functional Magnetic Resonance Imaging – Methods

Denis Schluppeck

Visual Neuroscience Group
University of Nottingham, UK

4/4

This lecture

1. Spatial and temporal properties of fMRI
   (+ linearity, convolution)
2. Signal and Noise
   (+ Fourier domain, convolution)
3. Preprocessing of fMRI data
   (+ common software tools, registration)
4. Statistics + experimental design
   (+ linear regression, GLM,
   multiple comparisons)

Matlab

```
>> if ~geek
```

http://tinyurl.com/5kcvqv – ***links to tutorials***

```
>> if (geek || keen || phd==fmri )
```

http://web.mit.edu/18.06/www/Course-Info/Tcodes.html

```
help Tcodes; help project; help lsq;
```

Quick recap: *preprocessing [+ spatial filtering]*

Many software packages...

AFNI * | SPM * | FSL *

FreeSurfer | *caret5* | SurfRelax

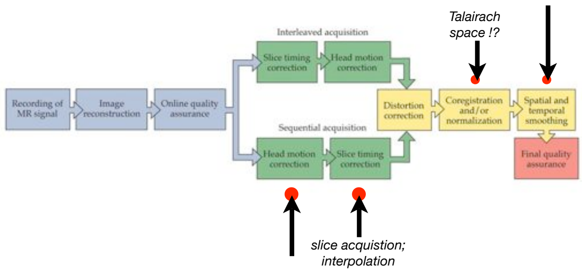
BrainVoyager | Octave | Matlab

C/C++
GNU Scientific Library

SciPy.org
Python

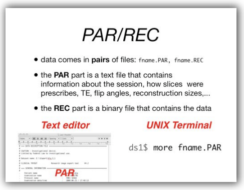
mrTools
Stanford / NYU tools

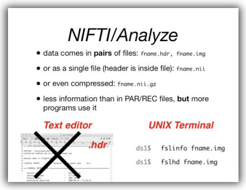
Data Preprocessing



File formats...

two file formats used here: PAR/REC, NIFTI/Analyze

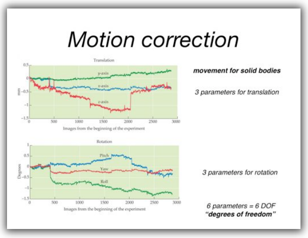




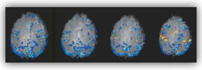
on scanner... **...most tools use this**

`ds1$ ptoa`

Motion correction



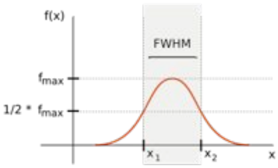
... avoid, rather than deal with!



what motion artefacts tend to look like...

Spatial filtering

... done by convolving image (at each timepoint) with a filter / kernel – often Gaussian



2.3548

$$\text{fwhm} = 2\sqrt{2\ln 2}\,\sigma$$

$$\sigma = \frac{1}{2\sqrt{2\ln 2}}\text{fwhm}$$

see also: *http://en.wikipedia.org/wiki/fwhm*

Filter sizes

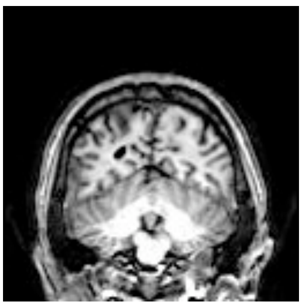
sigma [mm]	fwhm [mm] (~ 2.35 sigma)
1	2.35
3	7.10
5	11.77
0.59	**1**
1.77	**3**
2.94	**5**
5.89	**10**
standard deviation	full-width at half-maximum

2.3548

$$\text{fwhm} = 2\sqrt{2\ln 2}\,\sigma$$

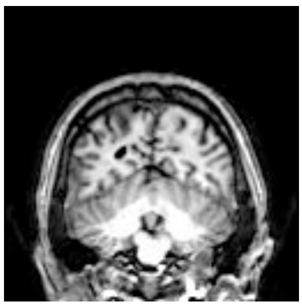
$$\sigma = \frac{1}{2\sqrt{2\ln 2}}\text{fwhm}$$

Spatial filtering



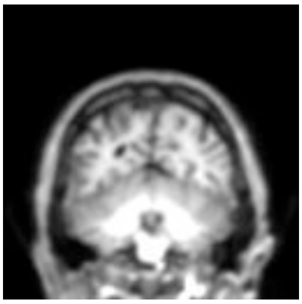
1.5mm inplane
128 x 128 matrix
no filtering

Spatial filtering



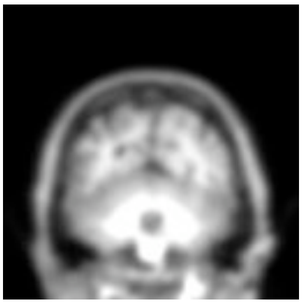
1.5mm inplane
128 x 128 matrix
1mm fwhm gauss

Spatial filtering



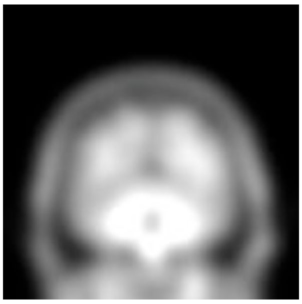
1.5mm inplane
128 x 128 matrix
3mm fwhm gauss

Spatial filtering



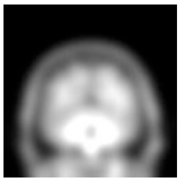
1.5mm inplane
128 x 128 matrix
5mm fwhm gauss

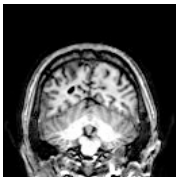
Spatial filtering



1.5mm inplane
128 x 128 matrix
10mm fwhm gauss

Spatial filtering





improves SNR
required for some statistics (Gaussian Random Fields)
increases overlap between subjects
does not preserve edges (blurs in “non-GM tissues”)
combines across sulci (anatomy!)
reduces peak values (e.g. when blurring statistical images)

Experimental design + Statistics

many slides courtesy of D.J. Heeger, NYU

Possible designs

- **Block design:** fixed sequence of different blocks

[A, B, A, B, ...]	*alternating*
[A, rest, B, rest, A, ...]	*alternating with ‘null’*
[A, C, B, A, A, C, D, rest, A, ...]	*randomized*

- **Event-related designs:** different type of ‘trials’ are presented in randomized order

*[A, **r**, B, **r**, B, ...]*	*r={2-5s}*	*rapid, event-related*
*[A, **r**, B, **r**, B, ...]*	*r={12-15s}*	*sparse*

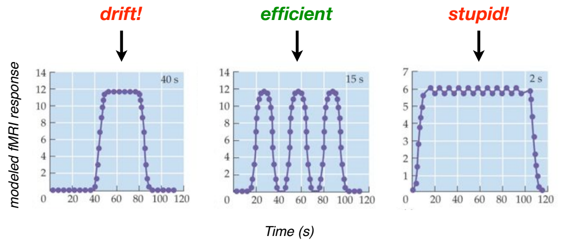
- **Mixed designs:** blocks (states) containing different trials / events.

Which to choose?

TABLE 11.1 Advantages and Disadvantages of Each Type of fMRI Experimental Design

	Advantages	Disadvantages
Blocked	Excellent detection power Useful for examining state changes Simple analysis	Poor estimation power Insensitive to shape of hemodynamic response Potential problems with selection of conditions
Event-related	Good estimation power Allow determination of change from baseline Very flexible analysis strategies Best for post hoc trial sorting	Can have reduced detection power Sensitive to errors in predicted HDR Refractory effects can influence analyses
Mixed or semi-random	Best combination of detection and estimation Can dissociate transient and sustained components of activity	Most complicated analyses Relies on assumptions of linearity

Picking the right timing for block designs...



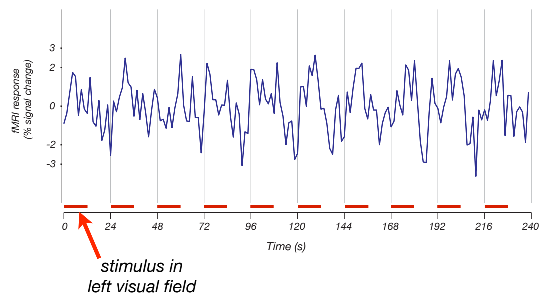
see e.g. Birn, et al. (2002) NeuroImage

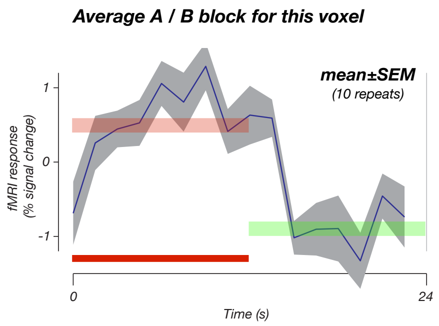
Simple (block) design

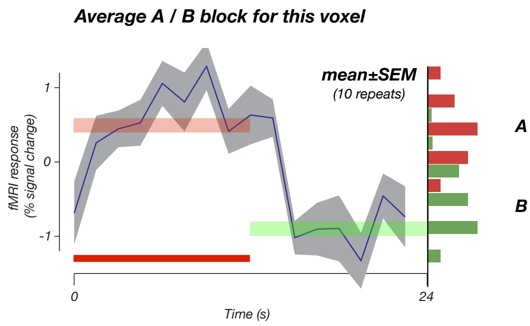
- Example visual experiment:
- Alternate Xs of (**A**) left visual field with Xs of (**B**) right visual field [repeat, say, 10 times]

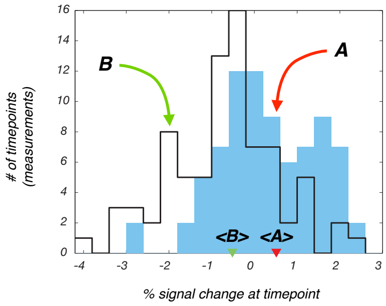
  dots, gratings, movies, ...

Example voxel in visual cortex









...is this a "significant" response?

- *Are the distributions different from each other?*
- at each voxel, calculate a statistic **(e.g. Student's *t*)**
- calculate means for 2 conditions `mean(A), mean(B)`
- and standard error of differences between them
  `sqrt( [var(A)+var(B)]./n ) % n = #samples in each group`

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}$$

difference in means

SE of difference

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2 + s_2^2}{n}}$$

Statistical significance

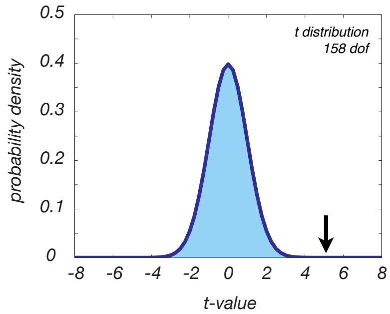
```
mA = mean(A); % => +0.5134
mB = mean(B); % => -0.5145
n = 80; %  # of samples in each group
semAB = sqrt( ( var(A)+var(B) )./80 );        0.1927
t = (mA - mB)./semAB;                         5.330
```

- could we have observed that specific *t* value by chance? (*inference*)
- null hypothesis H_0: *difference in means is 0*
  degrees of freedom: 2n-2

```
p = 1 - tcdf( 5.330, 160-2) % cumulative t
```

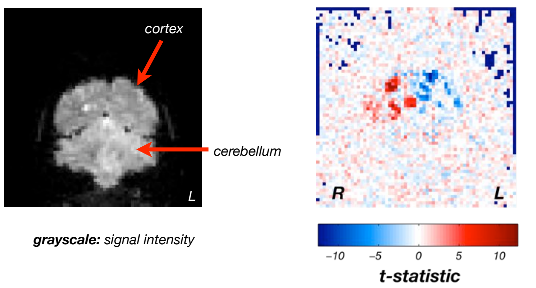
1.7e-7 = 0.00000017 ***i.e. reject*** H_0

... highly unlikely to be due to chance



☣ `1 - tcdf( 5.330, 160-2)` *is area under curve from 5.33 →∞*

1. **re-calculate at each voxel** in the data set to get a statistical parametric map (spm)
   [actual analyses use general linear model / multiple linear regression / non-parametric tests]
2. decide on a scheme for thresholding the statistical image
3. render result (co-registered to anatomy)
   [*optional:* superimpose on surface]
4. ... that's basically it



grayscale: *signal intensity*

higher response in B than A *higher response in A than B*

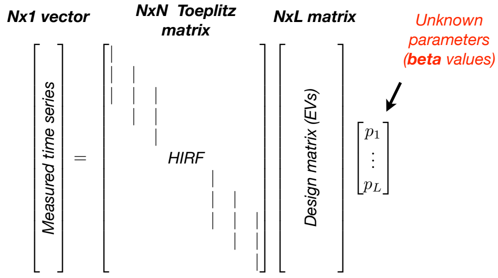
1. **re-calculate at each voxel** in the data set to get a statistical parametric map (spm)
   [actual analyses use **general linear model** / multiple linear regression / non-parametric tests]
2. decide on a scheme for **thresholding** the statistical image ("what is significant")
3. render result (co-registered to anatomy)
   [*optional:* superimpose on surface]
4. … that's basically it

estimation

inference

Estimation

Modeling the fMRI time series



***N:** number of time points in the time series*
***L:** number of regressors in the design matrix*

EV = explanatory variables

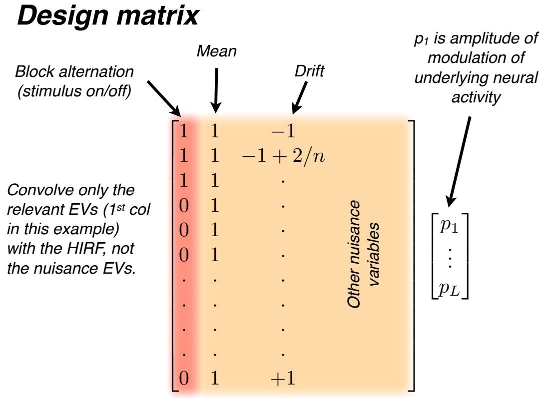
Matrix multiplication

A is a 2 by 2 matrix
$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 2.5 \\ 3.2 \end{bmatrix}$$
x is a vector (2 by 1 matrix)

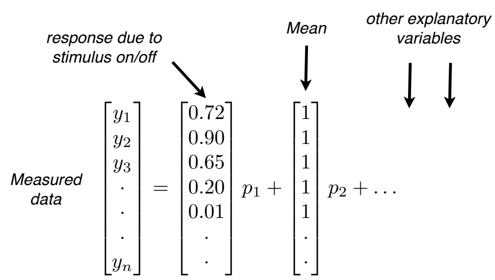
$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} 2.5 + \begin{bmatrix} 0 \\ 2 \end{bmatrix} 3.2 = \begin{bmatrix} 2.5 \\ 6.4 \end{bmatrix}$$

weighted sum of columns … $ax_1 + bx_2$

Design matrix



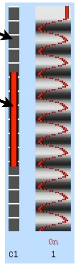
data = linear combination of effects



In FSL / SPM

time: 10s ticks

indicates scale of high-pass filter



design matrix is often represented as an image

low numbers: dark
high numbers: light

"regressors"

example: *12s on, 12s off*
TR 1.5s, 10 cycles

General Linear Model

Nx1 vector ***NxL matrix*** ***Lx1 vector***

*Measured data (**y**)*

$$\left[\text{Measured time series}\right] = \left[\begin{array}{c}\text{known matrix} \\ \text{with more rows} \\ \text{than columns} \\ (\boldsymbol{X})\end{array}\right] \begin{bmatrix} p_1 \\ \vdots \\ p_L \end{bmatrix}$$

matrix form

$$\mathbf{y} = \mathbf{X}\mathbf{p}$$

N: *number of time points in the time series*
L: *number of regressors in the design matrix*

How to solve for *p*?

$$\mathbf{y} = \mathbf{X}\mathbf{p}$$

*Unlikely to find **exact** solution, because we have more equations than unknowns.*

$$\mathbf{p}_{\text{opt}} = \mathbf{X}^{\#}\mathbf{y}$$

*... where $\mathbf{p}_{opt}$ are the (best) parameter estimates and **#** means pseudoinverse.*

One parameter example

Nx1 vector ***NxN Toeplitz*** ***Nx1 design***

$$\left[\text{Measured time series}\right] = \left[\text{HIRF}\right] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix} [p]$$

unknown parameter

remember, this is convolution...

One parameter example

Nx1 ***Nx1***

$$\left[\text{Measured time series}\right] = \left[\text{HIRF * block alternation}\right] [p]$$

unknown parameter

Solve...

$$\vec{y} = \vec{x}p$$

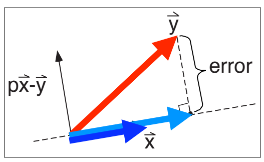
*... where **x, y** are vectors and p is a number.*

Least-squares regression (for ref)

Solution 3 (using the orthogonality principle). The error vector for the best p is perpendicular to x:

$$\vec{x} \cdot \vec{e} = 0$$
$$\vec{x} \cdot (p\vec{x} - \vec{y}) = 0.$$

measured data
'direction of model'
best fit (scaled by p)



Multiple parameters

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\mathbf{p} \\ \mathbf{X}^{\mathbf{T}}\mathbf{y} &= \mathbf{X}^{\mathbf{T}}\mathbf{X}\mathbf{p} \\ (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y} &= \mathbf{p}_{\text{opt}} \end{aligned}$$

projection matrix

*... where **y, p** are vectors and **X** is the known matrix. and $[\bullet]^T$ is transpose and $[\bullet]^{-1}$ is the matrix inverse.*

pseudoinverse (X)

```
>> p = pinv(X)*y
>> p = X \ y
```

! Orthogonality

$$(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}$$

- *Make sure the design matrix makes sense!*
- *Is X^TX always invertible? If not, why not?*
- *What is the interpretation for the values corresponding to each element of p_{opt}? Is the meaning of each value independent of the other elements?*

Simple example

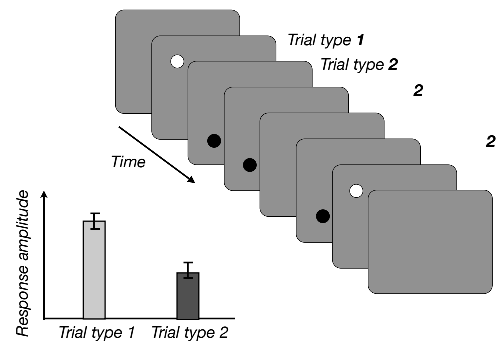
$$\mathbf{y} = \mathbf{X}\mathbf{p}$$

$$\begin{bmatrix} 3 \\ 2 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \leftrightarrow \begin{vmatrix} 3 \\ 2 \\ 0.1 \end{vmatrix} = \begin{vmatrix} 1 \\ 0 \\ 0 \end{vmatrix} p_1 + \begin{vmatrix} 0 \\ 1 \\ 0 \end{vmatrix} p_2$$

$$\mathbf{X}^{\mathbf{T}}\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 0.1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

Event-related fMRI experiment



Event-related design matrix

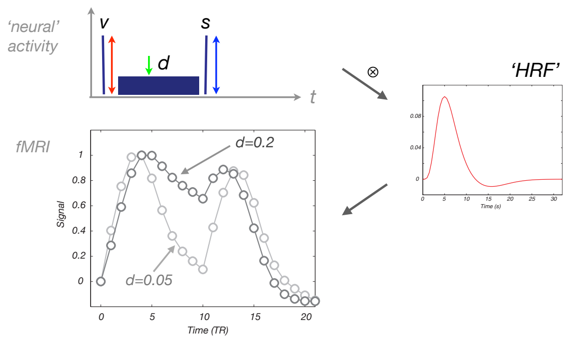
Time of each type 1 trial

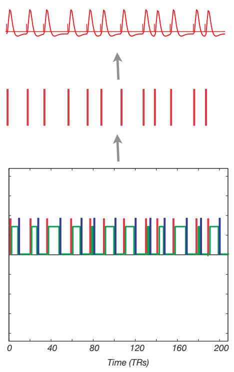
Trial type 2

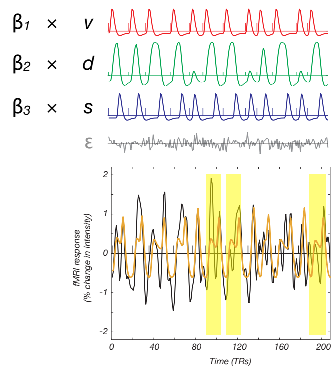
$$\begin{bmatrix} 1 & 0 & \\ 0 & 0 & \\ 0 & 1 & \\ 1 & 0 & \\ 0 & 1 & \\ 0 & 0 & \cdots \\ \vdots & \vdots & \\ 0 & 1 & \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_L \end{bmatrix}$$

p_1 is 'neural' response to trial type 1

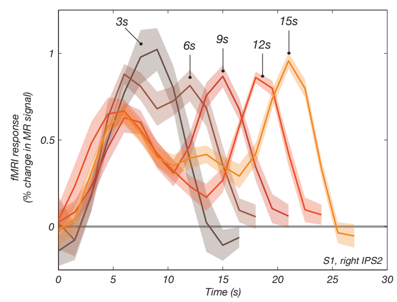
Estimating Delay Period Activity

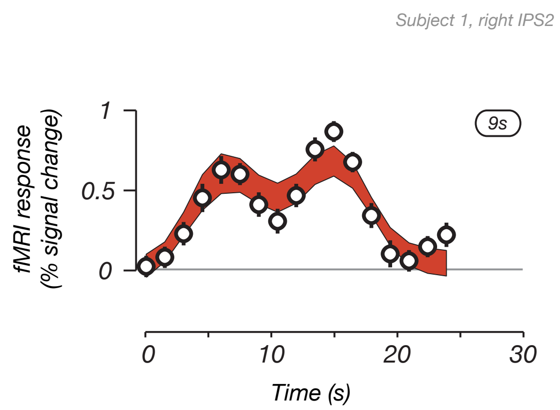


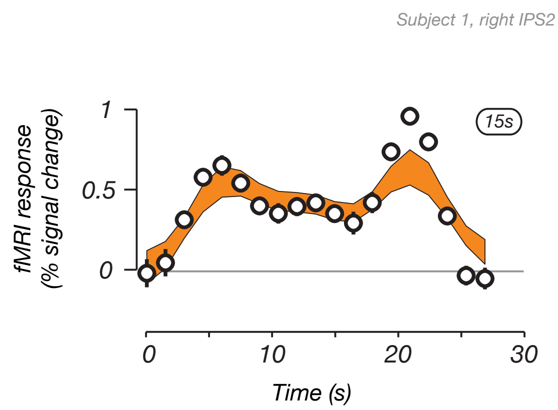




Measured fMRI response

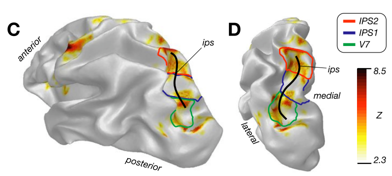






Usually...

1. this is looking at time course data from (functionally) predefined areas
2. ... could also show a map of the parameter estimates (cf. block-design data)



Event-related analysis (2)

- **can also estimate** stimulus-locked (trial-triggered) responses for different event types **without making assumptions about the HIRF**
- similar matrix-algebra magic
- Not enough to cover this here, but if you are interested, let me know...

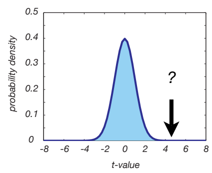
Inference

Statistics

1. *How well does the model fit the data?*
2. *What are the confidence intervals/error bars on the parameter estimates?*
3. *Are the parameter estimates different from zero? Different from each other?*
4. *Which of the regressors contribute to fitting the data?*

GLM: calculate t

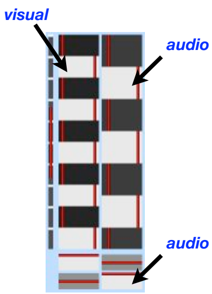
- ... the ratio of **mean** to **standard error** of our parameter estimates (p, or beta) = t
- **error:** look at residuals
- **do inference on these t values:** is the t observed at this voxel large enough?



GLM: contrasts

- calculate individual statistical maps
- **compare:** by contrasting them
- **e.g.**
  [1 0] for 'responds to visual'
  [0 1] for 'responds to audio'

  [1 -1] for 'responds more to visual than audio'



FSL: t →p → Z

for a v clear explanation:
http://www.fmrib.ox.ac.uk/fsl/feat5/glm.html

"Thresholding"

- either **voxelwise** or taking into account the fact that voxels are not independent (clustering, GRF).
- voxelwise: **corrected** versus **uncorrected?**
- Because we are computing many statistical tests, we will get many false positives ***tens of thousands!***

  This is called the **multiple comparisons problems**

Kinds of error...

data say...

... we conclude		signal present	signal absent	
	YES	*hit*	*false positive*	← *controlled by* ***alpha e.g. p<0.05***
	NO	*miss*	*correct reject*	

Kinds of error...

data say...

... we conclude		signal present	signal absent	
	YES	*hit*	*false positive*	← *controlled by* ***alpha e.g. p<0.05***
	NO	*miss*	*correct reject*	

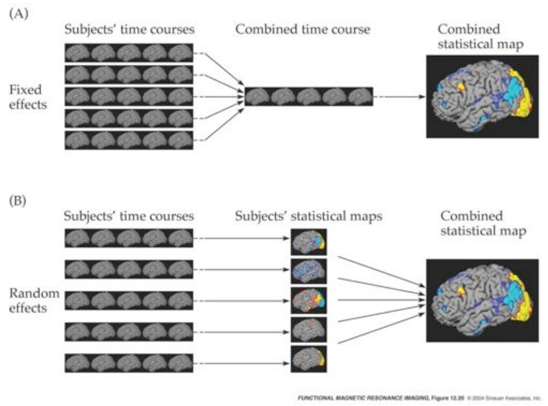
With 10,000 tests, we may have 500 false positives

Corrections for multiple comparisons

- **Bonferroni:** divide alpha by number of tests... 0.05 (5e-2) becomes 0.000005 (5e-6) with 10,000 tests .... very conservative.
- **Resel:** resolution elements. After smoothing, roughly the number of independent elements in data set (use this instead of voxels)
- **Gaussian Random Field theory**

Multi-subject analysis

- **Normalize brains anatomically:** affine, e.g. Talairach, MNI, or non-rigid transformations...
- **Fixed-effects analysis:** assume brains are "the same" across subjects (more sensitive)
- **Random-effects:** allow between-subject variability in pattern of responses as another factor in your analysis (more conservative, but more likely to be "true")

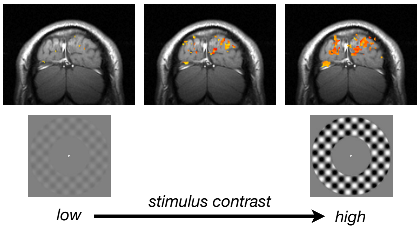


Some thoughts...

Beware of...

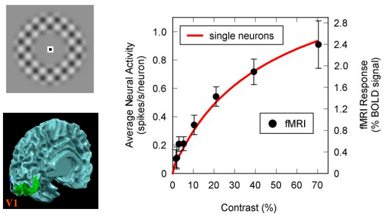
- ... statistical thresholds
- ... circular reasoning
- ... finding the "cortical locus for cognitive ability X"

Beware of statistical thresholds



Alternative

...plot parameter estimates with error bars



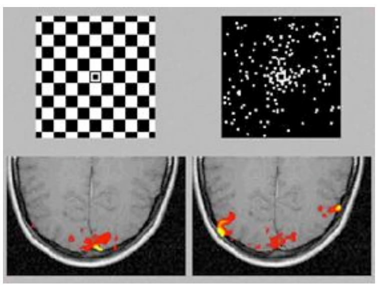
Heeger et al, Nature Neurosci (2000)

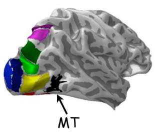
Circular reasoning

1. ***Hypothesis:*** *that there is a cognitive process called groking that is localized to a functionally specialized brain area.*
2. ***Design*** *an experiment with two tasks, one of which you believe imposes a greater load on groking.*
3. ***Run the experiment*** *and find sure enough that there is a brain area that responds more strongly during high grok load trials then low load trials.*

What can you conclude from this?

Case study: visual area MT/ V5 and motion perception





Geoff Boynton
Alex Huk

Cortical area MT is specialized for visual motion perception

- *Neurons in MT are* ***selective*** *for motion direction.*
- *Neural responses in MT are* ***correlated*** *with the perception of motion.*
- *Damage to MT or temporary inactivation* ***causes*** *deficits in visual motion perception.*
- *Electrical stimulation in MT* ***causes*** *changes in visual motion perception (Newsome).*
- *Computational* ***theory quantitatively explains*** *both the responses of MT neurons and the perception of visual motion.*
- *Well-defined* ***pathway*** *of brain areas (cascade of neural computations) underlying motion specialization in MT.*

Additional Resources

- The Oxford FMRI book
- FSL course (Oxford, all lecture materials online) **http://www.fmrib.ox.ac.uk/fslcourse/**
- SPM book: **http://www.fil.ion.ucl.ac.uk/spm/doc/books/hbf2/**
- Random Field Theory tutorial MRC-CBU (Cambridge) **http://imaging.mrc-cbu.cam.ac.uk/imaging/PrinciplesRandomFields**
- Additional slides [will be on website]...

Thanks + see you next term!

Least-squares regression (for ref)

*Find **p** to make (**x_n p**) as close as possible to y_n for all n.*
*That is, choose **p** to minimize:*

$$\min_p \sum_{n=1}^{N} (y_n - p x_n)^2$$

Or, in vector notation:

$$\min_p \|\vec{y} - p\vec{x}\|^2$$

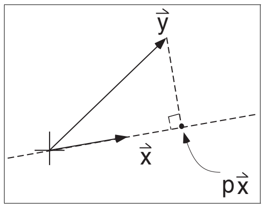
***Solution 1 (using calculus).** Take the derivative of the above expression, set it equal to zero, and solve for p:*

$$p_{\text{opt}} = \frac{\vec{y}^T \vec{x}}{\vec{x}^T \vec{x}}.$$

Least-squares regression (for ref)

***Solution 2 (using geometry).** Find the scale factor **p** such that the scaled vector **p x** is as close as possible (in Euclidean distance) to **y**. Geometrically, we know that the scaled vector should be the projection of y onto the line in the direction of x:*

$$p\vec{x} = (\vec{y} \cdot \hat{x})\hat{x} = \frac{(\vec{y} \cdot \vec{x})}{\|\vec{x}\|^2}\vec{x}$$



Least-squares regression (for ref)

Solution 3 (using the orthogonality principle). The error vector for the best p is perpendicular to x:

$$\vec{x} \cdot (p\vec{x} - \vec{y}) = 0.$$