode the spatial position is to have finely tuned receivers that each collects data for one spatial position only. There are however several problems with this approach. One problem is that it would be expensive to use so many receivers and another problem is that the spatial resolution would be equal to the number of receivers. Since the protons only can spin with one frequency, it is neither possible to directly translate the spin frequency to more than a 1D position.

The solution that is used instead is to navigate in k -space, the frequency domain, with the help of three different gradient fields, one in the x -direction, one in the y -direction and one in the z -direction. In order to get the 2D (or 3D) position, we have to know the whole history of the applied gradient fields. If we sample one slice at the time, we can encode the z -position as the position of the slice that we excite.

2.3 How to create an image

How to create an image from a MR scanner can be divided into 4 steps.

1. Place the object in a strong magnetic field B_0 .
2. Send radiowaves into the object to excite the protons, this produces the alternating B_1 field.
3. Receive radiowaves transmitted by the object while varying the magnetic field. Store the data and magnetic field parameters. If needed, send more radiowaves to get more data.
4. Reconstruct the image or volume, normally by using an inverse fast Fourier transform (FFT).

Step 1

The first step leads to that the spin of the protons align to the strong magnetic field.

n_+ number of protons with low energy (parallel)

n_- number of protons with high energy (anti-parallel)

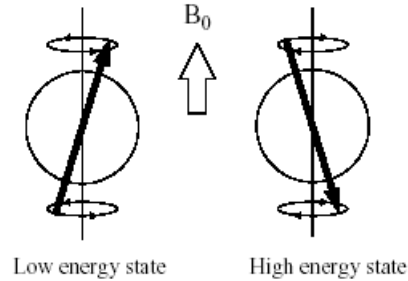


Figure 2: *Each proton can either be in the high-energy (anti-parallel) state or the low-energy state (parallel). There will be a small excess of hydrogen nuclei in the low energy state and this results in a magnetic vector pointing in the direction of the B_0 field.*

The difference between these numbers is given by

$$\frac{n_+}{n_-} = e^{-\frac{\Delta E}{kT}} \approx 1.000007 \quad (1)$$

where ΔE is the energy difference between the two states, T is the absolute temperature and k is the Boltzmann constant ($1.38 \cdot 10^{-23} J/K$). This difference is very small but it causes a bulk magnetization vector \mathbf{M} , along the direction of the B_0 field, i.e. along \hat{z} .

Step 2

In the second step we add the RF-pulses to form the B_1 field. The strength of the B_1 field is in the order of 50 mT and is perpendicular to the B_0 field. This makes the magnetization vector start precessing with the *Larmor* frequency w

$$w = \gamma B_0 \quad (2)$$

where γ is the gyromagnetic ratio and B_0 is the strength of the big magnet, $\gamma = 42.58$ MHz / T for the hydrogen nucleus. The magnetization is thus flipped down from z to the $x - y$ plane.

The flip angle α is calculated as

$$\alpha = \gamma B_1 \tau \quad (3)$$

if the RF-pulse is rectangular, τ is the time length of the RF pulse. If the pulse is not rectangular, the flip angle is calculated as

$$\alpha = \int_0^\tau \gamma B_1(t) dt \quad (4)$$

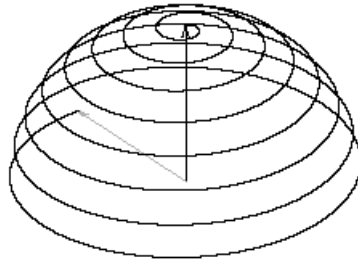


Figure 3: *When the RF-pulse is applied the magnetic vector is flipped down in the x - y -plane, along the indicated trajectory.*

In order for the receiver to pick up any signal, it has to be tuned to the Larmor frequency.

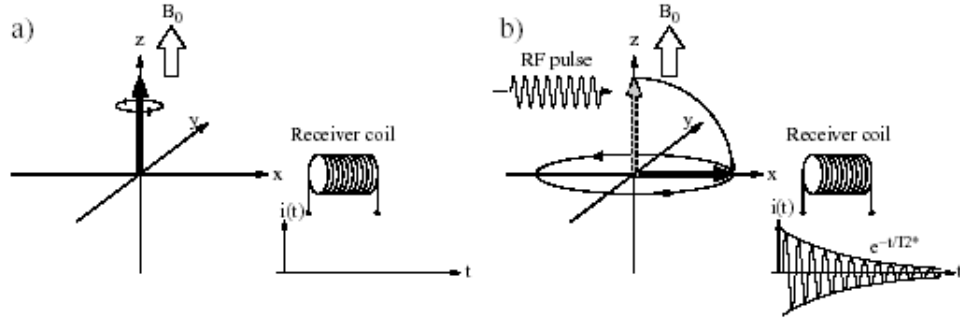


Figure 4: When the magnetic vector is flipped down in the x - y -plane it induces a current in the receiver coil. This current will decay due to the different relaxation processes.

Step 3

In the third step we add gradient fields G , such that that the total magnetic field B changes locally, and thereby the frequency also changes.

$$B(\mathbf{x}) = B_0 + G(\mathbf{x}) = B_0 + g\mathbf{x}^T\mathbf{n}, \quad \mathbf{x} = (x \ y \ z)^T \quad (5)$$

This gives the local magnetization, m

$$m(\mathbf{x}, t) = m_0(\mathbf{x})e^{-i\omega t} \quad (6)$$

where

$$\omega = \omega_0 + \Delta\omega(\mathbf{x}) = \omega_0 + \gamma G(\mathbf{x}) = \omega_0 + \gamma g\mathbf{x}^T\mathbf{n} \quad (7)$$

such that

$$m(\mathbf{x}, t) = m_o(\mathbf{x})e^{-i(\omega_0 + \gamma g\mathbf{x}^T\mathbf{n})t} \quad (8)$$

Now look at the phase of m (the integral of the frequency)

$$\varphi_{absolute} = \int_0^t \omega_0 + \gamma g\mathbf{x}^T\mathbf{n} \, d\tau = \omega_0 t + \gamma g\mathbf{x}^T\mathbf{n} t \quad (9)$$

The term $\omega_0 t$ is spatially constant and is demodulated by the receiver. The term $\gamma g\mathbf{x}^T\mathbf{n} t$ is the relative phase that we are interested in.

The signal $S(t)$ that the receiver sees is the sum over the whole object Ω_x

$$S(t) = \int_{\Omega_x} m_o e^{i\varphi} d\mathbf{x} = \int_{\Omega_x} m_o e^{i\mathbf{x}^T\mathbf{n}\gamma g t} d\mathbf{x} = \int_{\Omega_x} m_o e^{i\mathbf{x}^T\mathbf{u}t} d\mathbf{x} \quad (10)$$

where \mathbf{u} is the three dimensional frequency.

This means that what we measure actually are samples of m in the Fourier-domain, also known as the frequency domain or k-space.

A constant gradient gives a constant speed sample trajectory in k-space. Now we can change g and \mathbf{n} over time to search k-space in an appropriate way.

The phase of the spins at position \mathbf{x} and time t is given by

$$\varphi(\mathbf{x}, t) = \int_0^t \gamma g(\tau) \mathbf{x}^T \mathbf{n}(\tau) d\tau \quad (11)$$

The Fourier transform of $m(\mathbf{x})$ can be written as

$$F\{m(\mathbf{x})\} = \int_{-\infty}^{\infty} m(\mathbf{x}) e^{-i\varphi(\mathbf{x}, t)} d\mathbf{x} = \int_{-\infty}^{\infty} m(\mathbf{x}) e^{-i \int_0^t \gamma (g_x(\tau)x + g_y(\tau)y + g_z(\tau)z) d\tau} d\mathbf{x} \quad (12)$$

If we compare this to the more common way to write the multidimensional Fourier transform

$$F\{m(\mathbf{x})\} = \int_{-\infty}^{\infty} m(\mathbf{x}) e^{-i2\pi \mathbf{x}^T \mathbf{u}} d\mathbf{x} \quad (13)$$

where

$$\mathbf{u} = (k_x \ k_y \ k_z)^T \quad (14)$$

we get that the positions of the sample in k-space at timepoint t , $k_x(t)$, $k_y(t)$ and $k_z(t)$, are given by

$$k_x(t) = \frac{\gamma}{2\pi} \int_0^t g_x(\tau) d\tau, k_y(t) = \frac{\gamma}{2\pi} \int_0^t g_y(\tau) d\tau, k_z(t) = \frac{\gamma}{2\pi} \int_0^t g_z(\tau) d\tau \quad (15)$$

where g_x , g_y and g_z are the gradient fields in the x-, y-, and z-direction respectively.

Step 4

In order to get an image from the collected data, the image has to be reconstructed. How the reconstruction is performed depends on how the data have been sampled. If the data have been sampled on a regular cartesian grid, it is sufficient to use an inverse 2D FFT. Otherwise more sophisticated reconstruction algorithms have to be applied.

When the inverse FFT has been applied, one might think that we now have an image of real valued data. This is however seldom the case, mostly due to imperfections in the magnetic fields and the electronics. The image data is thus normally still complex valued, and the most common approach is to simply return the image as the magnitude of the complex valued data. There are however applications where it is necessary to use both magnitude and phase of the complex valued data.

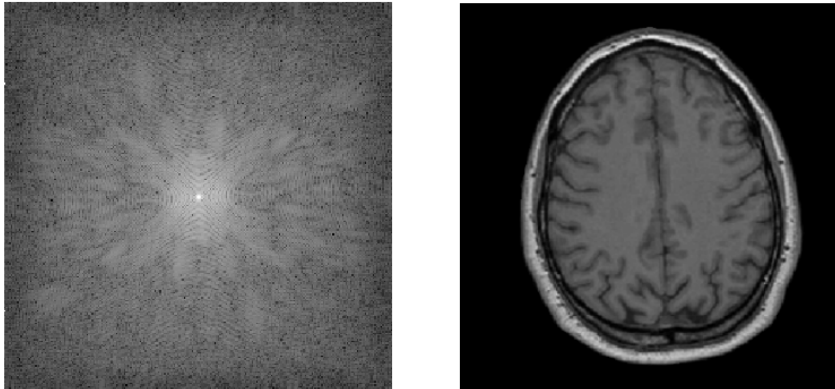


Figure 5: **Left:** *The logarithm of the magnitude of the sampled k-space.* **Right:** *The reconstructed image, using an inverse 2D FFT.*

Slice selection

In order to only excite a slice, and not the whole object, a gradient (in the z-direction) is added during the RF pulse. This leads to that only the tissue where the Larmor frequency and the radiowave frequency are the same will be excited. The thickness of the slice is controlled by the bandwidth of the RF pulse and the strength of the gradient.

2.4 Sampling in k-space

Since we are sampling in k-space, instead of in the image domain, we have to use the inverted sampling theorem.

$$\Delta k_x \leq \frac{1}{W_x} \quad \Delta k_y \leq \frac{1}{W_y} \quad (16)$$

where Δk_x and Δk_y are the distances between the samples in the x- and the y-direction respectively. W_x and W_y is the physical size of the object, given in millimeters and not in pixels. If we want to increase the field of view (FOV) we have to sample more dense in k-space. If we want to increase the spatial resolution, we have to sample further out in k-space. Note that we sample in the continuous k-space, therefore there is no π or $-\pi$.

If we make the FOV too small, we will get spatial aliasing. Compare this to when we get aliasing in the frequency domain if the sampling frequency is too low.

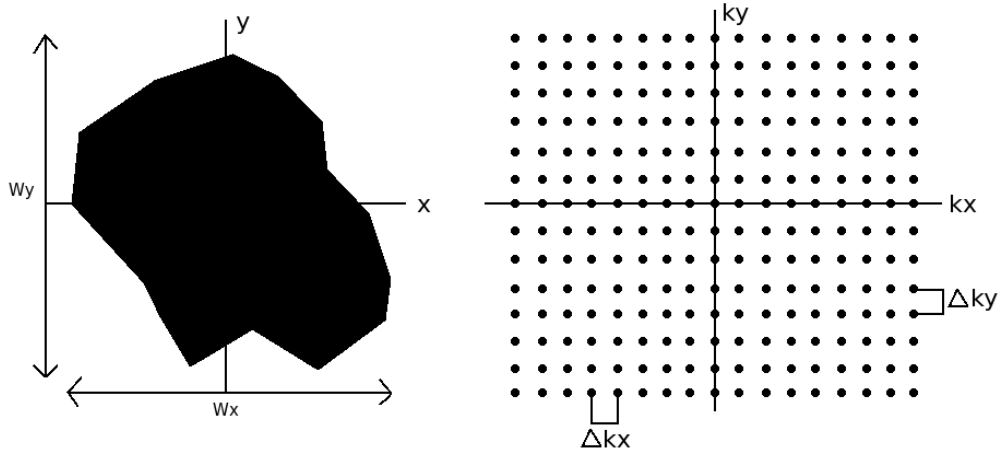


Figure 6: **Left:** The object to be scanned, with width W_x and height W_y in mm (not pixels). **Right:** The sampled k-space, each dot represents a sample. In order to avoid spatial aliasing for this object, the distance between the samples have to be less than Δk_x and Δk_y respectively.

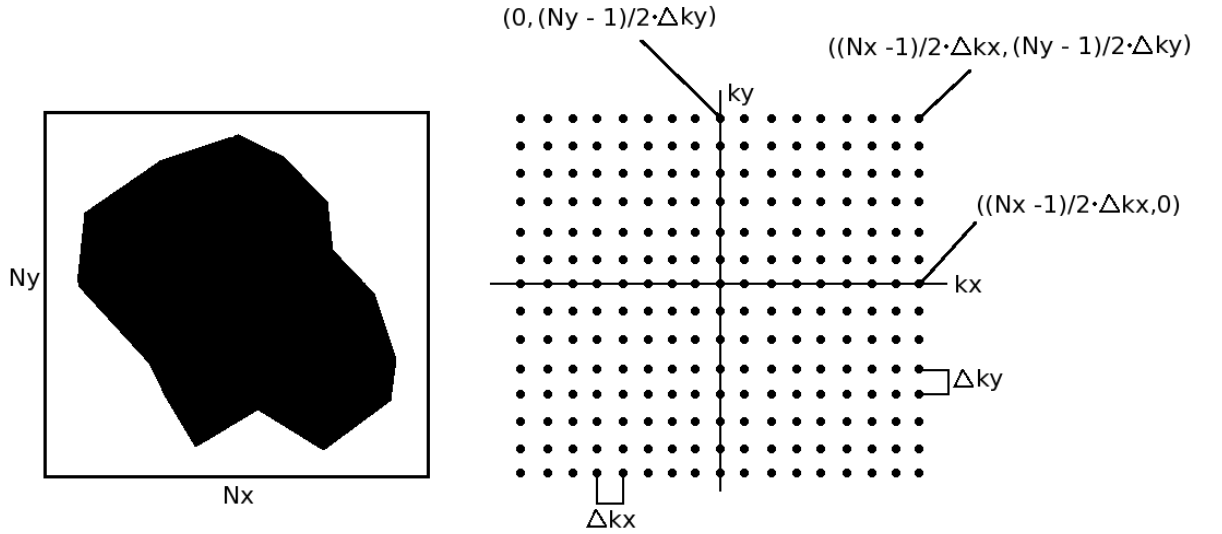


Figure 7: **Left:** We want the resulting image to have a resolution of N_x times N_y pixels. **Right:** We set the distance between the samples in k -space to $\Delta k_x = \frac{1}{W_x}$ and $\Delta k_y = \frac{1}{W_y}$ respectively. We use an odd number of samples in each direction, such that the position of the DC component is well defined. The samples to the far right in k -space thus have the position $(\frac{N_x-1}{2} \cdot \Delta k_x, \Delta k_y \cdot y)$ and the samples at the top have the position $(\Delta k_x \cdot x, \frac{N_y-1}{2} \cdot \Delta k_y)$. If we want to increase the resolution of the image, this means that we have to sample further out in k -space, and this takes more time. In order to move Δk in each timestep, this means that the strength of the gradient field should be Δk . Note the absence of π , since we sample in the continuous k -space.

2.5 Relaxations

The spins will not continue to precess forever. The signal that is measured in MRI is a function of 3 parameters, proton density (PD), relaxation due to energy loss (T_1) and relaxation due to phase incoherence (T_2). T_1 is the spin-lattice relaxation time that describes the longitudinal relaxation, i.e. how long time it takes for the magnetic moment to return from the x-y-plane to the \hat{z} direction after an RF pulse. T_2 is the spin-spin relaxation time that describes the transverse relaxation, i.e. how long time it takes for the protons to come out of phase in the x-y-plane after an RF pulse. For all tissues, T_2 is always shorter than T_1 .

The real values of T_1 and T_2 differs with the type of tissue. A coarse division can be made by dividing all the tissue into fluids (cerebrospinal fluid, synovial fluid, oedema), water-based tissues (muscle, brain, cartilage, kidney) and fat-based tissues (fat, bone marrow). The different types of tissue will have different intensities in the final image, depending on the scanner settings. Below is an example of real T_1 and T_2 values for different tissue types.

Tissue type	T_1	T_2
Fluids	1500 - 2000 ms	700 - 1200 ms
Water-based	400 - 1200 ms	40 - 200 ms
Fat-based	100 - 150 ms	10 - 100 ms

Depending on the given task, we want to produce images with different weighting. The images can be PD-weighted, T_1 -weighted or T_2 -weighted. If an image is T_1 -weighted it means that we want the T_1 contrast to be as big as possible, i.e. that we can see a difference in the image for tissues with different T_1 relaxation time.

All the effects described above are summarized in the Bloch equation, which is defined as

$$\frac{d\mathbf{M}}{dt} = \gamma \mathbf{M} \times \mathbf{B} - \frac{M_x \hat{x} + M_y \hat{y}}{T_2} - \frac{(M_z - M_z^0) \hat{z}}{T_1} \quad (17)$$

The Bloch equation describes the time dependent behaviour of \mathbf{M} in the presence of an applied magnetic field \mathbf{B} . M_z^0 is the thermal equilibrium value for \mathbf{M} in the presence of B_0 only.

From the Bloch equation we can derive the expressions for the longitudinal

$$M_z(t) = M_z^0(1 - e^{-t/T_1}) + M_z(0)e^{-t/T_1} \quad (18)$$

and the transverse relaxation

$$M_{xy}(t) = M_{xy}(0)e^{-t/T_2} \quad (19)$$

Time zero means directly after the RF-pulse. These relaxations describe what happens with the magnetized spin system after the RF-pulse, when it returns to its thermal equilibrium.

The longitudinal relaxation describes the *recovery* of the magnetization to the \hat{z} direction. This relaxation depends on T_1 and is called spin-lattice relaxation. After the time T_1 , M_z has *regained* 63% of its thermal equilibrium value.

The transverse relaxation describes the *destruction* of the phase coherence in the x-y-plane. This relaxation depends on T_2 and is called spin-spin relaxation. After the time T_2 , M_{xy} has *lost* 63% of its original value.

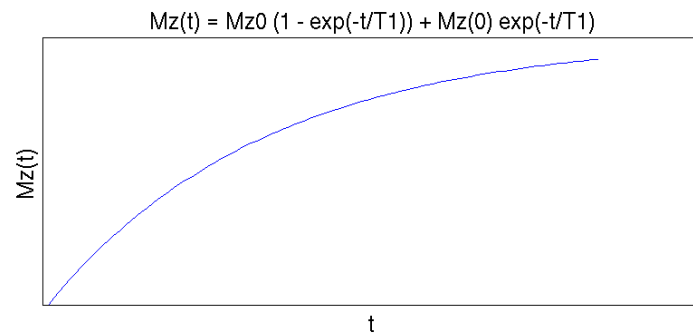


Figure 8: After the RF pulse the magnetic moment returns back from the x-y-plane to the z-direction. This is called the spin-lattice relaxation. After the time T_1 , 63% of the original z-component has been restored.

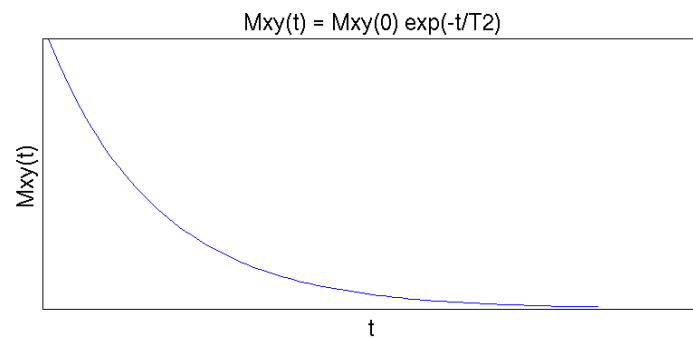


Figure 9: After the RF pulse the spins start to dephase. This is called the spin-spin relaxation. After the time T_2 , 63% of the original x-y-component has been lost.

After some time, the relaxations comes into a steady state (ss). The expressions for the relaxations in the steady state can be derived from the Bloch equation.

The longitudinal relaxation depends on the flip angle α and on the repetition time T_R

$$M_z^{ss} = \frac{M_z^0 \left(1 - e^{-\frac{T_R}{T_1}}\right)}{1 - \cos(\alpha) e^{-\frac{T_R}{T_1}}} \quad (20)$$

The transverse relaxation also depends on the echo time T_E

$$M_{xy}^{ss} = \frac{M_z^0 \left(1 - e^{-\frac{T_R}{T_1}}\right)}{1 - \cos(\alpha) e^{-\frac{T_R}{T_1}}} \sin(\alpha) e^{-\frac{T_E}{T_2^*}} \quad (21)$$

When a spin echo is used instead of a gradient echo, these expressions can be simplified since the flip angle α is always 90 degrees for a spin echo. The T_2^* is also changed to T_2 .

From these expressions we see that the repetition time is related to the T_1 relaxation, and the echo time is related to the T_2 relaxation. We have to wait a certain time T_R before we can apply another RF pulse, otherwise there is no z-component to flip down in the x-y-plane. We have to wait a certain time T_E until the echo is formed, in order to have a signal to readout.

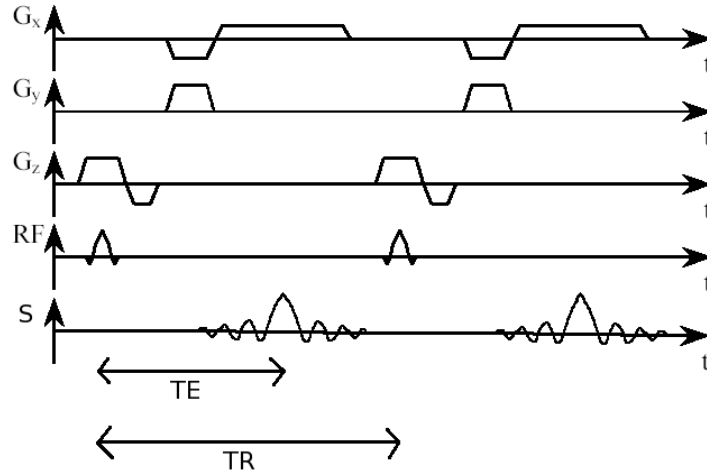


Figure 10: *The gradient fields, G_X G_Y G_z , for sampling of 2 lines using a gradient echo. RF is the radio frequency pulse and S is the signal. The time between each RF pulse is called the repetition time, T_R , and the time between the pulse and the formation of the echo is called the echo time, T_E .*

2.6 Echo types

Directly after the RF pulse is applied the magnetic moment starts to return to the \hat{z} direction and the spins start to dephase. This means that the signal strength decays with time. One might therefore think that it is best to start the sampling directly after the RF pulse when the signal is strong, but there are however at least two problems with this approach. One problem is that the coil has to be switched from being a transmitter to a receiver, and this takes some time. The main problem is though that if we sample directly after the pulse, we will not be able to see any difference in the image between the different tissue types. Since the different tissue types have different relaxation rates, we must wait a small time such that there will be a difference in the actual relaxation. The difference in the relaxation is what we see as a difference in the grey scale values in the image.

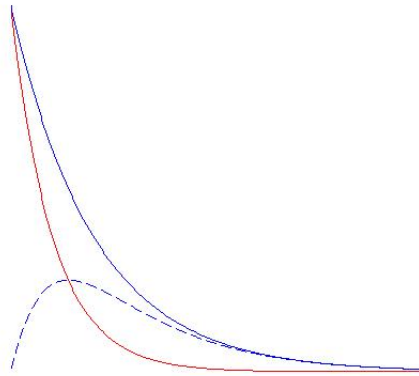


Figure 11: *Directly after the RF pulse is applied the signal is strong, but there is however no difference in relaxation between the different tissue types (blue and red lines) at this timepoint and we can thus not separate the different tissue types in the image. To get contrast in the image, we have to wait a short time before the sampling is started, in order to maximize the contrast (dashed line). If we wait for too long, the signal will be lost.*

By waiting a small time after the RF pulse, the different tissue types get a different relaxation but it also results in a dephasing of the spins. In order to get an as strong signal as possible, the spins have to be rephased again and this is done by forming an echo. There are two ways to form an echo, either by using a pulse (spin echo, SE) or by using a gradient field (gradient echo, GE). Spin echoes are also called RF echoes. Gradient echoes are also called field echoes.

To form a gradient echo, we first apply a negative gradient lobe immediately after the excitation pulse. This causes rapid dephasing of the spins.

We then apply a positive gradient, this results in a rephasing of the spins. After some time the spins will be in phase again and we have our echo. Gradient echos however suffer from the inhomogeneity of the magnetic field and will thus give worse image quality. This is denoted by T_2^* relaxation instead of T_2 . The good thing with gradient echos are though that we can vary the flip angle and thus shorten the repetition time T_R .

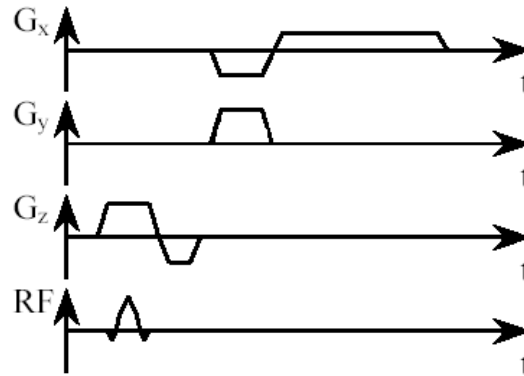


Figure 12: *The gradient fields for sampling of a line using a gradient echo. The slice selective gradient field G_z is applied at the same time as the RF pulse, to excite a specific slice. G_y is used to select the current line to sample in k -space. G_x is used to sample along this line. G_z is often called G_{ss} since the slice selection is done in the z -direction.*

To create a spin echo, we let the spins dephase naturally after the 90° pulse and then apply another pulse, that is 180° instead of 90° , such that the phase angles are reversed. This will result in that the spins rephases again and after some time we have our echo. By using the phase reversal trick, the echo height will only depend on T_2 and not on the magnetic field inhomogenities or tissue susceptibilities. Spin echoes give better image quality but takes longer time since the flip angle always has to be 90 degrees.

To speed up the sampling process, there are methods called turbo spin echo and turbo gradient echo, where several echoes are applied for one excitation. This speeds up the acquisition process, but results in worse image quality.

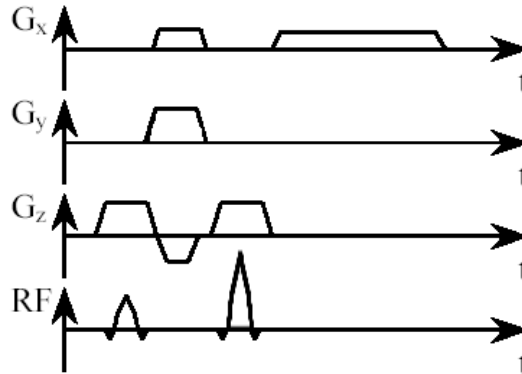


Figure 13: *The gradient fields for sampling of a line using a spin echo. Note that there are two RF-pulses, one that is 90° and one that is 180° , in order to reverse the phase angles.*

Below are the resulting weighting of the images for different repetition times and echo times, for a spin echo sequence. The flip angle is always 90° for a spin echo sequence. A short T_R here means less than 750 ms, and long T_R means more than 1500 ms. A short T_E here means less than 40 ms, and long T_E means more than 75 ms.

	Short T_E	Long T_E
Short T_R	T_1 -weighted	Not useful
Long T_R	PD-weighted	T_2 -weighted

Below are the resulting weighting of the images for different echo times and flip angles, for a gradient echo sequence. The T_R is always short (less than 750 ms) for gradient echo sequences compared to spin echo sequences. A short T_E here means less than 15 ms, a long T_E means more than 30 ms. A small flip angle here means less than 40° and a large flip angle means more than 50° .

	Short T_E	Long T_E
Small flip angle	PD-weighted	T_2 -weighted
Large flip angle	T_1 -weighted	Not useful

2.7 Sampling patterns

When the sampling is performed in k -space, different sampling patterns can be used. The most common is to use a cartesian sampling pattern, since the reconstruction then can be performed as an inverse fast Fourier transform (FFT). Sometimes it is however better to sample with other patterns, such as a spiral pattern. The reconstruction then becomes much harder since a normal FFT can not be used.

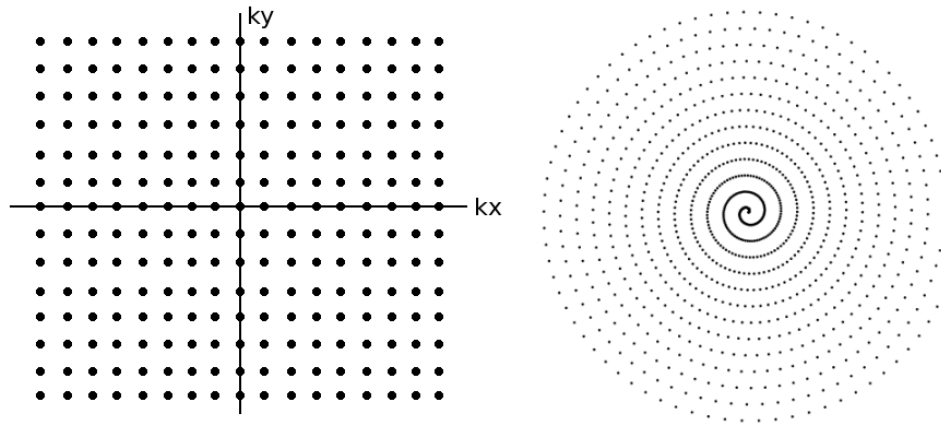


Figure 14: **Left:** Sampling k -space using a cartesian sampling pattern. **Right:** Sampling k -space using a spiral sampling pattern.

In some cases, the normal sampling in k -space, with one excitation per line, is far too slow. In functional MRI (fMRI) for example, where the objective is to study brain activity, it is necessary to get a volume of the brain every or every other second. In this case a completely different sampling approach has to be used. Instead of exciting the protons for every line of k -space, a whole slice is sampled after one excitation, see Figure 15 for the sampling pattern and Figure 16 for the gradient fields. The resulting image quality is much worse than with normal sampling, but we do not have to wait the set repetition time for each line, but only for each slice. A slice of a T_1 -weighted volume, that took about 5 minutes to acquire, and a slice of a fMRI-volume, that took about 2 seconds to acquire, are given in Figure 17.

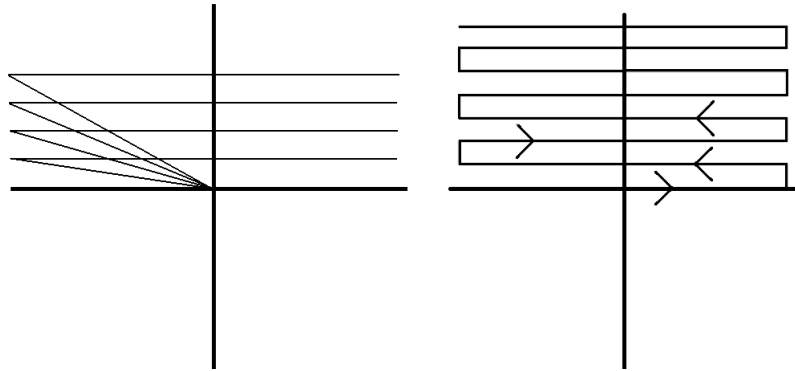


Figure 15: **Left:** For normal MRI, one line of k -space is sampled for each excitation. Each time the protons are excited we return back to origo. We have to wait the set repetition time for each line. **Right:** For echo planar imaging, the entire slice is sampled after one excitation. This results in a much faster sampling, but the image quality is much worse.

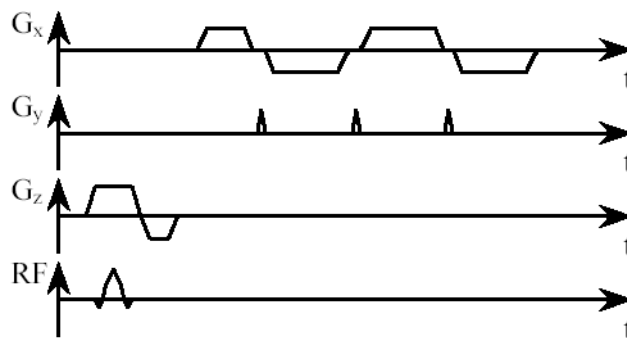


Figure 16: The gradient fields for echo planar imaging, using a gradient echo. We first excite the protons and select the current slice. There are only small blips in G_y , to change the line in k -space. G_x is alternating between positive and negative lobes, resulting in that we move to the left in k -space for half of the lines and to the right for the other half, as seen in Figure 15.

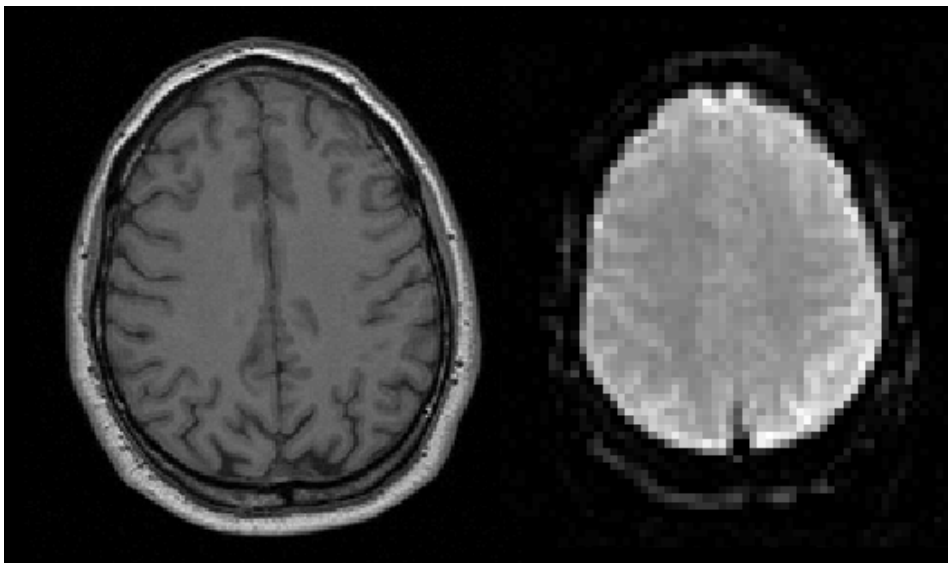


Figure 17: **Left:** A slice of a high resolution T_1 -weighted volume. The voxels have a size of $1 \times 1 \times 1$ mm. A gradient echo sequence have been used, with one excitation per line. The volume has a resolution of $240 \times 240 \times 140$ voxels and took about 5 minutes to acquire. **Right:** A slice of a low resolution fMRI volume, which is T_2^* -weighted. The voxels have a size of $3 \times 3 \times 3$ mm. A gradient echo EPI sequence has been used, with one excitation per slice. The volume has a resolution of $80 \times 80 \times 20$ voxels and took 2 seconds to acquire.

2.8 Properties of the Fourier transform

There are some properties of the Fourier transform that are very useful when sampling in k-space. The most important property is that the Fourier transform of a real valued signal is *Hermitian*. For a 1D signal this means that

$$f(-x) = f(x)^* \quad (22)$$

i.e. that the complex conjugate of the function is equal to the original function with the variable changed in sign. This property is valid for any number of variables, in 2D it can be written as

$$f(-x, -y) = f(x, y)^* \quad (23)$$

This means that we only have to sample half of the k-space to be able to reconstruct the image. Another way to look at it is that in the image domain we have $N_x * N_y$ real values. The Fourier transform of an image is though complex valued, such that we have 2 values in each pixel. For the number of values to be equal in the two domains, we can remove half of the samples in k-space. If we sample the top half of k-space, we can thus calculate what the bottom part will be. The top half is first flipped upside down, then it is flipped from left to right and finally we take the complex conjugate of the values. If we put the sampled top half and the calculated bottom half together, we have the total k-space.

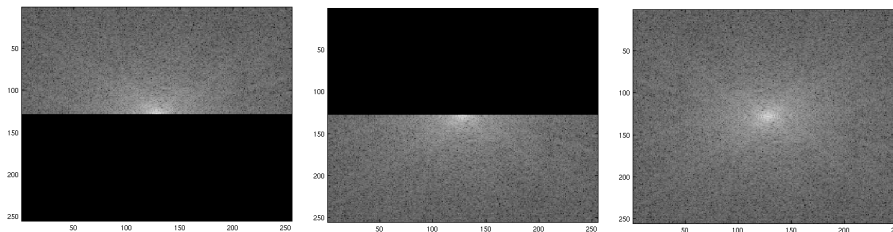


Figure 18: **Left:** The top half of the k-space has been sampled. The image shows the logarithm of the magnitude of k-space. **Middle:** The top half of k-space is first flipped upside down, then it is flipped from left to right and finally complex conjugated. **Right:** The two halves of k-space are put together to create the total k-space.

2.9 3D scanning methods

In order to scan a volume of data, several approaches can be used. The easiest approach is to consider a volume as a number of 2D slices. We excite and sample one slice at the time and then move on to the next slice. This approach is called multi 2D (M2D) but it is however rather slow since we have to wait the set repetition time T_R before we can sample the next line in k-space. In order to circumvent this another approach, called multislice (MS), is used. The difference is that we do not sample a complete slice at the time. Instead we sample one line of the first slice, then we sample the same line in the next slice and so on. The advantage with this method is that is much faster since the waiting by the repetition time can be used to excite and sample the same line in the other slices. The last approach is to use true 3D sampling, where the whole volume is excited, instead of one slice at the time.

2.10 Running the MR scanner

The area of MRI can be quite overwhelming and abstract in the beginning. To make it a bit more concrete, this section describes how a real MR scan is performed. The patient first has to sign a protocol to assure that he or she does not have any metallic implants that can be dangerous in the MR scanner. The patient is then put on the bunk and moved into the core of the big magnet. The examination is started with a survey scan, in order to know how to place the slices in the real scan. Before the real scan is started, the operator has to know what kind of tissue that is important to see in the images. From this information, the operator will set the necessary parameters, for example if the image will be T_1 -weighted or T_2 -weighted, what repetition time to use and what echo time to use, the resolution of the voxels and so on. When the scan is finished, the images are normally stored in a picture archiving system (PACS) from where the doctors can access the images.

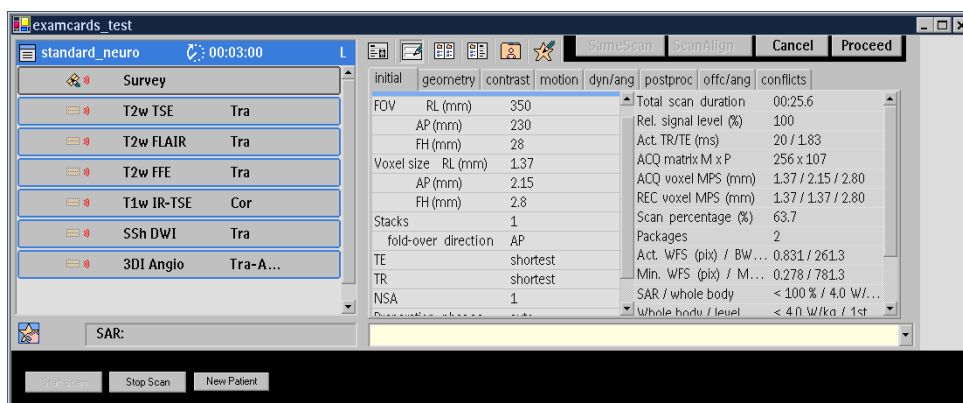


Figure 19: A screenshot of the user interface to the Philips Achieva MR scanner. To the left are the different scans. The first scan is a survey scan that is used in order to place the slices for the real scans. The second scan, T_2w TSE, is a T_2 -weighted turbo spin echo scan. In the middle are the settings that can be made by the operator. This is for example where you set the field of view (FOV), i.e. how large area the image will cover. Here RL stands for right-left, AP stands for anterior-posterior and FH stands for feet-head. Below the FOV you can set the size of the voxels. To the right you can see all the current settings, giving for example the total scan time, the repetition time and echo time and the size of the image in pixels.

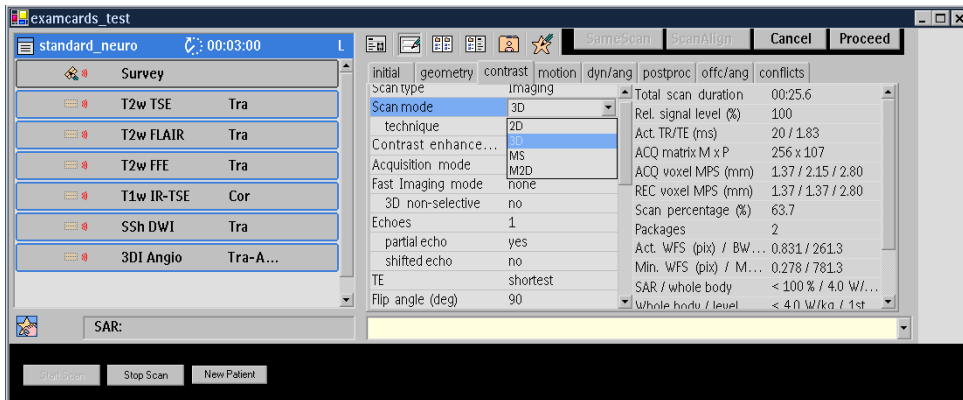


Figure 20: Selecting the scan mode, 3D, 2D, multi slice (MS) and multiple slices (multi 2D, M2D).

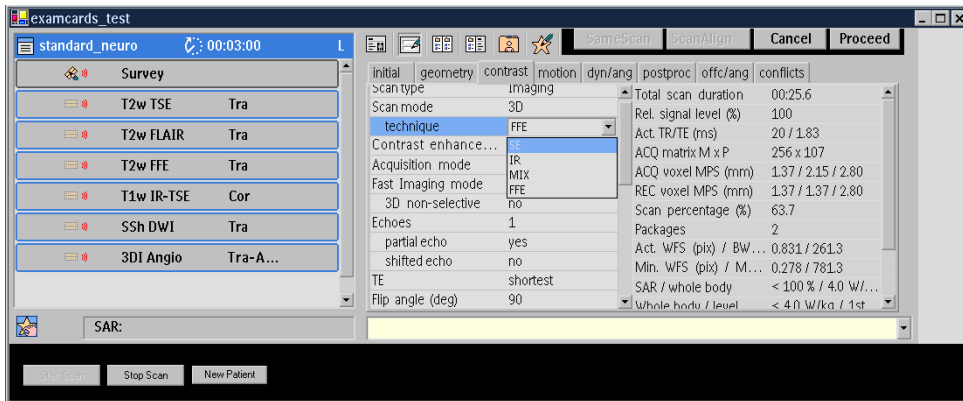


Figure 21: Selecting the scan technique to be used, spin echo (SE), inversion recovery (IR), mix, gradient echo (FFE).

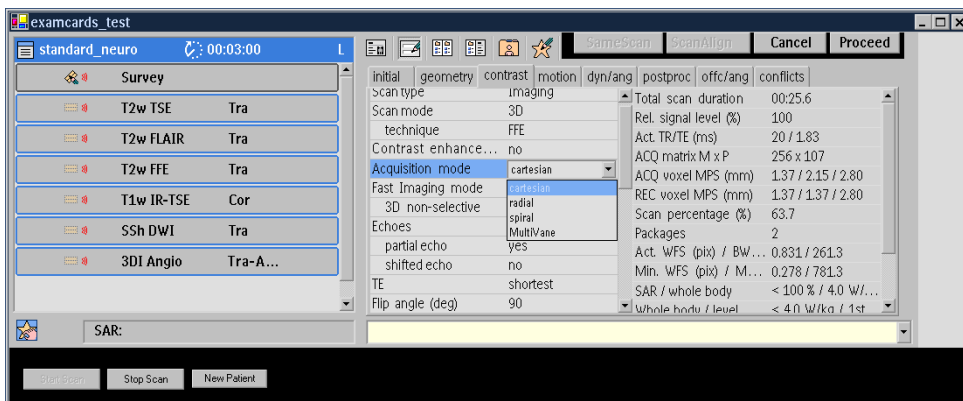


Figure 22: Selecting the sampling pattern, cartesian, radial, spiral or multi-vane.

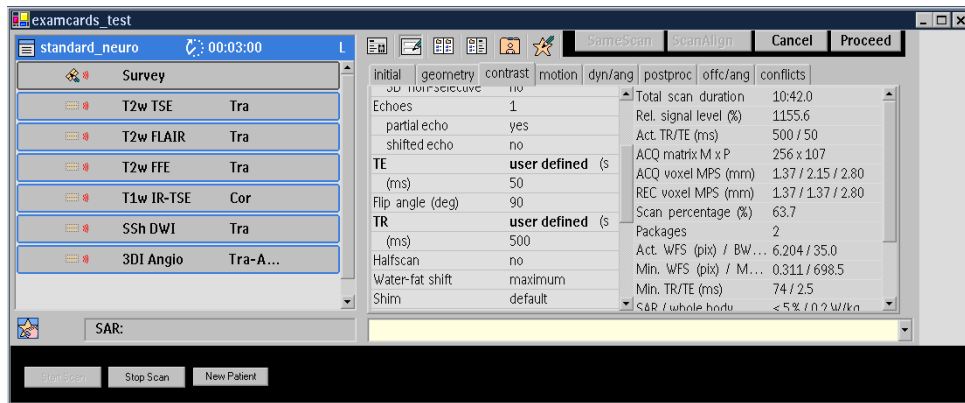


Figure 23: *Setting of the echo time (TE), repetition time (TR) and flip angle.*

3 Functional Magnetic Resonance Imaging

3.1 Purpose and history of fMRI

The main purpose of functional magnetic resonance imaging (fMRI) is to study brain activity, in order to learn more about how the brain works. fMRI is also used as a clinical tool prior to brain tumor surgery, to prevent removal of important brain areas. The advantage with fMRI compared to other techniques, is that it is non invasive, does not require any harmful radiation (compared to SPECT, single photon emission computed tomography) and has a higher spatial resolution.

fMRI was started in 1990 by Ogawa that observed that blood vessels became more visible as the amount of oxygen in the blood increased. The work was continued by Belliveau that used a contrast agent to create functional images. In 1992 Ogawa et al. published the first functional images using the BOLD signal. Today it is possible to perform the statistical analysis while the subject is in the scanner, such that the subject can look at it's own brain activity. It is also becoming more common with different kinds of brain computer interfaces, where the subject can control some system with it's brain activity.

3.2 fMRI compared to EEG

Before fMRI was introduced, the most common way to measure brain activity was with the help of electroencephalography (EEG) where electrodes are connected to the skull to measure the electrical activity. This is contrary compared to fMRI, where differences in blood flow are measured, instead of the electrical activity itself.

The advantage with EEG compared to fMRI is that it has a very high temporal resolution since it is common to have a sampling rate of 5 kHz, compared to fMRI where the sampling rate normally is 0.5 - 1 Hz. fMRI on the other hand has a very high spatial resolution, since the voxels that are collected normally are of the size 3 x 3 x 3 mm, compared to EEG where 32 - 256 electrodes is used. An fMRI volume can consist of 50 000 - 150 000 voxels and about 20% of these are normally inside the brain. To get a better signal in EEG, the electrodes are sometimes placed directly on the brain.

When EEG data is analyzed, it is normally lowpass filtered and down-sampled from 5 kHz to around 100 Hz, since there are no brain waves of higher frequencies. If 19 electrodes are used, according to the international 10-20 system, the number of useful samples are thus 1900 per second. For fMRI, a volume that is collected every second can consist of for example 10 000 brain voxels, giving 10 000 useful samples per second. The area of fMRI is also still developing, while EEG is quite old.

There have been some work on trying to combine fMRI and EEG, to get a high temporal resolution and a high spatial resolution at the same

time. The problem is though that the gradient fields in the MR scanner will introduce large artifacts in the EEG data. These artifacts can be removed to some extent, but it is still not trivial how to combine the data to achieve the high temporal and high spatial resolution. It is also necessary to use a special EEG cap that can be used in the MR scanner.

3.3 The BOLD signal and the balloon model

The main idea with BOLD fMRI is that the intensity in the acquired images depends on the level of oxygen in the blood. We thus have a blood oxygen level dependent signal (BOLD). When the brain activity increases, the oxygen consumption in the neurons increases. The body increases the cerebral blood flow to compensate for this, but overcompensates the blood flow such that the blood flow increases more than the oxygen extraction. This leads to that the amount of oxygenated blood increases and the amount of deoxygenated blood decreases. This leads to a decrease of the magnetic susceptibility, which in turn increases the T_2^* relaxation and this gives a slightly higher signal in the MRI image.

The difference for voxels with and without activity is however so small that it is impossible to see any activity by simply comparing images from a brain during rest and activity. The approach that is used instead is to let the subject follow a *stimulus paradigm*, and then perform a statistical analysis to compare the brain voxels at activity and rest.

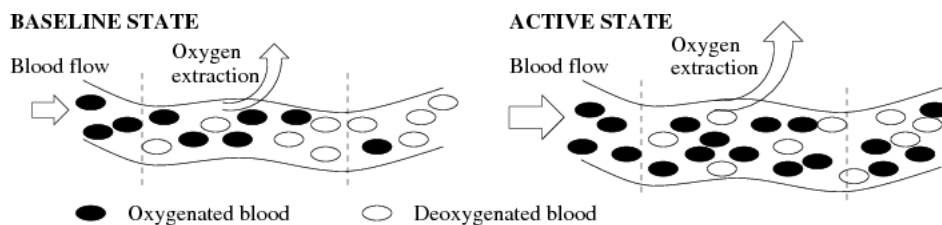


Figure 24: **Left:** The blood flow, amount of oxygenated blood and oxygen extraction during rest, the baseline state. **Right:** The blood flow, amount of oxygenated blood and oxygen extraction during activity. The neurons use more oxygen when they are active. The body increases the cerebral blood flow to compensate for this, but overcompensates such that the amount of oxygenated blood is larger than in the baseline state. The relation between oxygenated and deoxygenated blood will affect the magnetic properties of the blood and can be seen as a small change in the intensity in the images of the brain.

When a stimulus is sent to the brain, for example by activating one hand for 20 seconds, there will be a response with a certain appearance. The response will not come directly, but is delayed 3-8 seconds. First there

is an initial small dip, then there is a small overshoot and in the end there is a post stimulus undershoot.

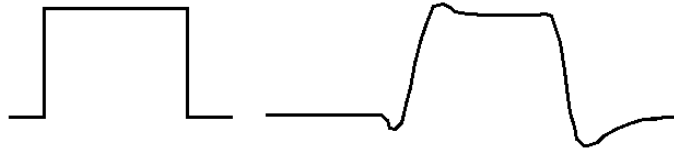


Figure 25: **Left:** The applied stimulus, for example activity for 20 seconds. **Right:** The BOLD response from the applied stimulus, the response is delayed 3-8 seconds and starts with a small initial dip. Then there is a small overshoot and a post stimulus undershoot in the end. The small initial dip is said to be due to that the amount of oxygen decreases first before the body has increased the blood flow. The difference compared to the baseline is normally not more than a few percent.

A stimulus sent to the brain, for example in the form of a square wave, will first be handled by the brain and then by the MR-scanner. The MR-scanner will capture images of the brain at different timepoints. The brain and the MR-scanner can together be seen as a black box to which we send a signal, the stimulus, and then get an image sequence from the MR-scanner from which we try to locate the brain activity.

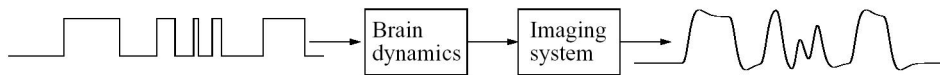


Figure 26: The known stimulus paradigm passes two main systems, the brain and the MR-scanner, before it can be measured as intensity variations in the acquired MRI-images.

The balloon model tries to describe how the blood flow in the brain reacts to a stimulus. The BOLD signal in the balloon model is defined as $\frac{\Delta S}{S}$ and can be calculated with the following equation.

$$\frac{\Delta S}{S} = V_0 \left(k_1(1 - q) + k_2 \left(1 - \frac{q}{v}\right) + k_3(1 - v) \right)$$

To calculate the BOLD signal we need to know the normalized venous blood volume v and the normalized total deoxyhemoglobin content q . k_1 , k_2 and k_3 are three constants.

The differential equations for v and q are in the model given by the following expressions

$$\begin{aligned}\frac{dq}{dt} &= \frac{1}{\tau_0} \left(f_{in}(t) \frac{E(t)}{E_0} - f_{out}(v) \frac{q(t)}{v(t)} \right) \\ \frac{dv}{dt} &= \frac{1}{\tau_0} (f_{in}(t) - f_{out}(v))\end{aligned}$$

$f_{in}(t)$ is in the known stimulus paradigm and $f_{out}(v)$ can be modelled to vary linearly or exponentially with v . Observe that $f_{in}(t)$ is a function of time and that $f_{out}(v)$ is a function of volume. E is the net extraction of O_2 from the blood as it passes through the capillary bed and it can be modelled with the following equation

$$E(f_{in}(t)) = 1 - (1 - E_0)^{\frac{1}{f_{in}(t)}} \quad (24)$$

For more information about the Balloon model, see the article by Buxton et al.

3.4 fMRI experiments

In order to perform a fMRI experiment, the stimulus paradigm first has to be designed. The paradigm is used to tell the subject what to do, and to know when the activity starts and stops. The design of the fMRI experiments can be divided into two groups, block designs and event related designs.

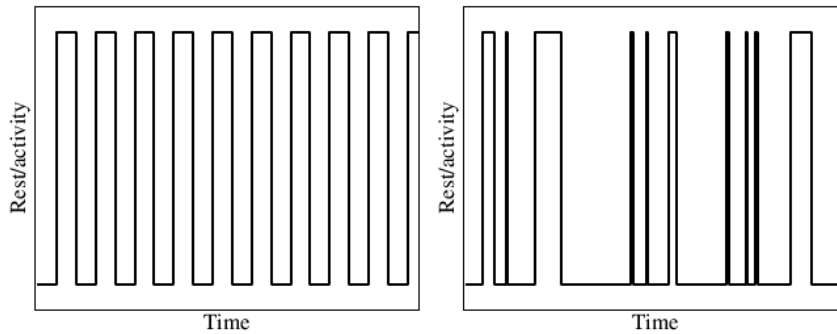


Figure 27: **Left:** A block stimulus paradigm where the periods of activity and rest normally are equally long, for example 20 seconds. **Right:** An event related paradigm where the active periods consists of spikes or blocks with varying length.

During the experiment, volumes of the brain are collected continuously. It is common to collect a volume every or every other second, giving a temporal resolution of 0.5 - 1 Hz. The subject receives instructions through virtual reality goggles or through earphones. It can however be hard to use the earphones, since the MR scanner is very noisy.

3.5 fMRI Analysis

There are many ways to perform the statistical analysis of the fMRI data, some examples are principal component analysis (PCA), independent component analysis (ICA), correlation etc. The most common approach is though to use the general linear model (GLM) that is implemented in the statistical parametric mapping (SPM) software by Friston et al. This results in a t-test value for each voxel and then a significance level is set, for example 95% or 99%. Each voxel time series is tested to examine if there is a significant difference of the values during rest and activity. The voxels that have a t-test value that is significant are considered to be active. Since the t-tests are performed separately, there is always a risk of that voxels that not are active still will be considered to be active. If the length of the rest and activity periods are the same, the t-test value can be translated to a correlation value, that states the correlation between the stimulus paradigm and the current timeseries.

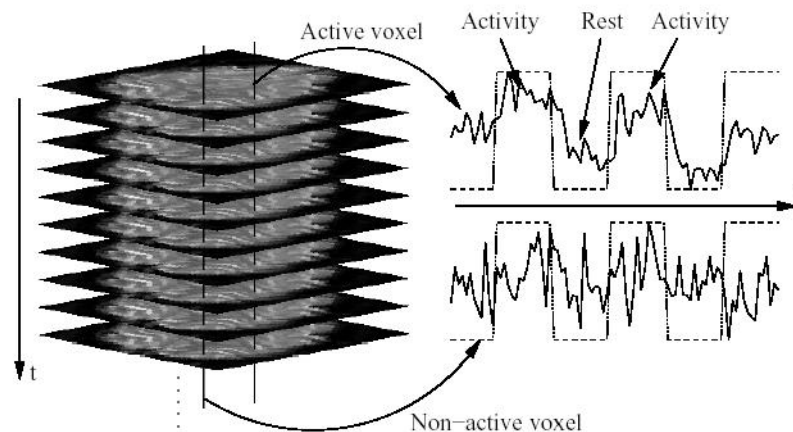


Figure 28: *In fMRI a number of slices is collected repeatedly during the experiment. It is common to have a volume of data for each second, the fMRI data is thus 4D. Each voxel will have a timeseries, if this timeseries is similar to the stimulus paradigm, the voxel is considered to be active.*

3.6 Preprocessing of the fMRI data

Before the statistical analysis is performed, different preprocessing steps are normally applied to improve the statistical analysis. Some of these preprocessing steps are further described in later sections.

- Since it is possible to move the head during the experiment, and the fact that an experiment normally is several minutes long, it is necessary to perform *motion correction* of the collected fMRI volumes such that they are aligned.

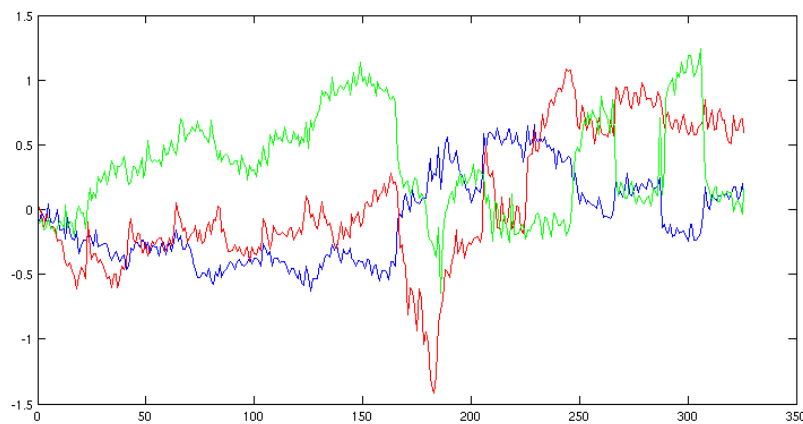


Figure 29: In fMRI it is common to compensate for head movement prior to the statistical analysis. This plot shows the estimated translations of the head from a 5 minute long fMRI experiment. The blue curve is the movement in the x -direction, the red curve is the movement in the y -direction and the green curve is the movement in the z -direction. The maximum movement is only about 1.5 mm, but it is sufficient to ruin the fMRI analysis.

- The fMRI volumes are normally collected by sampling of one slice at the time, using echo planar imaging (EPI). This means that the slices in the volume are not taken at the same time. If the volume consists of 20 slices and it takes 2 seconds to acquire the volume, there is a 0.1 s difference between each slice. In order to compensate for this, *slice timing correction* is applied.

- Due to imperfections in the scanner and brain activity that not can be controlled, there will be drifts and trends in the fMRI data that has to be removed prior to the statistical analysis. This is called *detrending* and can be done in many ways.

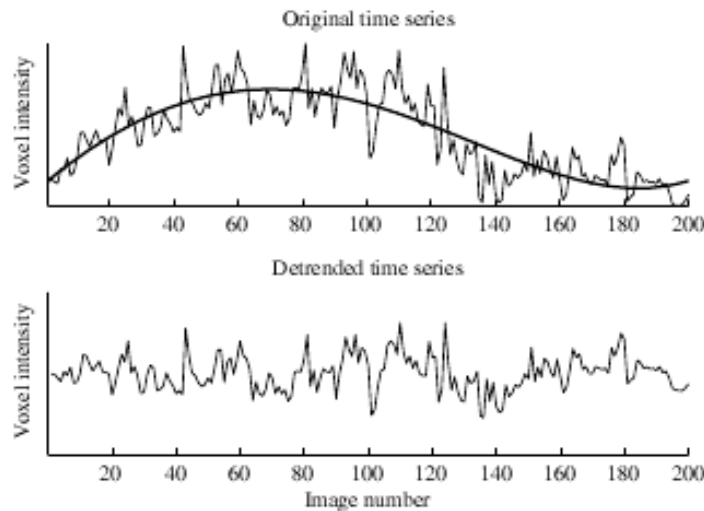


Figure 30: *In fMRI data it is common to have different kinds of trends, from the MR scanner and from the subject. These trends will disturb the statistical analysis and should be removed. **Top:** A timeseries of a voxel, it is obvious that there is a slow trend in the data. **Bottom:** The trend has been removed prior to the statistical analysis.*

- Prior to any statistical analysis, it is common to *normalize* the data such that it has zero mean and unit variance.

3.7 Motion correction

Since a fMRI experiment takes a couple of minutes to perform, there is always a risk that the subject will move the head during the experiment and this can result in false activation in the fMRI analysis. It is common to see activity close to the edge of the brain and the reason for this is that these voxels can change from being inside the brain and outside the brain if the subject moves. These voxels will thus have a much higher variance, that can be misinterpreted as activity, especially if the subject moves the head in pace with the stimulus paradigm. In order to compensate for the head movement, motion correction algorithms must be used, this is also known as image (or volume) registration.

The area of motion correction for MRI can be divided into prospective and retrospective motion correction. The idea with prospective motion cor-

rection is to compensate for the head movement *before* the sampling of the new volume, while retrospective motion correction tries to compensate for the movement *after* the acquisition of the volume.

Prospective motion correction tries to estimate the position of the head using a number of navigator echoes, and then adjust the gradients such that the k-space coordinate system is locked to the head of the patient, instead of to the MR scanner. This however normally requires that the MR scanner is reprogrammed.

The idea with image registration is to find the translation and rotation of each fMRI volume, compared to a reference volume, and then translate and rotate back the altered volume, using some kind of interpolation. Since the head can be considered to be a rigid body, no deformations are allowed in the registration.

The problem with fMRI sampling is however that the acquisition sequence that is normally used samples one slice at the time. It is thus possible to have the head in different positions in the different slices. Another way to look at it is that in order for the head movement to be compensated by a rigid 3D registration, all the movement has to be between the last slice of the previous volume and the first slice of the current volume.

3.8 The general linear model

The general linear model (GLM) is the most used statistical method for fMRI analysis and is implemented in a software package called SPM (statistical parametric mapping). A general linear model explains the variation in Y_j in terms of a linear combination of the explanatory variables, plus an error term. This can be written as

$$Y_j = x_{j1}\beta_1 + \dots x_{jl}\beta_l + \dots + x_{jL}\beta_L + \epsilon_j \quad (25)$$

where B_l are the unknown parameters corresponding to each of the L explanatory variables. This approach is also known as *regression analysis*.

In matrix form this can be written as

$$Y = X\beta + \epsilon \quad (26)$$

where Y are the observations, i.e. all the voxels in the fMRI dataset, β are the parameters that we want to find, ϵ are the errors that can not be explained by the model and X is the design matrix that is given by the experiment setup.

In order to find the parameters that minimize the least square error, one can show that the following equation must be fulfilled

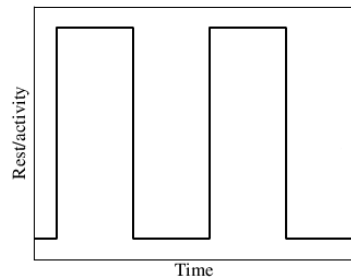
$$X^T Y = (X^T X)\hat{\beta} \quad (27)$$

where $\hat{\beta}$ are the estimated parameters. This gives the following expression for how to find the best parameters

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (28)$$

This is called the best linear unbiased estimator (BLUE).

The parameters that we want to find are the weights of the temporal basis functions that can be combined to fit the BOLD response of a voxel timeseries.



(a) Paradigm.

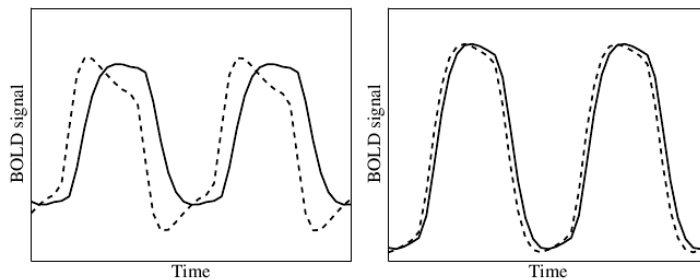


Figure 31: **Top:** *The stimulus paradigm.* **Bottom left:** *Two temporal basis functions, solid line and dashed line, that can be combined to different BOLD responses.* **Right:** *Two different resulting linear combinations of the two temporal basis functions.*

3.9 Canonical correlation analysis

The problem with the GLM is that we do not fully take advantage of the fact that if one voxel is active, there is a high probability that the neighbouring voxels also are active. To include information from the neighbouring voxels, and to increase the signal to noise ratio, it is common to apply an isotropic lowpass filter prior to the statistical analysis with the GLM. It is however a better idea to use adaptive filtering, such that lowpass filters with different shape and different size are used in different parts of the fMRI volumes. The GLM can however only find one set of parameters, in order to find the

best *temporal* parameters and the best *spatial* parameters at the same time we have to use another approach, one way to do this is to use canonical correlation analysis (CCA) instead.

CCA finds the linear combinations of *two* multidimensional variables that gives the highest correlation. We can think of it as correlation between the projection of two multidimensional variables x and y . The projections are given by $w_x^T x$ and $w_y^T y$, i.e. a linear combination of the different variables in each dataset.

Ordinary correlation ρ can be written as

$$\rho = \frac{E\{(x - \mu_x)^T(y - \mu_y)\}}{\sqrt{E\{(x - \mu_x)^T(x - \mu_x)\}E\{(y - \mu_y)^T(y - \mu_y)\}}} \quad (29)$$

if we now use that $x = w_x^T x$ and $y = w_y^T y$, and assume that the data has zero mean, we get

$$\rho = \frac{E\{w_x^T x y^T w_y\}}{\sqrt{E\{w_x^T x x^T w_x\}E\{w_y^T y y^T w_y\}}} \quad (30)$$

This can also be written as

$$\rho = \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}} \quad (31)$$

where C_{xy} is the correlation matrix between the variables, and C_{xx} and C_{yy} are the correlation matrices for x and y respectively.

To find the best linear combinations, we simply take the derivative of ρ with respect to w_x and w_y . We though have to make sure that the weight vectors have unit length, we denote this with \hat{w}_x and \hat{w}_y . It is then easy to show that the best weights are given by an eigenvalue problem

$$C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} = \lambda^2 \hat{w}_x \quad (32)$$

$$C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy} = \lambda^2 \hat{w}_y \quad (33)$$

such that \hat{w}_x and \hat{w}_y are found as eigen vectors and the correlation is the square root of the corresponding eigen value.

As temporal basis functions, we can for example use 3 sine waves and 3 cosine waves that are of the same frequency as the stimulus paradigm, twice the frequency and three times the frequency. By using both sine and cosine waves, we can create a sine wave with arbitrary phase, to compensate for the unknown BOLD delay.

As spatial basis functions, we can use different combinations of the neighbouring pixels or voxels. The most simple case is the pixel base but a filter base is better to use. If we use a number of filters with different orientation, a combination of them can result in a filter with arbitrary orientation.

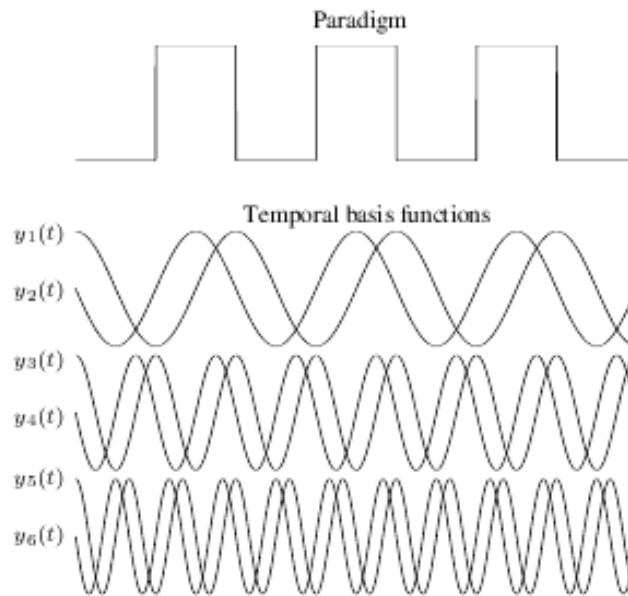


Figure 32: **Top:** The stimulus paradigm. **Bottom:** The six different temporal basis functions that can be combined to fit the BOLD response.

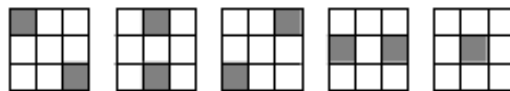


Figure 33: As spatial basis functions we can use different combinations of the neighbouring pixels.

If we use CCA for the fMRI analysis it will thus find the combination of temporal basis functions and the combination of spatial basis functions that gives the highest correlation with the timeseries of the current voxel. If we use too many basis functions however, CCA will find high correlations everywhere.

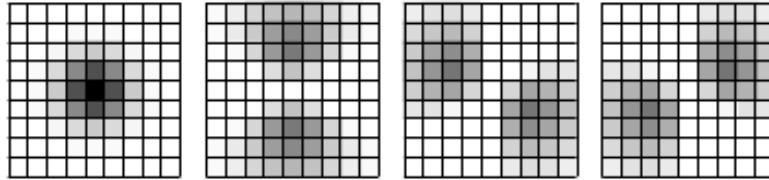


Figure 34: *A better approach is to use a number of filters with different orientation, that can be combined to a filter with arbitrary orientation.*

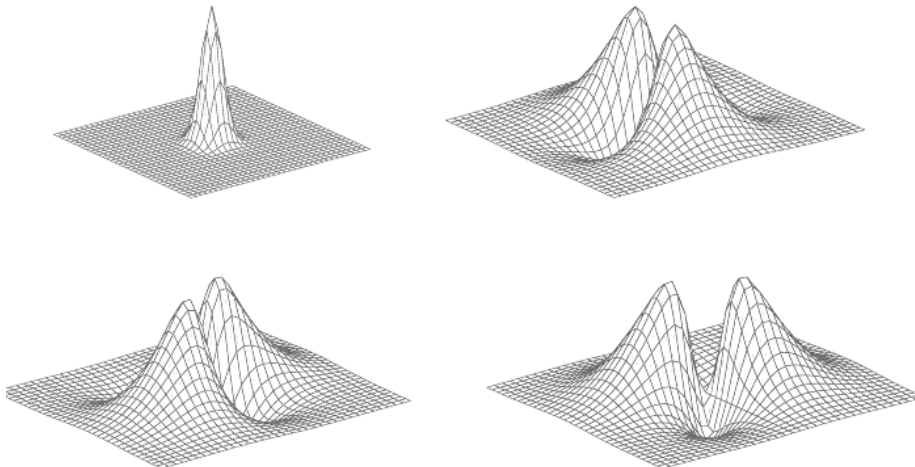


Figure 35: *The same filters as in the previous image, but now we look at the mesh of them.*

3.10 Visualization of brain activity

The statistical analysis of the fMRI data results in an activity map, where each voxel has an activity value that states how active that voxel was in the experiment. To visualize this, it is common to put the low resolution activity map into a high resolution T_1 -weighted MRI volume, to easier see the activity and the anatomical structure of the brain at the same time. Since the voxels in the fMRI volumes normally have a physical size of something like $3 \times 3 \times 3$ mm and the voxels in the T_1 -weighted volume normally have a physical size of $1 \times 1 \times 1$ mm, the activity map first have to be interpolated to the same resolution as the T_1 volume. Then it is also necessary to register the two volumes, for example by maximizing the mutual information between them.

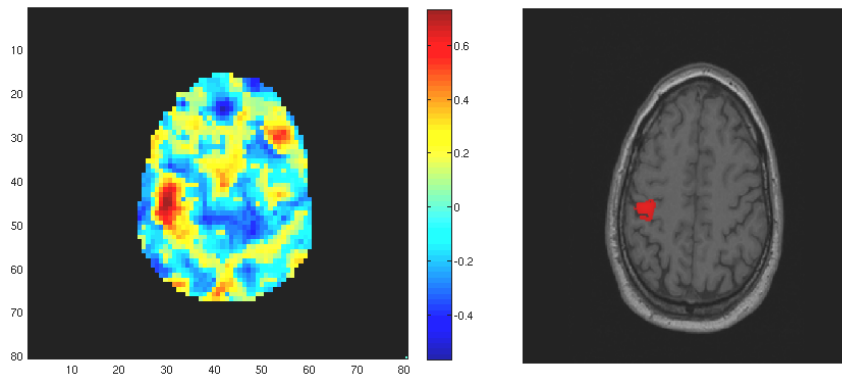


Figure 36: **Left:** The correlation map for a slice of the fMRI volume. The correlation is colour encoded and the scale is given by the colour bar.

Right: A threshold has been used and the activity map has been registered to the T_1 volume, to make it easier to see the activity and the anatomical structure at the same time. The subject periodically activated the right hand, and we can see activity in the left motor cortex.