DISCOS SPM course, CRC, Liège, 2009

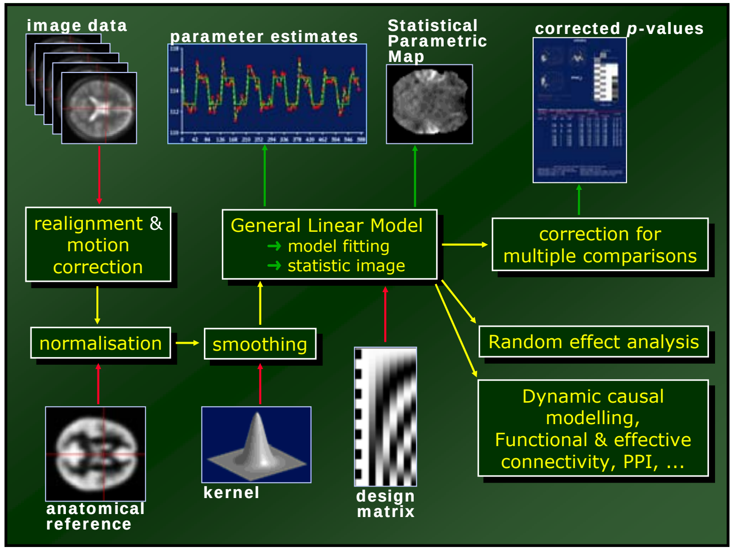
The General Linear Model, Part II

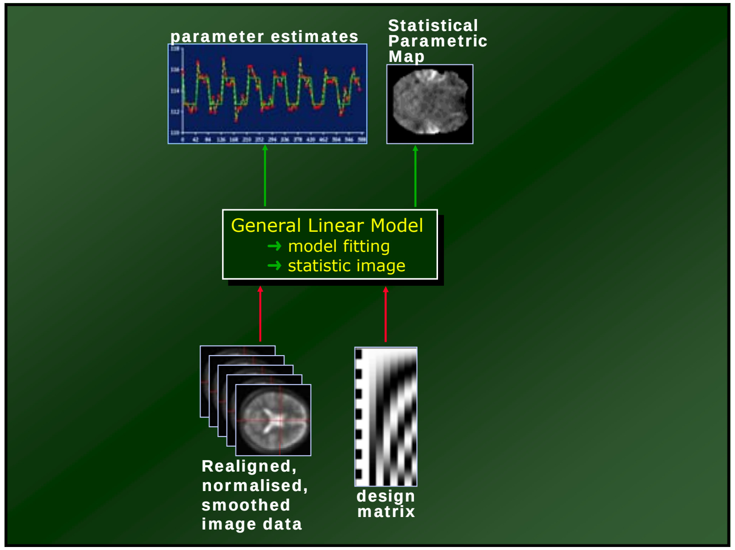
C. Phillips, Centre de Recherches du Cyclotron, ULg, Belgium

Based on slides from: T. Nichols, R. Henson, S. Kiebel, JB. Poline

Contents

- Introduction & recap
  - The General Linear Model
  - T test
- F-test and added variance
- Good & bad models
- Improved model
- HRF and ER fMRI
- « Take home » message





General Linear Model

What does it mean ?

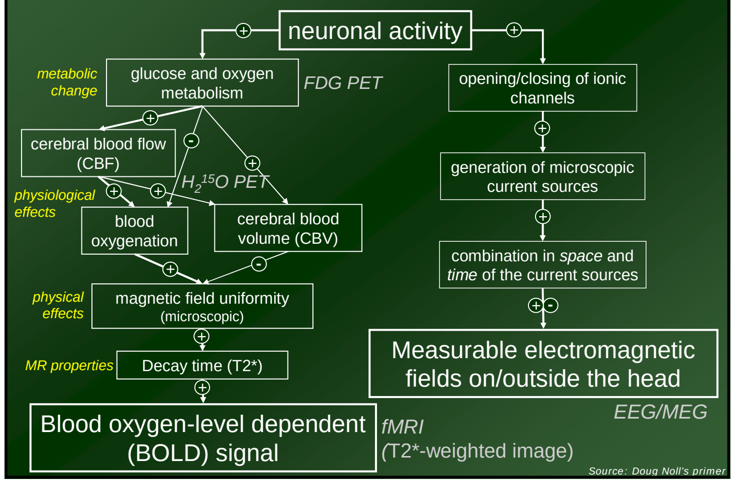
General
The model can be used to answer a wide variety of questions.

Linear
The model uses simple linear relationships between the variables.

Model
A set of equations are used to describe the data. Questions about the data can then be stated as mathematical expressions.

GLM is the basic model or general framework underlying the analysis of variance and multiple regression.

Functional neuroimaging signal



Source: Doug Noll's primer

Simple *f*MRI example dataset: can we do better ?

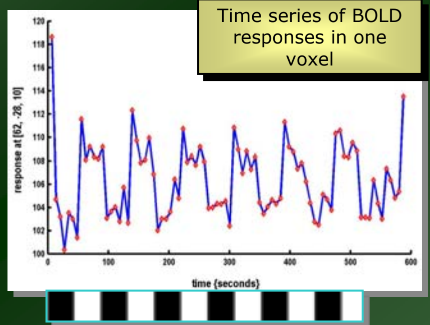
One session, one subject

Passive word listening versus rest

7 cycles of rest and listening

Each epoch 6 scans with 7 sec TR

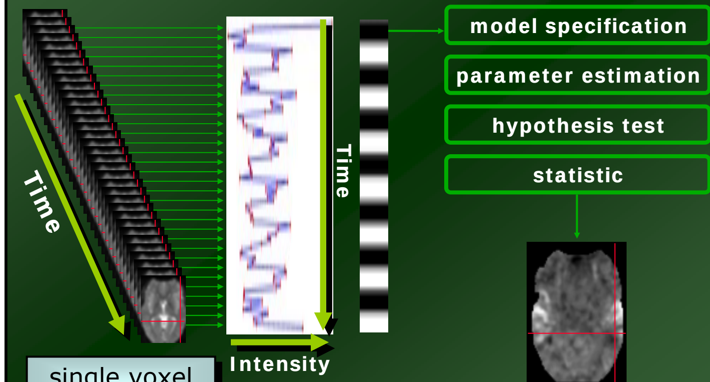
Time series of BOLD responses in one voxel



Stimulus function

Question: Is there a change in the BOLD response between listening and rest?

Voxel by voxel statistics…

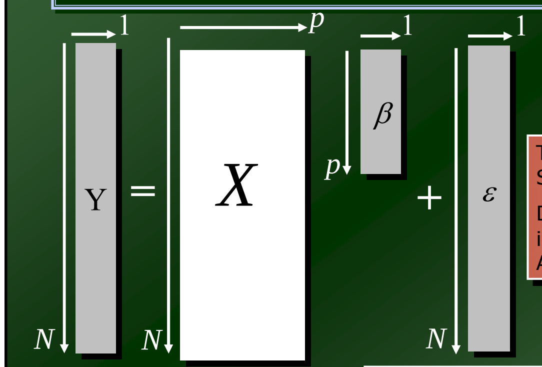


single voxel time series

statistic image *or* SPM

General Linear Model

$$Y = X\beta + \varepsilon$$



This is for a SINGLE voxel !

Design matrix **X** is the same for ALL voxels !

N: number of scans
p: number of regressors

Model is specified by
1. Design matrix **X**
2. Assumptions about ε

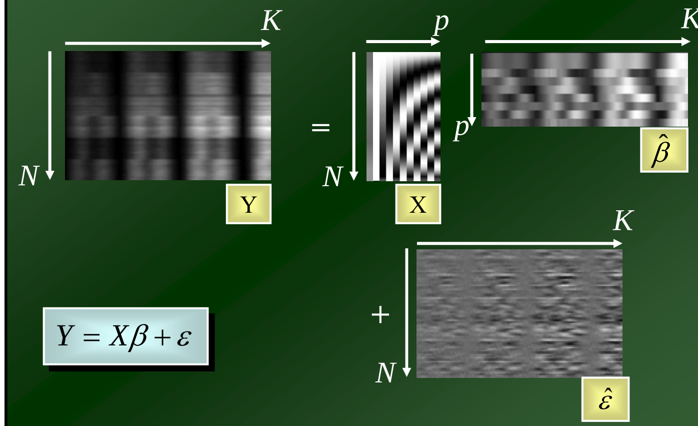
General Linear Model

- ***f*MRI time series:** $Y_1, \ldots, Y_s, \ldots, Y_N$
  - acquired at times $t_1, \ldots, t_s, \ldots, t_N$
- **Model:** Linear combination of basis functions

$$Y_s = \beta_1 f^1(t_s) + \ldots + \beta_l f^l(t_s) + \ldots + \beta_L f^L(t_s) + \varepsilon_s$$

- $f^l(.)$: ***basis*** **functions**
  - ***"reference waveforms"***
  - ***dummy variables***
- β_l: ***parameters*** *(fixed effects)*
  - amplitudes of basis functions *(regression slopes)*
- ε_s: ***residual errors***: $\varepsilon_s \sim N(0, \sigma^2)$
  - identically distributed
  - independent, or serially correlated
    (Generalised Linear Model → GLM)

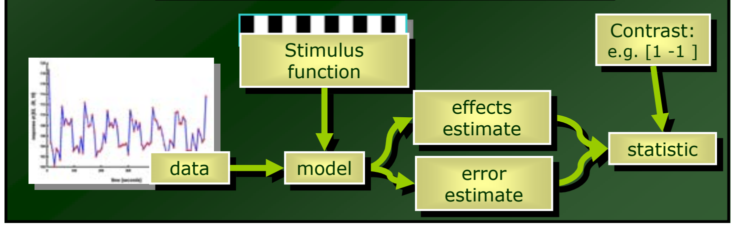
Mass univariate approach



$$Y = X\beta + \varepsilon$$

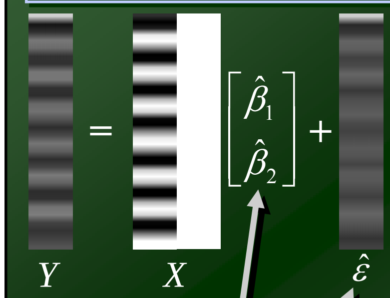
Why modelling?

Why?	Make inferences about effects of interest
How?	1. Decompose data into effects and error 2. Form statistic using estimates of effects and error
Model?	Use any available knowledge



Parameter estimation

$Y = X\beta + \varepsilon$



$$Y = X \begin{bmatrix} \hat\beta_1 \\ \hat\beta_2 \end{bmatrix} + \hat\varepsilon$$

$\hat\varepsilon = Y - X\hat\beta$

residuals

Assume iid error

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

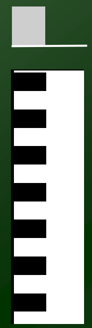
Estimate parameters

such that $\sum_{t=1}^{N} \hat\varepsilon_t^2$ minimal → **Least squares** parameter estimate

Model — contrasts — SPM{t}

Contrast **: specifies linear combination of parameter vector:** $c'\beta$

$c' = +1 \quad 0$



box-car amplitude > 0 ?

=

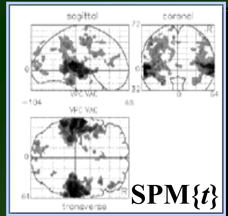
$\hat\beta_1 > 0$? ($\hat\beta_1$: estimation of β_1)

=

$1 \times \hat\beta_1 + 0 \times \hat\beta_2 > 0$?

=

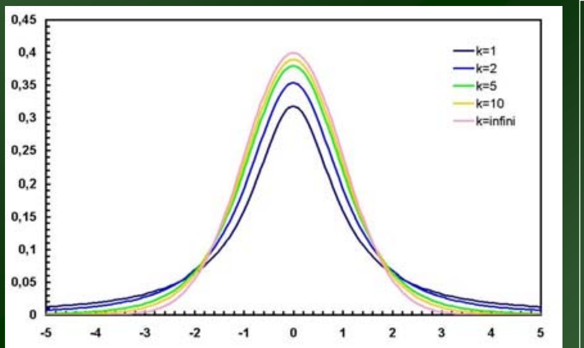
test H_0 : $c' \times \hat\beta > 0$?



$$T = \frac{\textit{contrast of estimated parameters}}{\sqrt{\textit{variance estimate}}} \rightarrow T = \frac{c'\hat\beta}{\sqrt{s^2 c'(X'X)^{+} c}}$$

t distribution

Depends on 1 set of degrees of freedom !



ν	95%
1	6.314
5	2.015
10	1.812
15	1.753
20	1.725
30	1.697
50	1.676
100	1.660
Inf	1.645

How is this computed ? (t-test)

Estimation [Y, X] [b, s]

$Y = X\beta + \varepsilon$ $\quad \varepsilon \sim \sigma^2 N(0,I)$ *(Y : at one position)*

$b = (X'X)^{+} X'Y$ *(b = estimation of β)* → *beta??? images*

$e = Y - Xb$ *(e = estimation of ε)*

$s^2 = (e'e/(n - p))$ *(s = estimation of σ, n: scans, p: parametres)* → *1 image ResMS*

Test [b, s², c] [c'b, t]

$\mathrm{Var}(c'b) = s^2 c'(X'X)^{+} c$ *compute for each contrast c*

$t = c'b / \mathrm{sqrt}(s^2 c'(X'X)^{+} c)$ *c'b → images con???*

compute the t images → images spm_t???

under the null hypothesis $H_0 : t \sim \mathit{Student}(df)$ $\quad df = n\text{-}p$

Contents

- Introduction & recap
- F-test and added variance
- Good & bad models
- Improved model
- HRF and ER fMRI
- « Take home » message

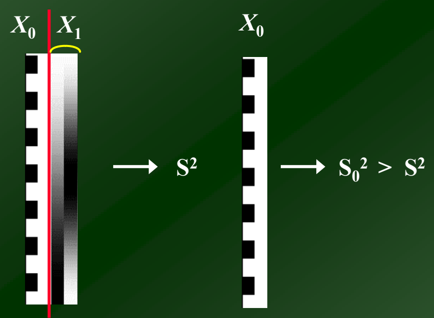
Model — contrasts — SPM{F}

Tests multiple linear hypotheses :

Does X_1 model anything ?

H_0: True model is X_0

X_0 X_1 X_0



This model ? Or this one ?

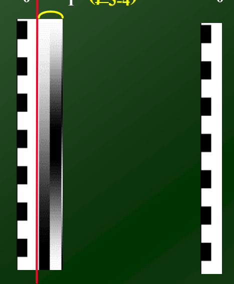
F = **additional variance accounted for by tested effects** / **error variance estimate**

$F \sim (S_0^2 - S^2) / S^2$

F test (SPM{F}) : a reduced model or ... multi-dimensional contrasts ?

Tests multiple linear hypotheses.

X_0 X_1 $(\beta_{3\text{-}4})$ X_0



This model ? Or this one ?

H_0: True model is X_0

$\downarrow$

H_0: $\hat{\beta}_{3\text{-}4} = (0\ 0)$

$\downarrow$

test H_0 : $c' \times \hat{\beta} = 0$?

$$c' = \begin{matrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}$$

How is this computed ? (F-test)

Estimation [Y, X] [b, s]

$Y = X\beta + \varepsilon$ $\qquad \varepsilon \sim N(0, \sigma^2 I)$

$Y = X_0 \beta_0 + \varepsilon_0$ $\qquad \varepsilon_0 \sim N(0, \sigma_0^2 I)$ $\qquad X_0$: X Reduced

Estimation $[Y, X_0]$ $[b_0, s_0]$ (not really like that)

$b_0 = (X_0'X_0)^+ X_0'Y$

$e_0 = Y - X_0 b_0$ $\qquad$ (e_0 = estimation of ε_0)

$s^2_0 = (e_0'e_0/(n - p_0))$ $\qquad$ (s_0 = estimation of σ_0, n: # time bins, p_0: # parameters)

Test [b, s, c] [ess, F]

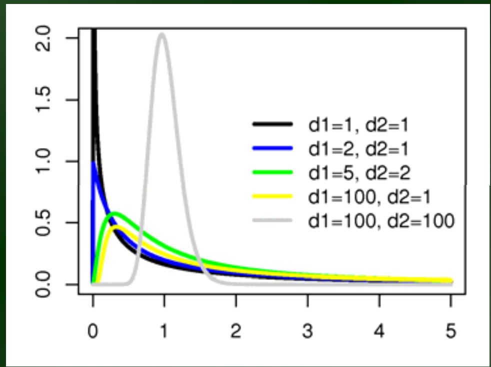
$F = (e_0'e_0 - e'e)/(p - p_0) / s^2$ $\rightarrow$ image $(e_0'e_0 - e'e)/(p - p_0)$: spm_ess???

$\rightarrow$ image of F : spm_F???

under the null hypothesis : $F \sim F(df1, df2)$, with $df1 = p - p_0$, $df2 = n - p$

F distribution

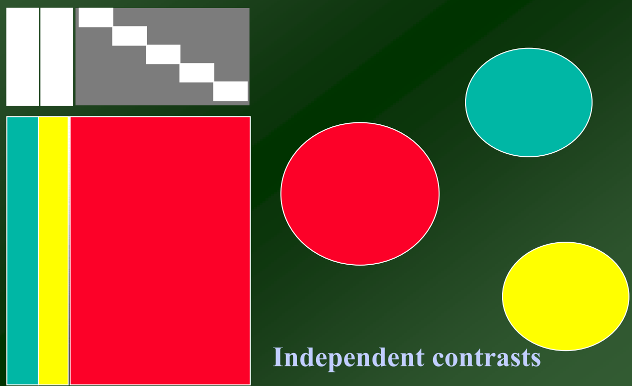
Depends on 2 sets of degrees of freedom !



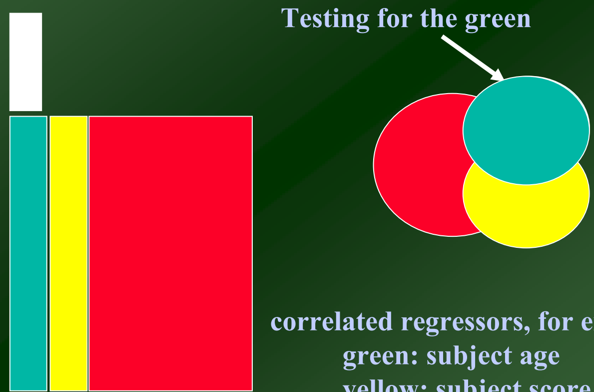
T and F test: take home ...

- *T tests are simple combinations of the betas; they are either positive or negative (b1 – b2 is different from b2 – b1)*
- *F tests can be viewed as testing for the additional variance explained by a larger model wrt. a simpler model, or*
- *F test the sum of the squares of one or several combinations of the betas*
- *in testing "single contrast" with an F test, for ex. b1 – b2, the result will be the same as testing b2 – b1. It will be exactly the square of the t-test, testing for both positive and negative effects, and the p-value will be twice as big.*

« Additional variance » : Again



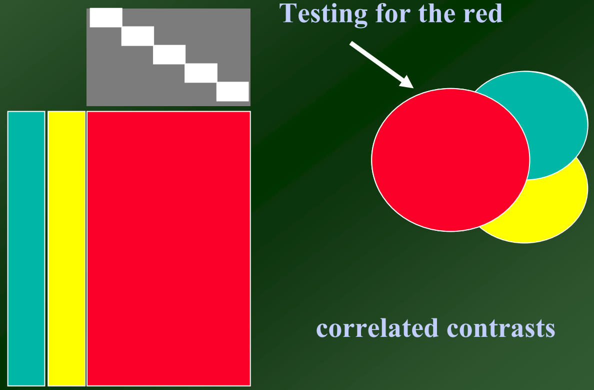
« Additional variance » : Again



correlated regressors, for examp
green: subject age
yellow: subject score

« Additional variance » : Again

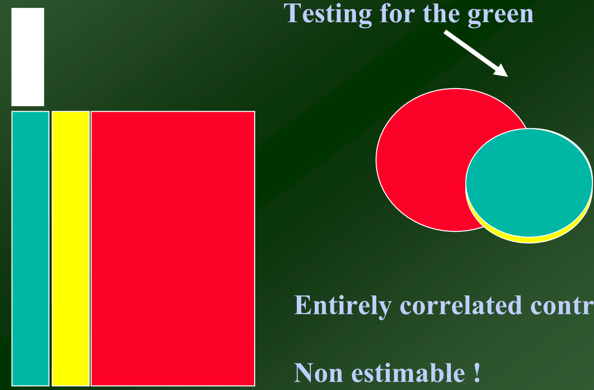
Testing for the red



correlated contrasts

« Additional variance » : Again

Testing for the green

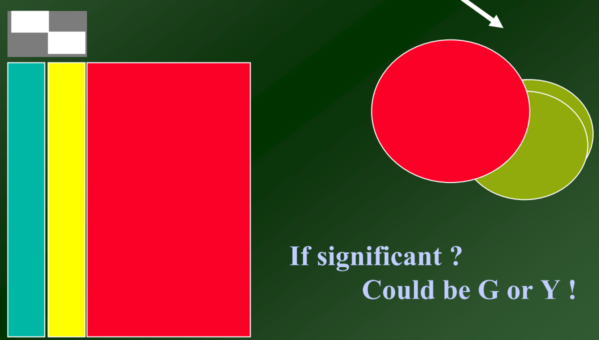


Entirely correlated contrasts ?

Non estimable !

« Additional variance » : Again

Testing for the green and yellov

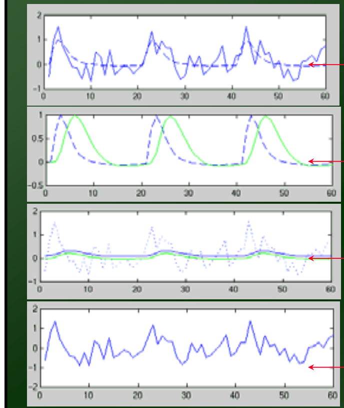


If significant ?
Could be G or Y !

Contents

- Introduction & recap
- F-test and added variance
- Good & bad models
- Improved model
- HRF and ER fMRI
- « Take home » message

A bad model ...



True signal (---) and observed signal

Model (green, peak at 6sec) and TRUE signal (blue, peak at 3sec)

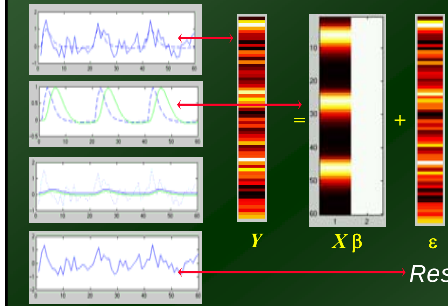
Fitting :
b1 = 0.2, mean = .11

Noise
(still contains some signal)

⇒ Test for the green regressor not significant

A bad model ...

$b_1 = 0.22$
$b_2 = 0.11$

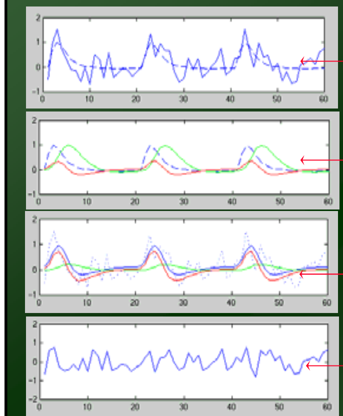


P(b1 = 0) = 0.1
(t-test b1>0)

P(b1 = 0) = 0.2
(F-test b1≠0)

Residual Variance = 0.3

A « better » model ...



True signal + observed signal

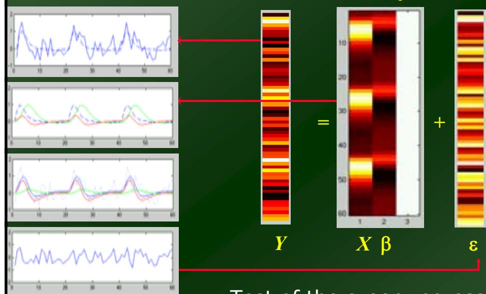
Model (green and red) and true signal (blue ---)
Red regressor : temporal derivative of the green regressor

Global fit (blue) and partial fit (green & red)
Adjusted and fitted signal

Noise (a smaller variance)

A better model ...

$b_1 = 0.22$
$b_2 = 2.15$
$b_3 = 0.11$



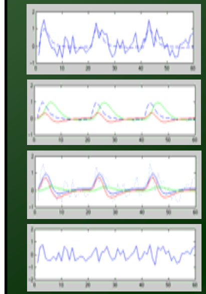
Residual Var = 0.2

P(b1 = 0) = 0.07
(t test b1>0)

P([b1 b2] = [0 0]) = 0.000001
(F test [b1 b2]≠[0 0])

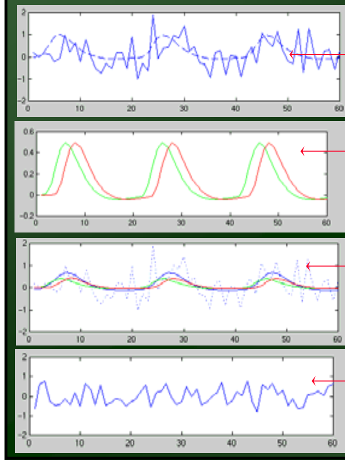
⇒ Test of the green regressor almost significant
⇒ Test F very significant
⇒ Test of the red regressor very significant

Summary ...



- *The residuals should be looked at ...(non random structure ?)*
- *We rather test flexible models if there is little a priori information, and precise ones with a lot a priori information*
- *In general, use the F-tests to look for an overall effect, then look at the betas or the adjusted signal to characterise the origin of the signal*
- *Interpreting the test on a single parameter (one function) can be very confusing: cf the delay or magnitude situation*

Correlation between regressors



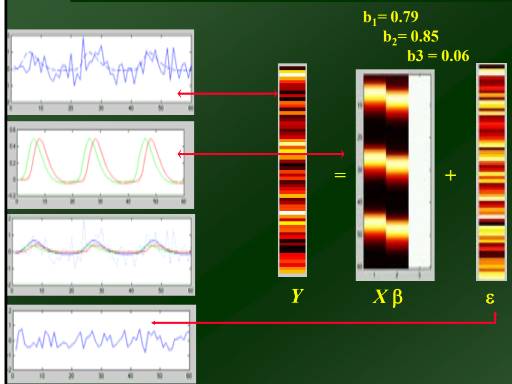
True signal

Model (green and red)

Fitting (blue : global fit)

Noise

Correlation between regressors



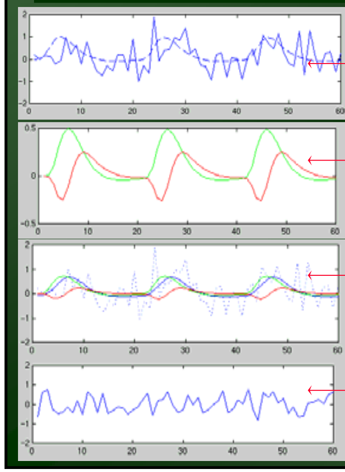
Residual var. = 0.2

P(b1 = 0) = 0.08
(t test b1>0)

P(b2 = 0) = 0.07
(t test b2>0)

P([b1 b2] = 0) = 0.002
(F test [b1 b2] ≠ 0)

Correlation between regressors - 2



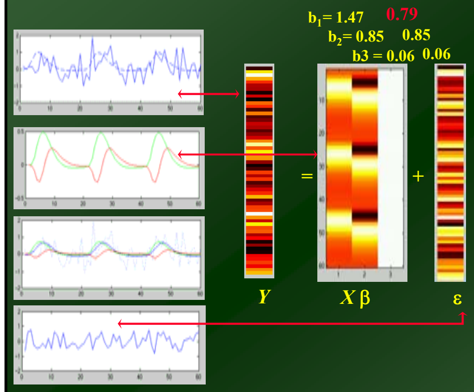
true signal

Model : red regressor orthogonalised with respect to the green one = remove every thing that can correlate with the green regressor

Fit

Noise

Correlation between regressors -2



Residual var. = 0.2

P(b1 = 0) = 0.0003
(t test b1>0)

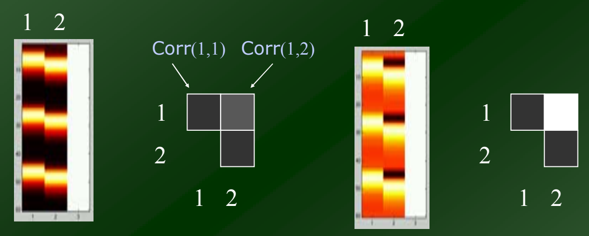
P(b2 = 0) = 0.07
(t test b2>0)

P([b1 b2] = 0) = 0.002
(F test [b1 b2] ≠ 0)

Design orthogonality : « explore design »

Black = completely correlated

White = completely orthogonal



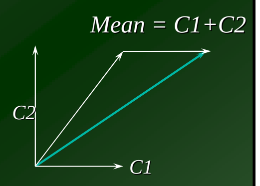
Beware : when there is more than 2 regressors (C1,C2,C3...), you may think that there is little correlation (light grey) between them, but C1 + C2 + C3 may be correlated with C4 + C5

"completely" correlated ...

$$Y = Xb + e$$

$$X = \begin{matrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{matrix}$$

Cond 1 Cond 2 Mean



Parameters are not unique in general !
Some contrasts have no meaning: NON ESTIMABLE

Example here :

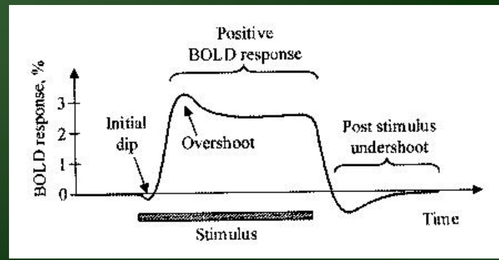
- c' = [1 0 0] is not estimable
  (= no specific information in the first regressor);
- c' = [1 -1 0] is estimable.

Summary ...

- *We are implicitly testing additional effect only, so we may miss the signal if there is some correlation in the model using t tests*
- *Orthogonalisation is not generally needed - parameters and test on the changed regressor don't change*
- *It is always simpler (when possible !) to have orthogonal (uncorrelated) regressors*
- *In case of correlation, use F-tests to see the overall significance. There is generally no way to decide where the « common » part shared by two regressors should be attributed to*
- *In case of correlation and you need to orthogonolise a part of the design matrix, there is no need to re-fit a new model : the contrast only should change.*

Contents

Hemodynamic Response Function



% signal change
= (point – baseline)/baseline
usually 0.5-3%

initial dip
-more focal and potentially a better measure
-somewhat elusive so far, not everyone can find it

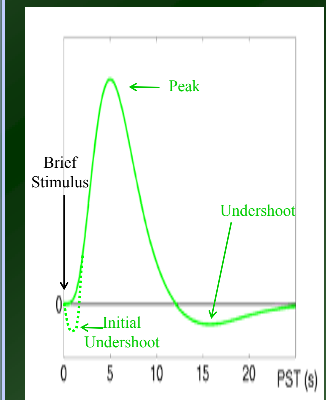
time to rise
signal begins to rise soon after stimulus start

time to peak
signal peaks 4-6 sec after stimulus begins

post stimulus undershoot
signal suppressed after stimulation ends

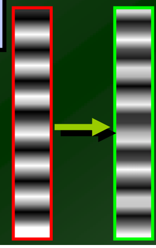
Haemodynamic response function

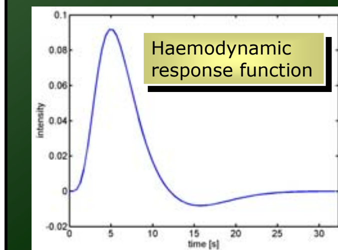
- Function of blood oxygenation, flow, volume (Buxton et al, 1998)
- Peak (max. oxygenation) 4-6s poststimulus; baseline after 20-30s
- Initial undershoot can be observed (Malonek & Grinvald, 1996)
- Similar across V1, A1, S1…
- … but differences across: other regions (Schacter et al 1997) individuals (Aguirre et al, 1998)

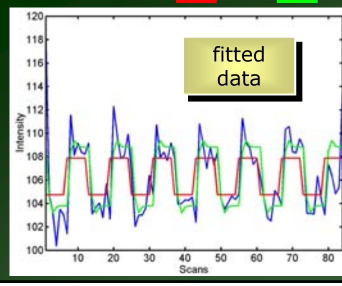


Improved model

Convolve stimulus function with model of BOLD response

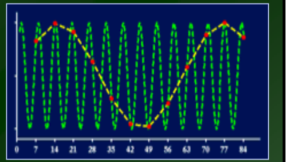


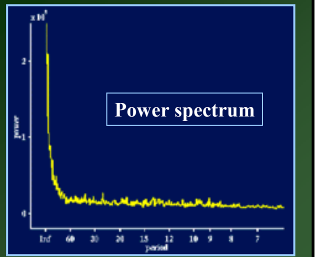




Low frequency nuisance effects

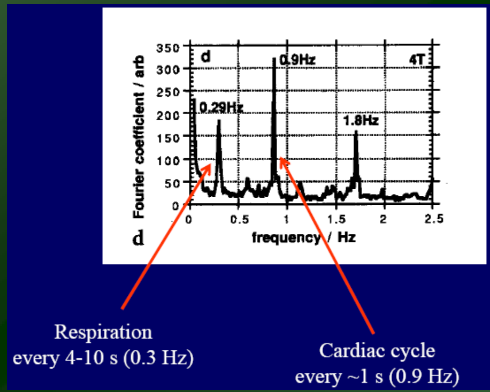
- Drifts
  - physical
  - physiological
- Aliased high frequency effects
  - cardiac (~1 Hz)
  - respiratory (~0.25 Hz)





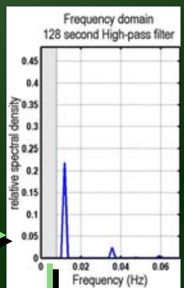
⇒ Power in the low frequencies

Physiological "noise"



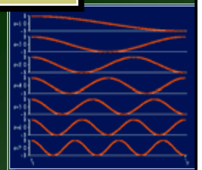
High pass filter

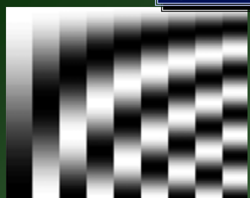
$$Y = X\beta + \varepsilon$$



high pass filter implemented by residuals of DCT set

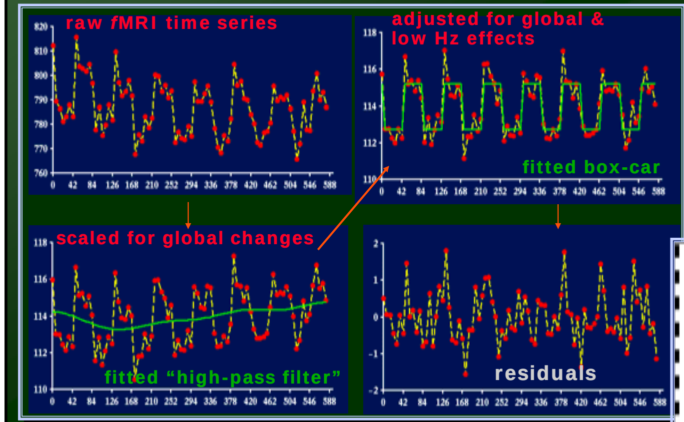
$$f_r(t) = \cos\left(r\pi \frac{t - t_1}{t_N - t_1}\right)$$

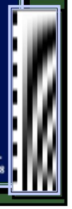




discrete cosine transform set

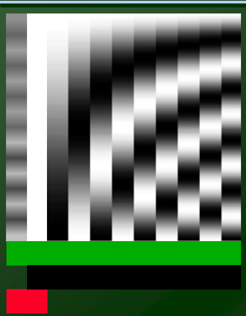
GLM fitted



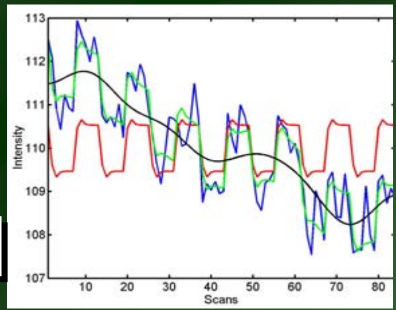


High pass filter

$Y = X\beta + \varepsilon$



data and three different models



Serial correlation (fMRI)

General Linear Model
⇒ Generalised Linear Model

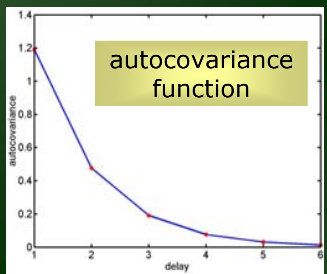
fMRI time series are auto-correlated:

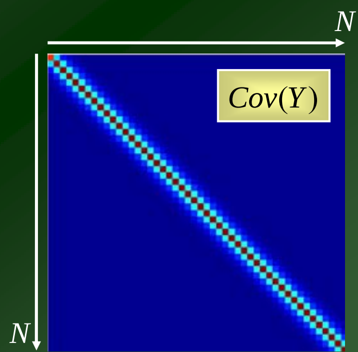
- **adaptation of general linear model necessary for valid test**
- **estimation of autocorrelation**
- **Optimal low-pass filter**

Serial correlation

$Y_t = aY_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim N(0, \sigma^2)$

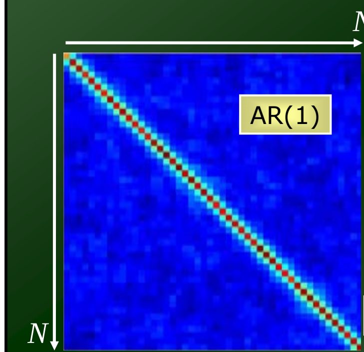
autoregressive process of order 1 (AR(1))

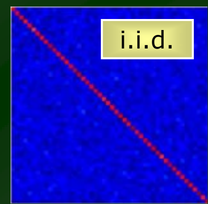




Error covariance matrix

$Y = X\beta + \varepsilon$ → $Cov(\varepsilon)$





sampled error covariance matrices (10^3 voxels)

Serial correlation

Serial correlations

$\underline{Y} = X\underline{\beta} + \varepsilon$ $\varepsilon \sim N(\underline{0}, \sigma^2 V)$ –intrinsic autocorrelation V

Problem:

Estimate $\sigma^2 V$ at each voxel and make inference about $\underline{c}^T\underline{\beta}$

Model:

Model V as linear combination of m variance components

$$V = \lambda_1 Q_1 + \lambda_2 Q_2 + \ldots + \lambda_m Q_m$$

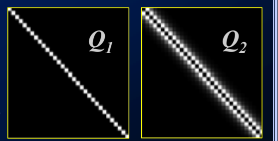
Assumptions:

V is the same at each voxel

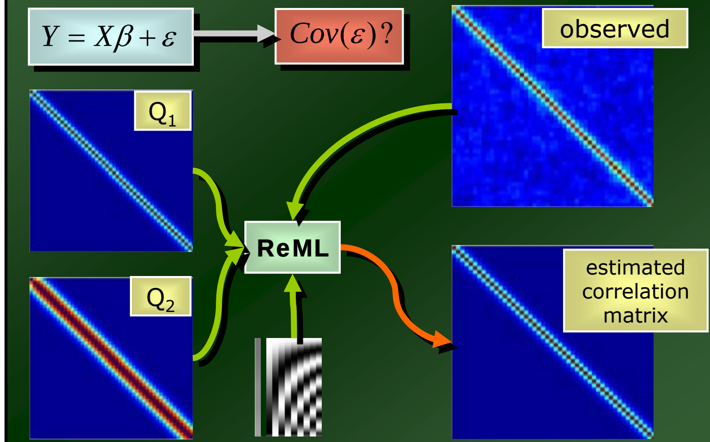
σ^2 is different at each voxel

Example:

For one fMRI session, use 2 variance components. Choice of Q_1 and Q_2 motivated by autoregressive model of order 1 plus white noise (AR(1)+wn)



Restricted Maximum Likelihood



Serial correlations…estimation

$\hat{\underline{\beta}} = (X^TX)^- X^T\underline{Y}$ – unbiased, ordinary least squares estimate

Compute sample covariance matrix of data at all activated voxels: $C_Y = \Sigma_k Y_k Y_k^T / K$

Important: Data Y_k must be high-pass filtered.

Model C_Y as $C_Y = X\underline{\beta}\,\underline{\beta}^T X^T + \Sigma \lambda_i Q_i$

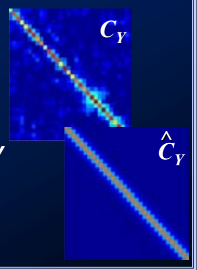
and estimate hyperparameters λ_i using "Restricted Maximum Likelihood" (ReML)

Estimate V by $V = N\, \Sigma \hat{\lambda}_i Q_i / trace(\Sigma \hat{\lambda}_i Q_i)$

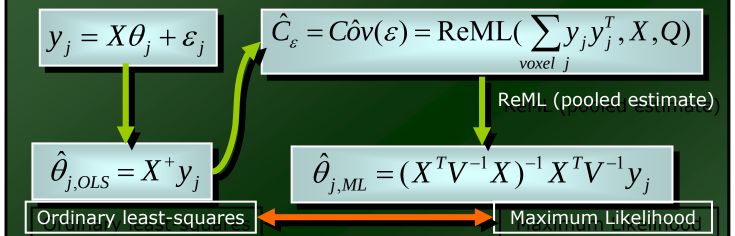
Estimate σ^2 at each voxel in the usual way by

$\hat{\sigma}^2 = (R\underline{Y})^T (R\underline{Y}) / \mathrm{trace}(R\,V)$ – unbiased

where $R = I - X(X^TX)^- X$



Estimation in SPM



- 2 passes (first pass for selection of voxels)
- more accurate estimate of V

Assume, at voxel j:

$$C_{\varepsilon,j} = \sigma_j V$$

$$t = \frac{c^T\theta}{\mathrm{SE}(c^T\theta)}$$

$$\mathrm{SE}(c^T\theta) = \sqrt{\hat{\sigma}^2 c^T (V^{-1/2}X)^- (V^{-1/2}X)^{-T} c}$$

Serial correlations...inference

Inference:

To test null hypothesis $c^T\beta = 0$, compute t-value by dividing size of effect by its standard deviation: $t = c^T\hat{\beta} / std[c^T\hat{\beta}]$

where $std[c^T\hat{\beta}] = sqrt(\sigma^2 c'(X^TX)^- X^T V X(X^TX)^- c)$

... but ... $std[c^T\hat{\beta}]$ *is not a* χ^2 *variable because of* V

Approximating χ^2 *distribution using* ***Satterthwaite approximation:***

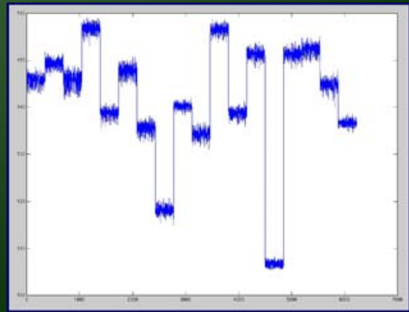
$$\mathrm{Var}[\sigma^2] = 2\sigma^4\,\mathrm{trace}(R\,V\,R\,V) / \mathrm{trace}(R\,V)^2$$

$$\nu = 2E[\sigma^2]^2/\mathrm{Var}[\sigma^2] = \mathrm{trace}(RV)^2 / \mathrm{trace}(RVRV)$$

– effective degrees of freedom

Use t-distribution with ν *degrees of freedom to compute p-value for* t

Absolute value of BOLD signal is meaningless



fMRI signal of an individual voxel across scans and sessions

⇒ **Scale the scans by the session global mean**

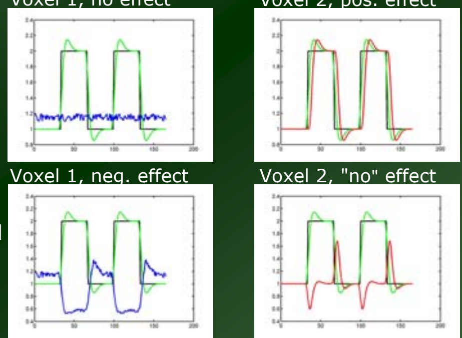
fMRI Global scaling artefact

Scale each scan by its own global mean ?

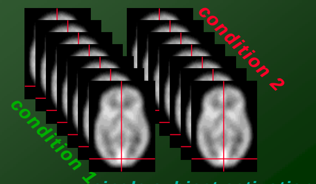
No scaling.

With scaling: voxel/global

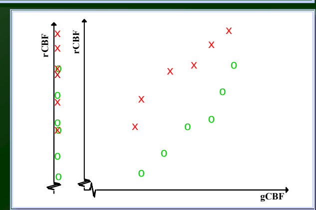
Stimulus Global Voxel 1 Voxel 2



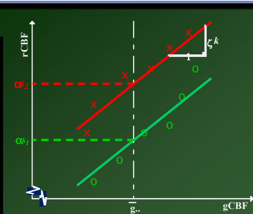
PET Global effects: AnCova...



single subject activation



- **AnCova**
  - **classic way to include a nuisance covariate into a comparison**
- **assumptions**
  - **linear / parallel**
  - **Constant across conditions**

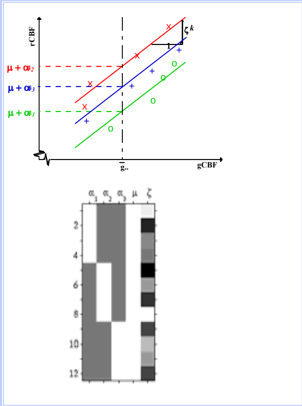


Single subject activation (AnCova)

$Y_{qj} = \alpha_q + \mu + \zeta\,(g_{qj} - \bar{g}_{..}) + \varepsilon_{jq}$

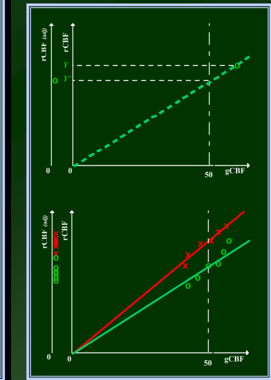
- $\varepsilon_{jq} \sim N(0,\sigma^2)$

- **Null hypothesis** *(at this voxel)*
  - $H_0 : \alpha_1 = \alpha_2$
- **parameter vector**
  - $\underline{\beta} = (\alpha_1, \alpha_2, \alpha_3, \mu, \zeta)^T$
- **contrast weights** $\underline{c}$
  - activation: $H_1 : \alpha_2 > \alpha_1$
    - $(-1\ {+1}\ 0)$
  - deactivation: $H_1 : \alpha_1 > \alpha_2$
    - $({+1}\ {-1}\ 0)$
- ***F*** **- test**
  - $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$



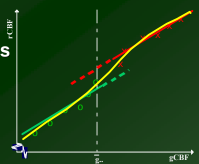
Proportional scaling by gCBF...

- **scale gCBF to 50ml/min/dl**
  - $Y'_{qj} = Y_{qj} \,/\, (g_{qj} / 50)$
  - → *statistics on adjusted data*
- **scales variance**
  - *e.g. single subject activation*
  - $Y'_{qj} = \alpha_q + \varepsilon_{qj}$
  - $\varepsilon_{qj} \sim N(0,\sigma^2)$
  - $Y_{qj} = \alpha_q\ (g_{qj}/50) + \varepsilon'_{qj}$
  - $\varepsilon'_{qj} \sim N(0, \sigma^2 \times (g_{qj}/50)^2)$

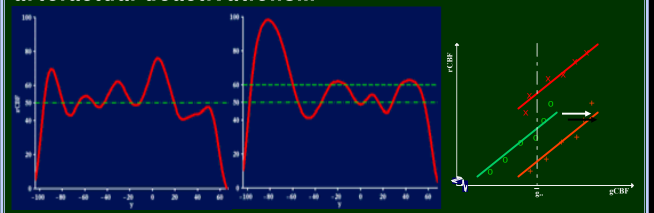


Confounded covariates...

- **E.g. global effects**
  - frequently correct for global changes
  - nuisance effect ?
  - global mean affected by response ?
- **Motion effects in *f*MRI**



artefactual deactivations...

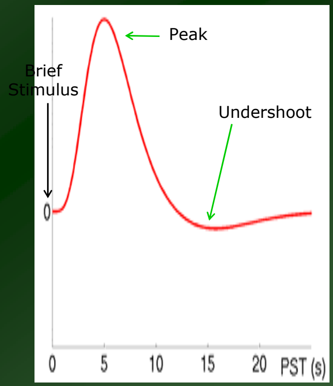


Contents

- Introduction & Recap
- F-test and added variance
- Good & bad models
- Improved model
- HRF and ER fMRI
- « Take home » message

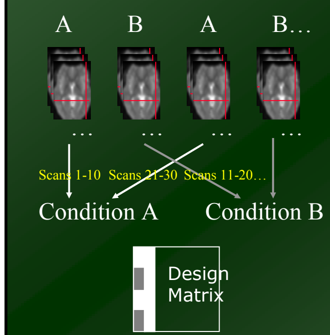
BOLD Impulse Response

- Function of blood oxygenation, flow, volume (Buxton et al, 1998)
- Peak (max. oxygenation) 4-6s poststimulus; baseline after 20-30s
- Initial undershoot can be observed (Malonek & Grinvald, 1996)
- Similar across V1, A1, S1…
- … but differences across: other regions (Schacter et al 1997) individuals (Aguirre et al, 1998)



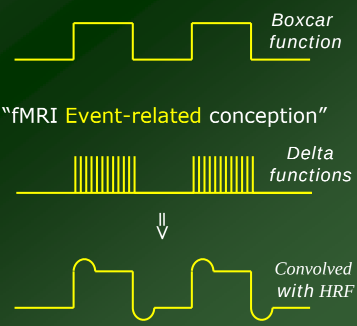
Epoch vs Event-related fMRI

"PET Blocked conception"
(scans assigned to conditions)



"fMRI Epoch conception"
(scans treated as timeseries)

"fMRI Event-related conception"

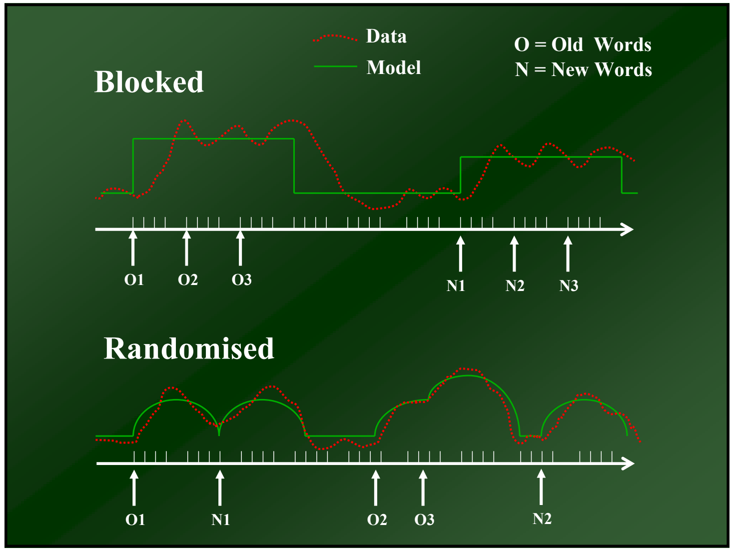


Overview

1. Advantages of efMRI
2. BOLD impulse response
3. General Linear Model
4. Temporal Basis Functions
5. Timing Issues
6. Design Optimisation

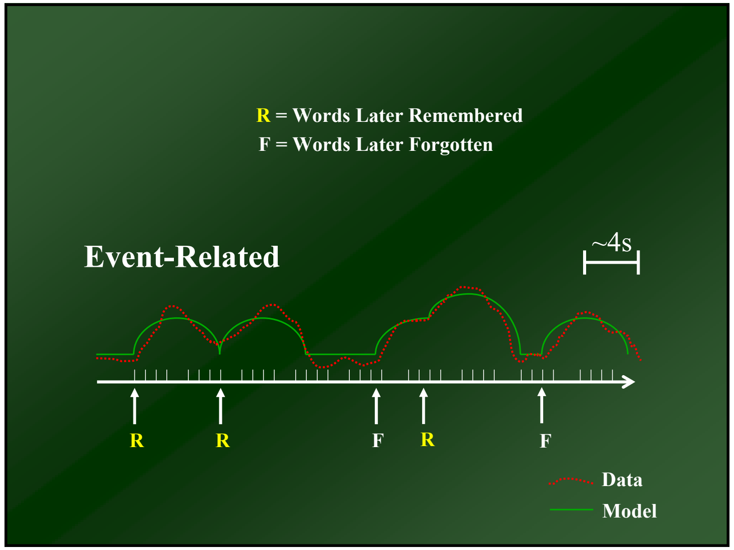
Advantages of Event-related fMRI

1. Randomised trial order
c.f. confounds of blocked designs



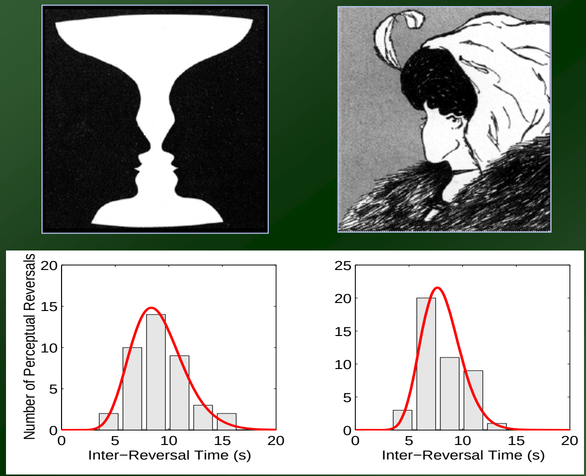
Advantages of Event-related fMRI

1. Randomised trial order
   c.f. confounds of blocked designs
2. Post hoc / subjective classification of trials
   e.g, according to subsequent memory



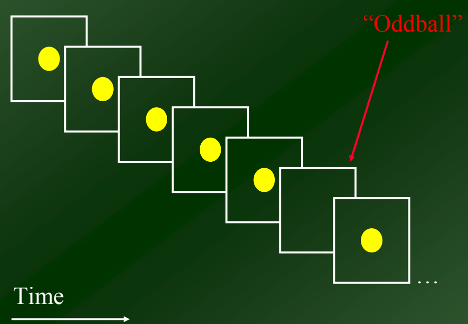
Advantages of Event-related fMRI

1. Randomised trial order
   c.f. confounds of blocked designs
2. Post hoc / subjective classification of trials
   e.g, according to subsequent memory
3. Some events can only be indicated by subject (in time)
   e.g, spontaneous perceptual changes



Advantages of Event-related fMRI

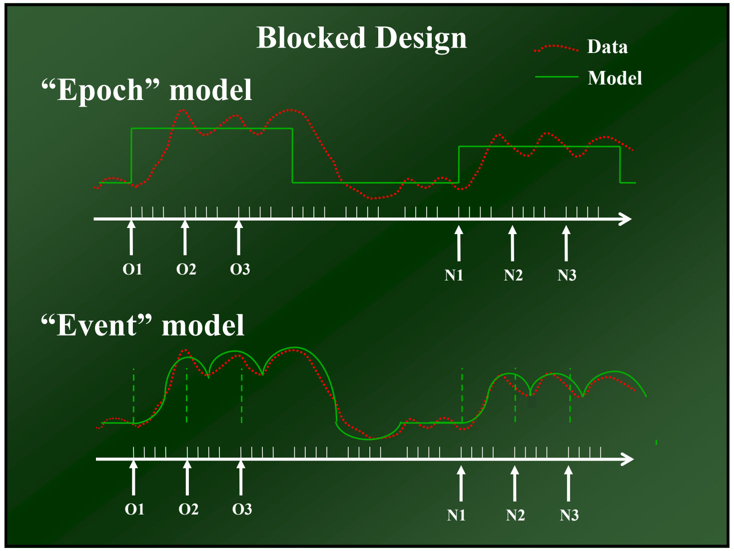
1. **Randomised trial order**
   c.f. confounds of blocked designs
2. **Post hoc / subjective classification of trials**
   e.g, according to subsequent memory
3. **Some events can only be indicated by subject (in time)**
   e.g, spontaneous perceptual changes
4. **Some trials cannot be blocked**
   e.g, "oddball" designs



Advantages of Event-related fMRI

1. **Randomised trial order**
   c.f. confounds of blocked designs
2. **Post hoc / subjective classification of trials**
   e.g, according to subsequent memory
3. **Some events can only be indicated by subject (in time)**
   e.g, spontaneous perceptual changes
4. **Some trials cannot be blocked**
   e.g, "oddball" designs

5. **More accurate models even for blocked designs?**
   e.g, "state-item" interactions



Disadvantage of Randomised Designs

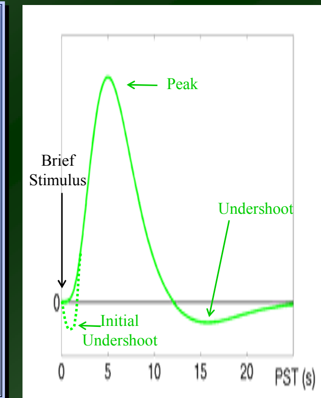
1. Less efficient for detecting effects than are blocked designs *(see later…)*
2. Some psychological processes may be better blocked (eg task-switching, attentional instructions)

Overview

1. Advantages of efMRI
2. BOLD impulse response
3. General Linear Model
4. Temporal Basis Functions
5. Timing Issues
6. Design Optimisation

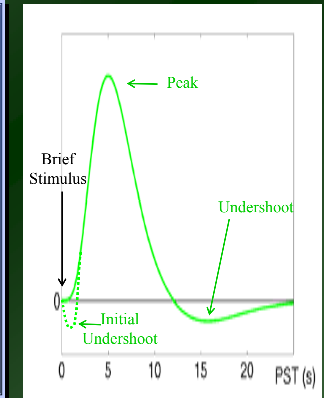
BOLD Impulse Response

- Function of blood oxygenation, flow, volume (Buxton et al, 1998)
- Peak (max. oxygenation) 4-6s poststimulus; baseline after 20-30s
- Initial undershoot can be observed (Malonek & Grinvald, 1996)
- Similar across V1, A1, S1…
- … but differences across: other regions (Schacter et al 1997) individuals (Aguirre et al, 1998)



BOLD Impulse Response

- Early event-related fMRI studies used a long Stimulus Onset Asynchrony (SOA) to allow BOLD response to return to baseline
- However, if the BOLD response is explicitly modelled, overlap between successive responses at short SOAs can be accommodated…
- … particularly if responses are assumed to superpose linearly
- Short SOAs are more sensitive…



Overview

1. Advantages of efMRI
2. BOLD impulse response
3. General Linear Model
4. Temporal Basis Functions
5. Timing Issues
6. Design Optimisation

General Linear (Convolution) Model

GLM for a single voxel:

$Y(t) = x(t) \otimes h(t) + \varepsilon$

$x(t)$ = stimulus train (delta functions)

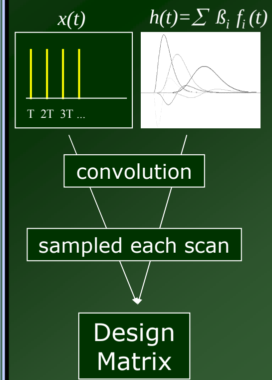
$x(t) = \sum \delta(t - nT)$

$h(t)$ = hemodynamic (BOLD) response

$h(t) = \sum \beta_i f_i(t)$

$f_i(t)$ = temporal basis functions

$Y(t) = \sum\sum \beta_i f_i(t - nT) + \varepsilon$



General Linear Model (in SPM)

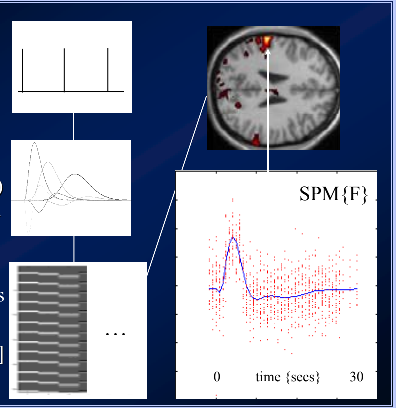
Auditory words every 20s

(Orthogonalised) Gamma functions $f_i(u)$ of peristimulus time u

Sampled every TR = 1.7s

Design matrix, **X**

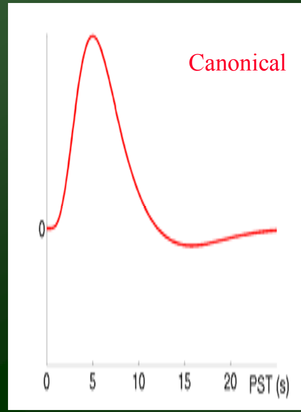
$[f_1(u) \otimes x(t) \mid f_2(u) \otimes x(t) \mid \ldots]$



Overview

1. Advantages of efMRI
2. BOLD impulse response
3. General Linear Model
4. Temporal Basis Functions
5. Timing Issues
6. Design Optimisation

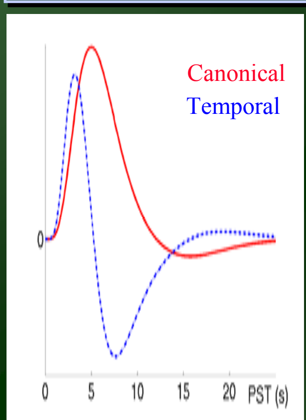
Temporal Basis Functions



Informed Basis Set (Friston et al. 1998)

- Canonical HRF (2 gamma functions)

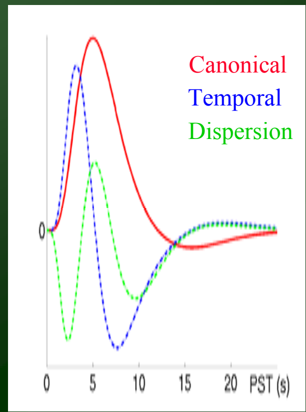
Temporal Basis Functions



Informed Basis Set (Friston et al. 1998)

- Canonical HRF (2 gamma functions) *plus* Multivariate Taylor expansion in:
  - time (*Temporal Derivative*)

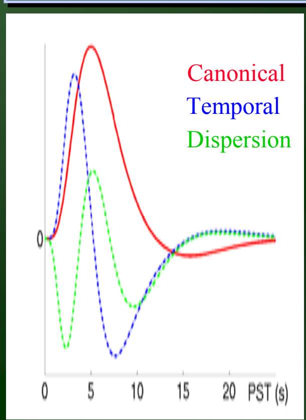
Temporal Basis Functions



Informed Basis Set (Friston et al. 1998)

- Canonical HRF (2 gamma functions) *plus* Multivariate Taylor expansion in:
  - time (*Temporal Derivative*)
  - width (*Dispersion Derivative*)

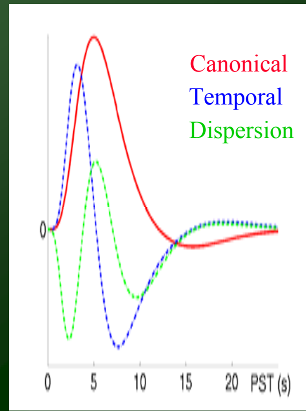
Temporal Basis Functions



Informed Basis Set (Friston et al. 1998)

- Canonical HRF (2 gamma functions) *plus* Multivariate Taylor expansion in:
  - time (*Temporal Derivative*)
  - width (*Dispersion Derivative*)
- "Magnitude" inferences via t-test on canonical parameters (providing canonical is a good fit…more later)

Temporal Basis Functions



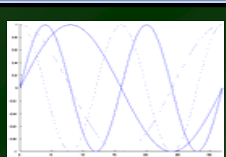
Informed Basis Set (Friston et al. 1998)

- Canonical HRF (2 gamma functions) *plus* Multivariate Taylor expansion in:
  - time (*Temporal Derivative*)
  - width (*Dispersion Derivative*)
- "Magnitude" inferences via t-test on canonical parameters (providing canonical is a good fit…more later)

- "Latency" inferences via tests on *ratio* of derivative : canonical parameters (more later…)

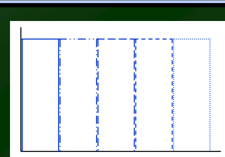
Temporal Basis Functions

- Fourier Set
  - Windowed sines & cosines
  - Any shape (up to frequency limit)
  - Inference via F-test



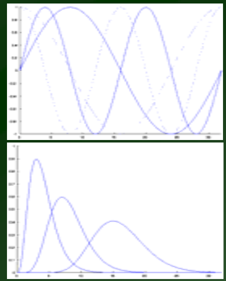
Temporal Basis Functions

- Finite Impulse Response (FIR)
  - Mini timebins (selective averaging)
  - Any shape (up to bin-width)
  - Inference via F-test



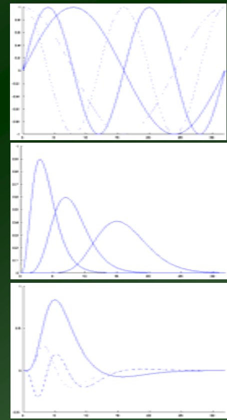
Temporal Basis Functions

- **Fourier Set**
  - Windowed sines & cosines
  - Any shape (up to frequency limit)
  - Inference via F-test
- **Gamma Functions**
  - Bounded, asymmetrical (like BOLD)
  - Set of different lags
  - Inference via F-test

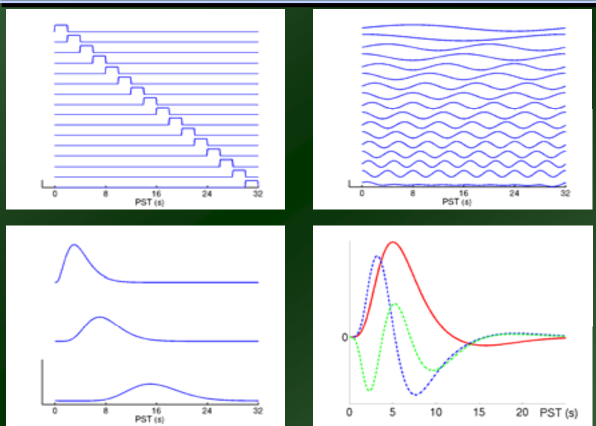


Temporal Basis Functions

- **Fourier Set**
  - Windowed sines & cosines
  - Any shape (up to frequency limit)
  - Inference via F-test
- **Gamma Functions**
  - Bounded, asymmetrical (like BOLD)
  - Set of different lags
  - Inference via F-test
- **Informed Basis Set**
  - Best guess of canonical BOLD response
  - Variability captured by Taylor expansion
  - "Magnitude" inferences via t-test...?

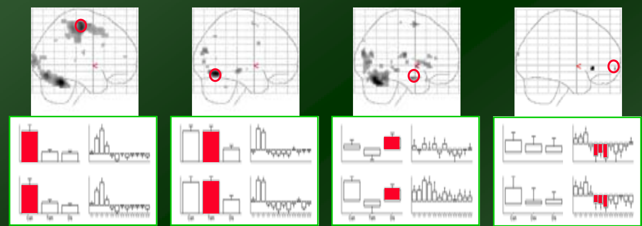


Temporal Basis Functions



Temporal Basis Sets: Which One?

In this example (rapid motor response to faces, *Henson et al, 2001*)...



Canonical + Temporal + Dispersion + FIR

...canonical + temporal + dispersion derivatives appear sufficient

...may not be for more complex trials (eg stimulus-delay-response)

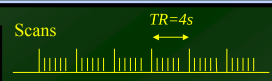
...but then such trials better modelled with separate neural components (ie activity no longer delta function) + constrained HRF (Zarahn, 1999)

Overview

1. Advantages of efMRI
2. BOLD impulse response
3. General Linear Model
4. Temporal Basis Functions
5. Timing Issues
6. Design Optimisation

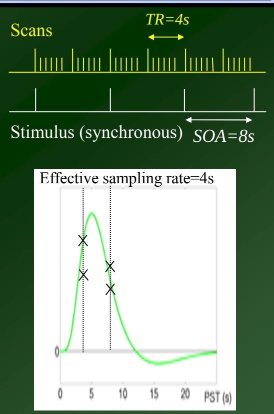
Timing Issues

- Typical TR for 48 slice EPI at 3mm spacing is ~ 4s



Timing Issues

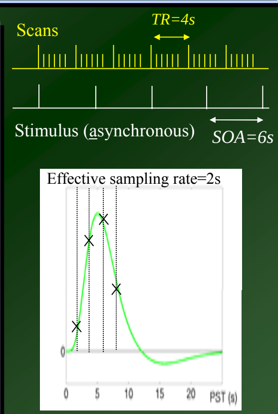
- Typical TR for 48 slice EPI at 3mm spacing is ~ 4s
- Sampling at [0,4,8,12…] post- stimulus may miss peak signal



Timing Issues

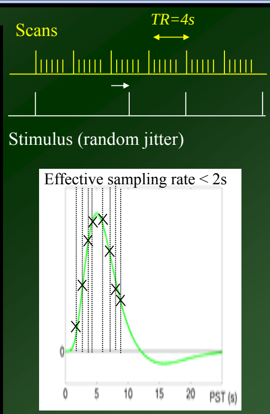
- Typical TR for 48 slice EPI at 3mm spacing is ~ 4s
- Sampling at [0,4,8,12…] post- stimulus may miss peak signal
- Higher effective sampling by:

  1. Asynchrony, *eg.* *SOA=1.5TR*



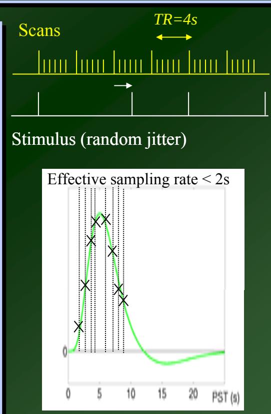
Timing Issues

- Typical TR for 48 slice EPI at 3mm spacing is ~ 4s
- Sampling at [0,4,8,12…] post-stimulus may miss peak signal
- Higher effective sampling by:
  1. Asynchrony, *eg. SOA=1.5TR*
  2. Random Jitter, eg. *SOA=(2±0.5)TR*



Timing Issues

- Typical TR for 48 slice EPI at 3mm spacing is ~ 4s
- Sampling at [0,4,8,12…] post-stimulus may miss peak signal
- Higher effective sampling by:
  1. Asynchrony, *eg. SOA=1.5TR*
  2. Random Jitter, eg. *SOA=(2±0.5)TR*
- Better response characte-risation (Miezin et al, 2000)



Timing Issues

- …but "Slice-timing Problem" (Henson et al, 1999)

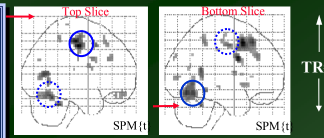
  Slices acquired at different times, yet model is the same for all slices

Timing Issues

- …but "Slice-timing Problem" (Henson et al, 1999)

  Slices acquired at different times, yet model is the same for all slices

  => different results (using canonical HRF) for different reference slices



Timing Issues

- …but "Slice-timing Problem" (Henson et al, 1999)

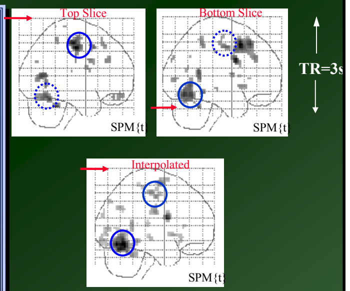
  Slices acquired at different times, yet model is the same for all slices

  => different results (using canonical HRF) for different reference slices

- Solutions:

1. Temporal interpolation of data

   … but less good for longer TRs



Timing Issues

- …but "Slice-timing Problem" (Henson et al, 1999)

  Slices acquired at different times, yet model is the same for all slices

  => different results (using canonical HRF) for different reference slices

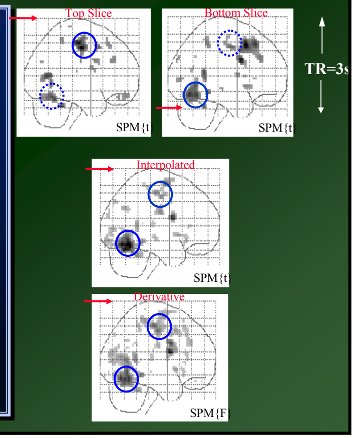
- Solutions:

1. Temporal interpolation of data

   … but less good for longer TRs

2. More general basis set (e.g., with temporal derivatives)

   … but inferences via F-test



BOLD Response Latency (Linear)

- Assume the real response, $r(t)$, is a scaled (by α) version of the canonical, $f(t)$, but delayed by a small amount dt:

$$r(t) = \alpha f(t+dt) \sim \alpha f(t) + \alpha f'(t)\, dt \qquad 1^{st}\text{-order Taylor}$$

- If the fitted response, $R(t)$, is modelled by the canonical + temporal derivative:

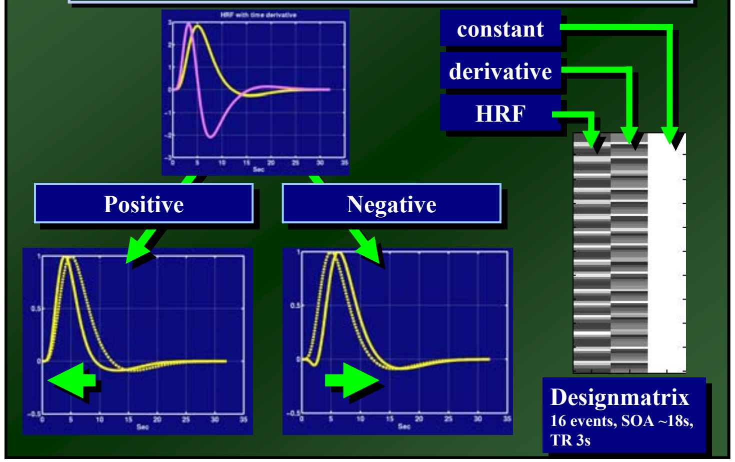
$$R(t) = \beta_1 f(t) + \beta_2 f'(t) \qquad \text{GLM fit}$$

- Then canonical and derivative parameter estimates, β_1 and β_2, are such that:

$$\Rightarrow \quad \alpha = \beta_1 \qquad dt = \beta_2 / \beta_1$$

ie, Latency can be approximated by the ratio of derivative-to-canonical parameter estimates (within limits of first-order approximation, +/- 1s)

BOLD Response Latency: example

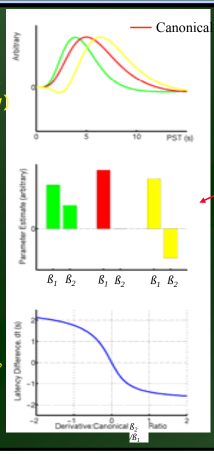


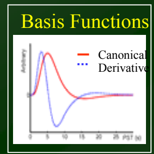
BOLD Response Latency (Linear)

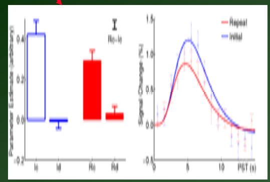
Delayed Responses (green/ yellow)

Parameter Estimates

Actual latency, *dt*, vs. $ß_2 / ß_1$



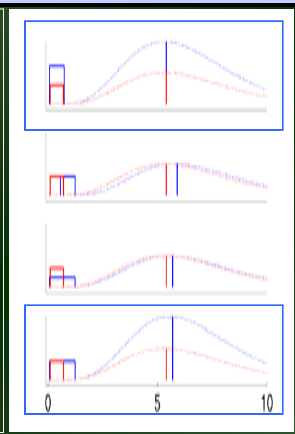




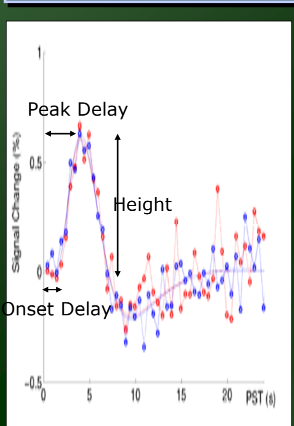
Face repetition reduces latency as well as magnitude of fusiform response

Neural Response Latency?

Neural	*BOLD*
A. Decreased	A. Smaller Peak
B. Advanced	B. Earlier Onset
C. Shortened (same integrated)	C. Earlier Peak
D. Shortened (same maximum)	D. Smaller Peak *and earlier Peak*

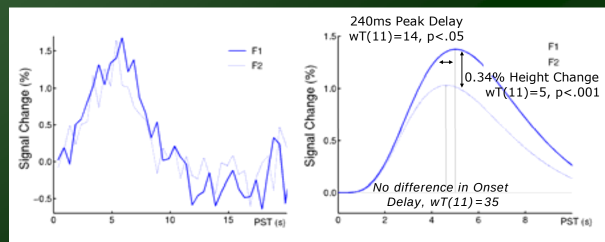


BOLD Response Latency (Iterative)

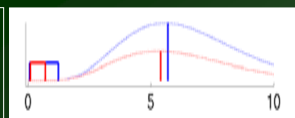


- Numerical fitting of explicitly parameterised canonical HRF *(Henson et al, 2001)*
- Distinguishes between *Onset* and *Peak* latency…
  …unlike temporal derivative…
  …and which may be important for interpreting neural changes (see previous slide)
- Distribution of parameters tested nonparametrically (Wilcoxon's T over subjects)

BOLD Response Latency (Iterative)



Neural
D. Shortened (same maximum)



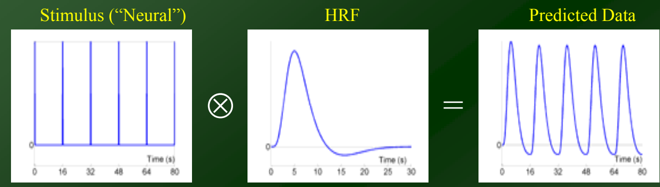
BOLD
D. Smaller Peak *and earlier Peak*

Most parsimonious account is that repetition reduces ***duration*** *of neural activity…*

Overview

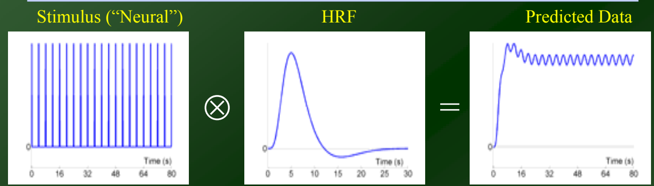
1. Advantages of efMRI
2. BOLD impulse response
3. General Linear Model
4. Temporal Basis Functions
5. Timing Issues
6. Design Optimisation

Fixed SOA = 16s



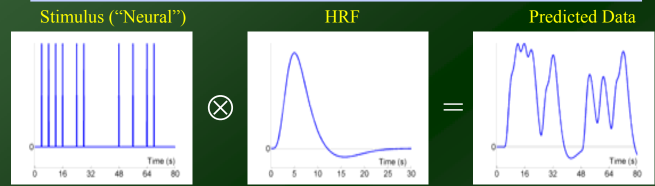
Not particularly efficient…

Fixed SOA = 4s



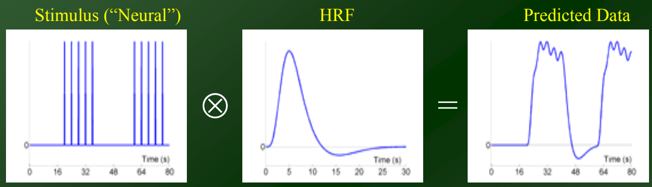
Very Inefficient…

Randomised, SOA_{min} = 4s



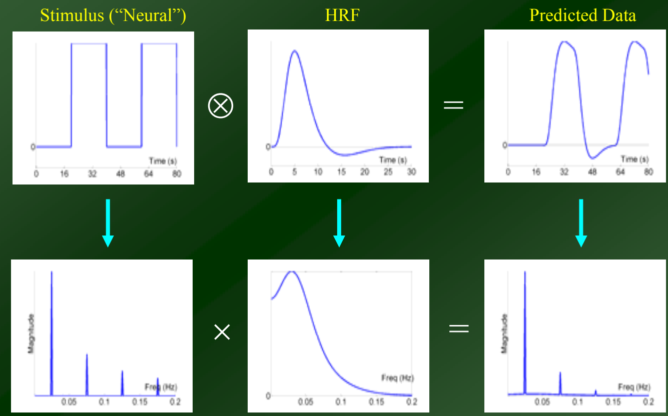
More Efficient…

Blocked, SOA_{min} = 4s



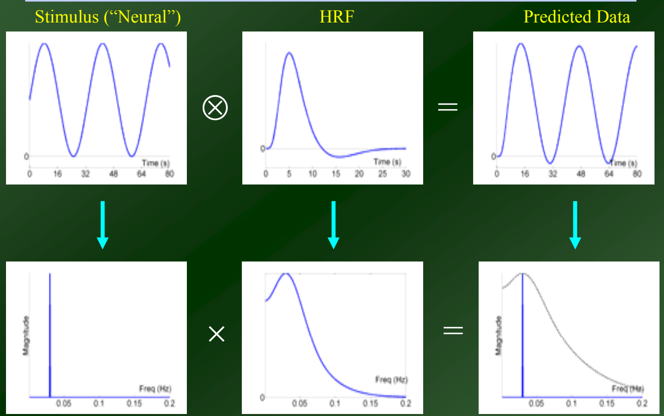
Even more Efficient…

Blocked, epoch = 20s



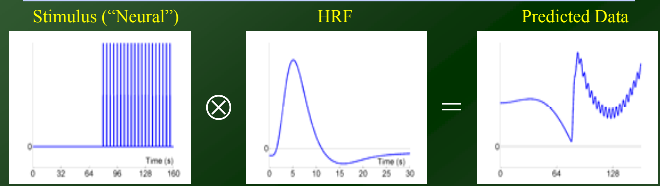
Blocked-epoch (with small SOA) and Time-Freq equivalences

Sinusoidal modulation, f = 1/33s

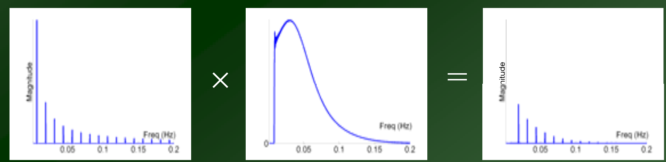


The most efficient design of all!

Blocked (80s), SOA_{min}=4s, highpass filter = 1/120s

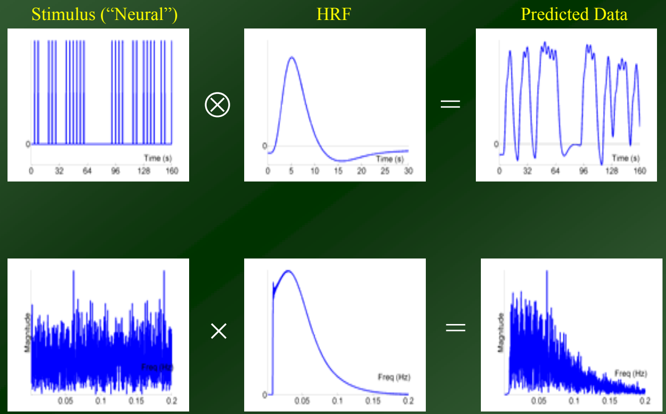


"Effective HRF" (after highpass filtering)
(Josephs & Henson, 1999)



Don't have long (>60s) blocks!

Randomised, SOA_{min}=4s, highpass filter = 1/120s



(Randomised design spreads power over frequencies)

Design Efficiency

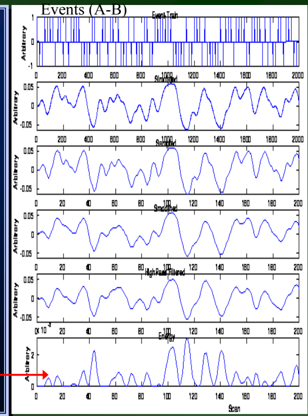
$T = c^T\beta / std(c^T\beta)$

$std(c^T\beta) = sqrt(\sigma^2 c^T(X^TX)^{-1}c)$ (i.i.d)

- **For max. T, want min. contrast variability (Friston et al, 1999)**
- **If assume that noise variance (σ^2) is unaffected by changes in X…**
- **…then want maximal efficiency, e:**

$e(c,X) = \{ c^T (X^TX)^{-1} c \}^{-1}$

- **= maximal bandpassed signal energy (Josephs & Henson, 1999)**

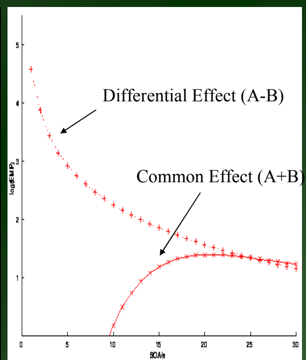


Efficiency - Multiple Event-types

- **Design parametrised by:**
  SOA_{min} **Minimum SOA**
  $p_i(h)$ **Probability of event-type *i* given history *h* of last *m* events**
- **With *n* event-types *$p_i(h)$* is a *$n^m \times n$ Transition Matrix***
- **Example: Randomised AB**

	A	B
A	0.5	0.5
B	0.5	0.5

=> ABBBABAABABAAA...



4s smoothing; 1/60s highpass filtering

Efficiency - Multiple Event-types

- **Example: Alternating AB**

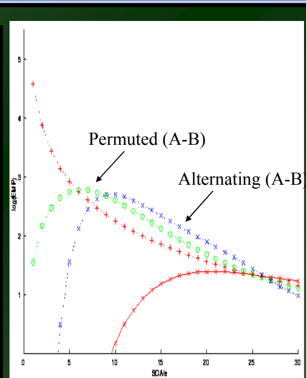
	A	B
A	**0**	**1**
B	**1**	**0**

=> ABABABABABAB...

- Example: Permuted AB

	A	**B**
AA	0	1
AB	0.5	0.5
BA	0.5	0.5
BB	1	0

=> **ABBAABABABBA...**



4s smoothing; 1/60s highpass filtering

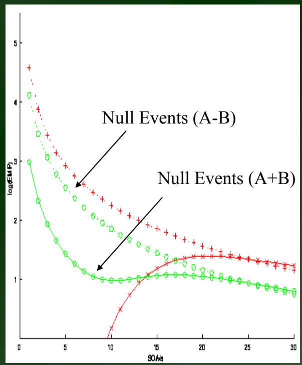
Efficiency - Multiple Event-types

- **Example: Null events**

	A	B
A	0.33	0.33
B	0.33	0.33

=> AB-BAA--B---ABB...

- **Efficient for differential *and* main effects at short SOA**
- **Equivalent to stochastic SOA (Null Event like third unmodelled event-type)**
- **Selective averaging of data (Dale & Buckner 1997)**



4s smoothing; 1/60s highpass filtering

Efficiency - Conclusions

- Optimal design for one contrast may not be optimal for another
- Blocked designs generally most efficient with short SOAs (but earlier restrictions and problems of interpretation…)
- With randomised designs, optimal SOA for differential effect (A-B) is minimal SOA (assuming no saturation), whereas optimal SOA for main effect (A+B) is 16-20s
- Inclusion of null events improves efficiency for main effect at short SOAs (at cost of efficiency for differential effects)
- If order constrained, intermediate SOAs (5-20s) can be optimal; if SOA constrained, pseudo-randomised designs can be optimal (but may introduce context-sensitivity)

Contents

- Introduction & Recap
- Good & bad models
- Improved model
- HRF and ER fMRI
- **« Take home » message**

Way to proceed

Prepare your questions.
ALL the questions !

Find a model which
- allows contrasts that translates these questions.
- takes into account ALL the effects (interaction, sessions,etc)

Devise task & stimulus presentation.

Acquire the data & analyse.

Not the other way round!!!

Three Stages of an Experiment

1. Sledgehammer Approach
   - brute force experiment : powerful stimulus & don't try to control for everything
   - look at was done before or by others
   - run a couple of subjects -- see if it looks promising
   - if it doesn't look great, tweak the stimulus or task
   - try to be a subject yourself so you can notice any problems with stimuli or subject strategies

Three Stages of an Experiment

1. Sledgehammer Approach
2. Real Experiment
   - at some point, you have to stop changing things and collect enough subjects run with the same conditions to publish it
   - how many subjects do you need
     - some psychophysical studies test two or three subjects, many studies test 6-10 subjects, random effects analysis requires at least 15 subjects,...
     - some subjects WILL be rejected, so acquire more than the minimum !
   - can run all subjects in one or two days
     - pro: minimize setup and variability
     - con: "bad magnet day" means a lot of wasted time
     - make sure all the data are treated the "same way". (script)

Three Stages of an Experiment

1. Sledgehammer Approach
2. Real Experiment
3. "Whipped Cream" experiment
   - after the real experiment works, then think about a "whipped cream" version
   - going straight to whipped cream is a huge endeavor, especially if you're new to imaging
   - and it gives you a second paper !