LECTURE 1

FUNDAMENTALS OF IMAGING

Lecture Notes by Assaf Tal

AN OVERVIEW OF IN-VIVO IMAGING

Imaging is Concerned With Mapping Meaningful (Physiological) Quantities Spatially

Imaging in general is concerned with mapping the spatial distribution of some parameter or set of parameters. For example, we might want to map the temperature T of a body as a function of position, T(**r**). This can be done using infrared cameras that image wavelength ~ 10 μm (these introduce questions of *penetration depth*, as IR radiation gets absorbed after a certain distance traveled in most matter).

One Rarely Maps The Quantity of Interest Directly

Rarely does one map a quantity directly, and most imaging is indirect: one maps a function *f* of the desired quantity, f(T(**r**)), and attempts to extract T out of it. An IR camera does not yield the temperature T directly. Rather, one measures radiation intensity in the infrared spectrum. The intensity depends on the body's temperature, and one needs to be able to connect the two: temperature ↔ intensity. If one is lucky, there exists a known as simple relation between the two, but sometimes this is nontrivial. For example, the amount of IR energy, f(T(**r**)) might be proportional to the temperature: f(T)=αT, which would imply that, at each point **r** of our final image,

$$f\left(T\left(\mathbf{r}\right)\right)=\alpha T\left(\mathbf{r}\right)$$

Of course, real life can be a lot more complicated, and there might not even exist a closed form expression (I'll calm you right now and state that in MRI there usually is one such expression).

The Hallmarks of a Good Image: SNR, Resolution, Time

From an engineering point of view, the **signal to noise ratio** (SNR), **spatial resolution** and total **acquisition time** are the defining characteristics of an image. These are all very intuitive quantities. First, resolution:

Image (10 x 10)

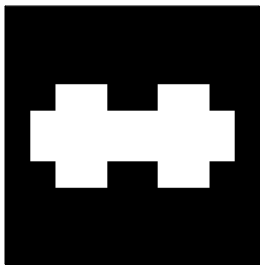
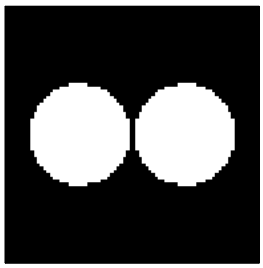


Image (30 x 30)

Image (80 x 80)

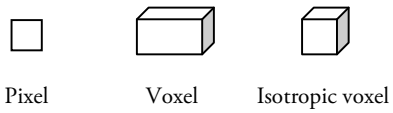


The images show the same object, two circles, with increasing image resolution (number of pixels). One talks of **pixels** when discussing 2D images, and **voxels** (=volumetric pixels) when discussing

3D images, and the **nominal spatial resolution**[1] is defined as the size of the pixel/voxel. The resolution could be different along different axes in space. An **isotropic voxel** is one having the same nominal resolution along all three spatial axes.

Modality	Typical Spatial Resolution (Linear, in mm)	Typical Temporal Resolution	Picks up signals from ...
MRI: Anatomical (3D)	1	2-5 minutes	Water
MRI: Spectroscopy	10	~10 minutes	Small metabolites
MRI: BOLD-fMRI	2-3	~sec	Oxygenated blood
Positron Emission Tomography	1-5	20-60 minutes	Radiotracers (FDG, ^{82}Rb-Chloride, etc)
Ultrasound	1	~10-100 ms	Tissue acoustic impedance
Computerized Tomography (CT)	1	few minutes	Tissue radiation absorption
Screen film radiography ("X-ray")	0.1		Tissue radiation absorption
Magnetoencephalography (MEG)	1	0.1-1 ms	Magnetic fields generated by Electrical currents
Electroencephalography (EEG)	Problematic ...	0.1-1 ms	Electrical currents

Table 1. Some typical imaging modalities in humans and their resolutions



Pixel Voxel Isotropic voxel

In general, one needs to specify the spatial resolution along each of the spatial axes of the voxel. Typical spatial resolutions of imaging modalities are shown in Table 1.

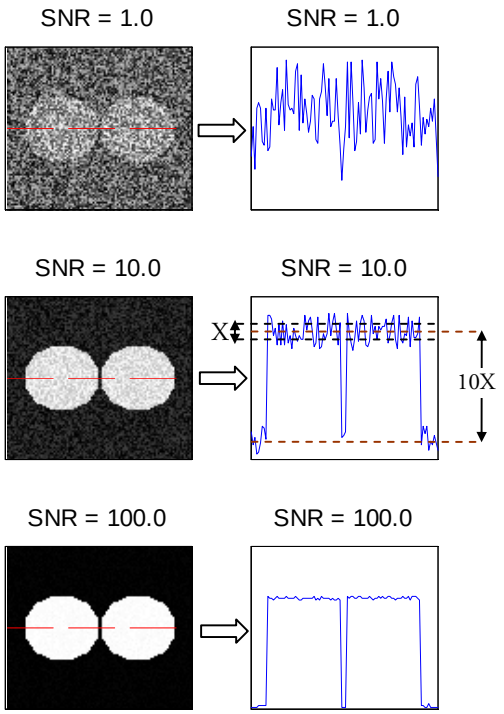
In many imaging modalities the **temporal resolution** is also important. For example, if we are interested in imaging the beating heart (~1 Hz, or ~ 1 beat per second), our imaging apparatus must be able to produce images at a much faster rate with a temporal resolution of ~10-100 ms.

The signal to noise is defined for a region, and is the ratio of the absolute value of the signal at a point divided by the standard deviation (SD) of the noise in the image. In many images, MRI included, noise is mostly random, so the noise's SD is calculated from an "empty" region containing no image and just noise. This is not always a perfect choice for reasons we'll encounter down the road, but it should give a rough estimate for the SNR.

$$SNR = \frac{f(T)}{\sigma}$$

A high SNR makes the image more discernable from the background. It also makes inverting f(T) (to obtain T) easier. This is illustrated in the figure below.

[1] Most people omit the "nominal" when talking about number of pixels or voxels. Later on when we introduce the point spread function we'll see that nominal resolution doesn't tell all the story.



Finally, many systems require a considerable amount of time to acquire 2D or 3D images of a quantity of interest. A simple X-ray image takes less than a second to obtain, if one does not factor in the time needed to position the patient. An MRI scan can taken from seconds to minutes to tens of minutes. CT scans and ultrasound scans also take minutes to complete, and additional time to be deciphered by a specially trained radiologist. Radiology of the sub-specialty of medicine which is concerned with reading medical images.

Sometimes, Contrast Beats All

One often images a particular imaged quantity, T, not because it is interesting in itself but because it changes between, say, healthy and pathological tissue. Temperature, for example, is known to increase when there is infection. One is then interested in resolving the difference:

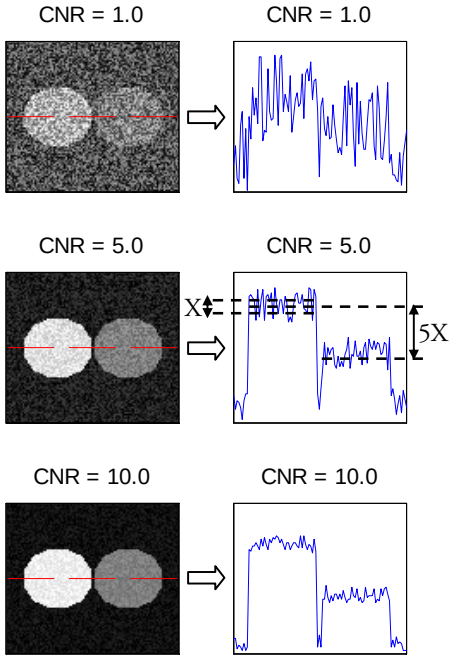
$$T_{\text{healthy}} - T_{\text{pathological}}$$

The ability to resolve this difference will depend on whether it is big or small in comparison to the "natural variation" in the image – that is, the noise. Therefore, the quantity we want to maximize is the ratio of the two, known as the **Contrast to Noise Ratio** (CNR):

$$CNR = \frac{f\left(T_{\text{healthy}}\right) - f\left(T_{\text{pathological}}\right)}{\sigma}$$

A CNR>>1 is necessary for reliable differentiation between the two tissues.

The following figure shows two spheres with increasing CNR:



It is interesting to note that it is much easier for us to detect a difference between the spheres in the CNR=1.0 case when viewing the 2D image compared to the 1D slice through the image. This is due to multiple factors, some of which physical and some of which are rooted in neuropsychology and the way our brain and vision work (e.g. in the 2D case we're "integrating" over a larger region which gives us a better "effective CNR").

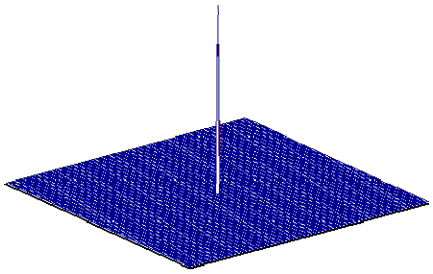
Quantities Are Averaged Over a Voxel

When one presents you with a map of some quantity T(**r**), you must keep in mind that the quantity is not really measured at an infinitely sharp point **r**, but almost always is averaged somehow over the entire pixel (2D) or voxel (3D). Therefore, when we speak of "the temperature at

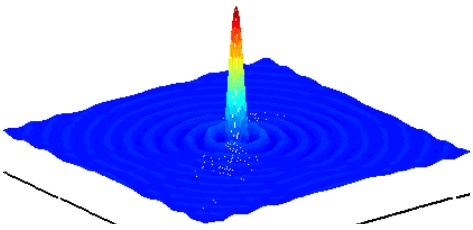
the point $\mathbf{r}_0$," we really mean "the temperature averaged over a voxel centered at point $\mathbf{r}_0$." One should always be careful when thinking of such quantities: if the voxel is much larger than the region in which temperature rises, the changes will be "washed out" by the averaging. This is known as a **partial volume effect**.

The Point Spread Function

Almost every imaging system can be analysed in terms of its **point spread function**, or PSF, which is the answer to the following question: if I put an infinitely narrow object in my imaging system (what physicists call a "delta function"), how will my final image look like? The resulting image is rarely an infinitely narrow object, but some finite, "blurred" version of it:



Input function: "delta" at $\mathbf{r}=0$



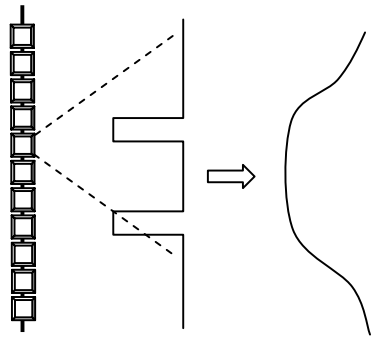
Ouput: The point spread function

One can now treat a "real" object as made out of many infinitely narrow "delta functions." The final image is obtained by converting each "point" in the original image into the PSF, and adding everything up. This process is known as a convolution.

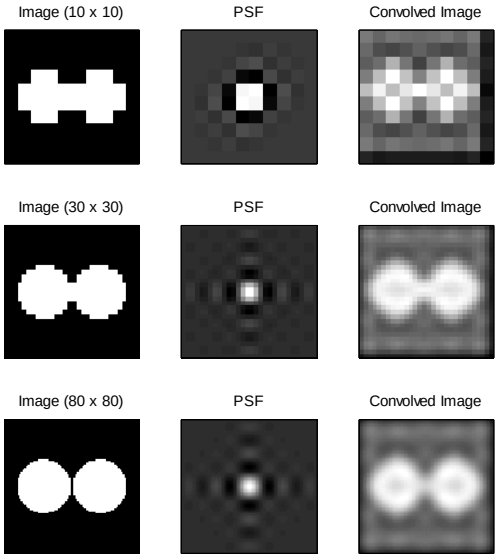
A **convolution** of an image $I(\mathbf{r})$ with a point spread function $PSF(\mathbf{r})$ is a transformation in which every point $\mathbf{r}_0$ in $I(\mathbf{r})$ is replaced by $PSF(\mathbf{r}-\mathbf{r}_0)$. Mathematically, we denote this convolution as $(I \otimes PSF)(\mathbf{r})$ and define it as:

$$(I \otimes PSF)(\mathbf{r}) = \int_{\substack{\text{all} \\ \text{space}}} I(\mathbf{r}')\, PSF(\mathbf{r}-\mathbf{r}')\, d\mathbf{r}'$$

We've remarked that nominal resolution only tells half the story, and the PSF is the other half. Even if the nominal resolution is very high, you still might not be able to resolve structures if your PSF is too wide. For example, suppose that you have many detectors that are very finely spaced apart in 1 mm intervals. However, each detector isn't very good and it captures light from a wide range of positions, about 1 cm wide. Even though you have many finely spaced detectors – your nominal resolution is high – you will still not be able to resolve structures that are closer than a centimeter, because they'll get blurred up by the PSF of the detectors:

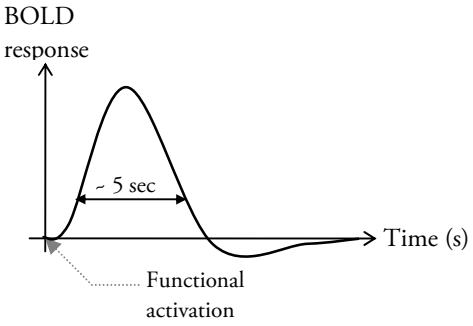


This is also illustrated in the following set of drawings which show what happens as we increase the spatial resolution of the two spheres presented in the previous sections, but convolved with the same unchanging PSF:

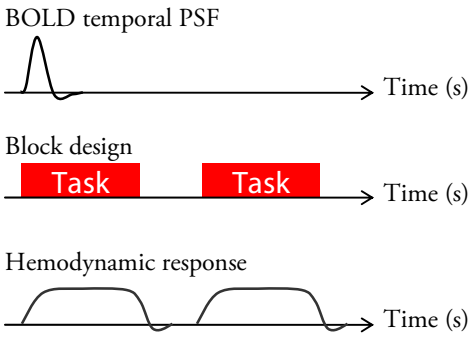


We see that past a certain point increasing the resolution does little to improve the final image.

The discussion focused on **spatial PSF**s which are easy to visualize. However, the imaging process also has a **temporal PSF**. A very famous example of this is obtained in blood oxygenation level dependent functional MRI (**BOLD-fMRI**) experiments, which measure hemodynamic changes in the brain in response to an external "task" or "stimulus". For example, a particular "input" is turned on - say, a particular flashing light – and an "MRI signal" is measured from a region of interest - say, the visual cortex. The visual cortex will almost always get activated in response to the flashing lights, meaning we will see a temporal change in the measured signal (the meaning of which we will discuss in-depth later on in the course). This signal usually looks as if it was convolved with a temporal PSF that has a width of several seconds, meaning that BOLD-fMRI has a point spread function a few seconds wide which looks like this:



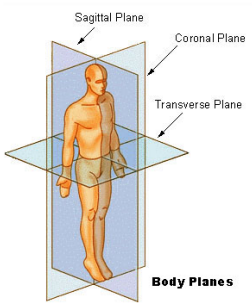
This means that, if we were to measure the response of a brain region to an external stimulus, we could predict it by convoluting the stimulus signal with the PSF:

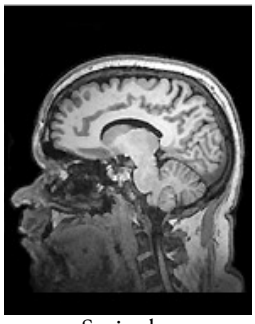


The brain however is not truly a linear response and thinking in terms of point spread functions and simple convolutions fails when tasks are spaced too closely together.

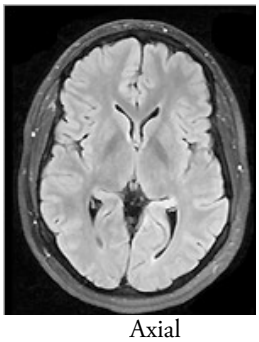
Image Orientation Names

The human body is three dimensional, yet images are usually displayed in 2D in some plane through the body. These planes are usually taken to be either **sagittal**, **axial** or **coronal**:

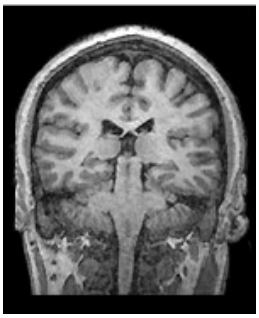




Sagittal



Axial



Coronal