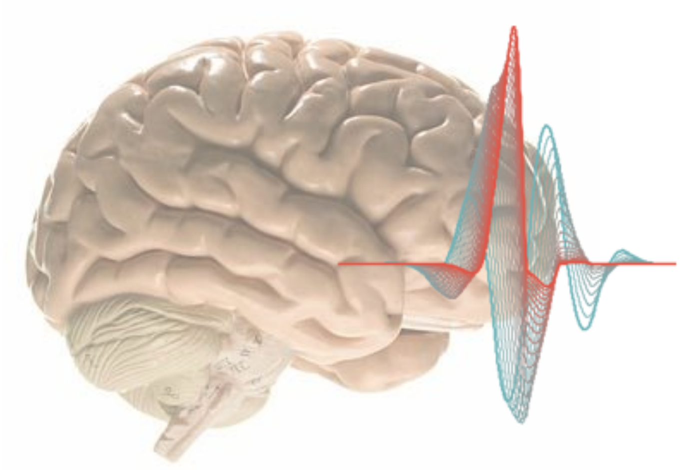


# fMRI statistical analysis: GLM, smoothing, hypothesis testing


Dimitri Van De Ville

Medical Image Processing Lab (MIPLab)

Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne
Department of Radiology and Medical Informatics, Université de Genève

http://miplab.unige.ch/

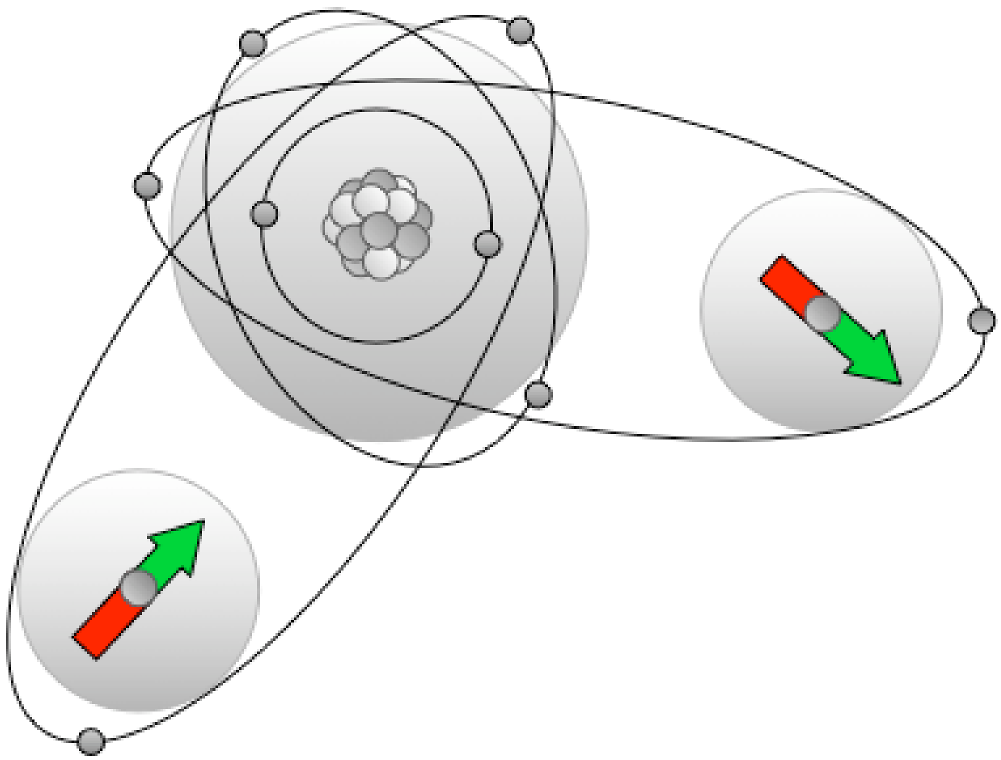

Montréal, September 14, 2009

# Overview

- FMRI in a nutshell
  - physics and neurovascular coupling
- Standard FMRI data analysis
  - confirmatory approach
  - generative model
    - general linear model and its extensions
    - fitting and forward inference
  - Gaussian smoothing
    - blessings and curses
- Perspectives and new avenues
  - backward inference
- Tomorrow’s lecture
  - wavelets enter the arena

# Magnetic Resonance Imaging

- Nuclear magnetic spin of hydrogen atom



  - can be aligned with a strong magnetic field (1.5T/3T/7T/...); no ionizing radiation
  - can be “manipulated” (in a spatially selective way) using clever radiofrequency pulses
  - non-invasive, exquisite soft-tissue contrast, versatile

# Measuring brain activity

- Electrical vs. blood oxygenation dependent signal

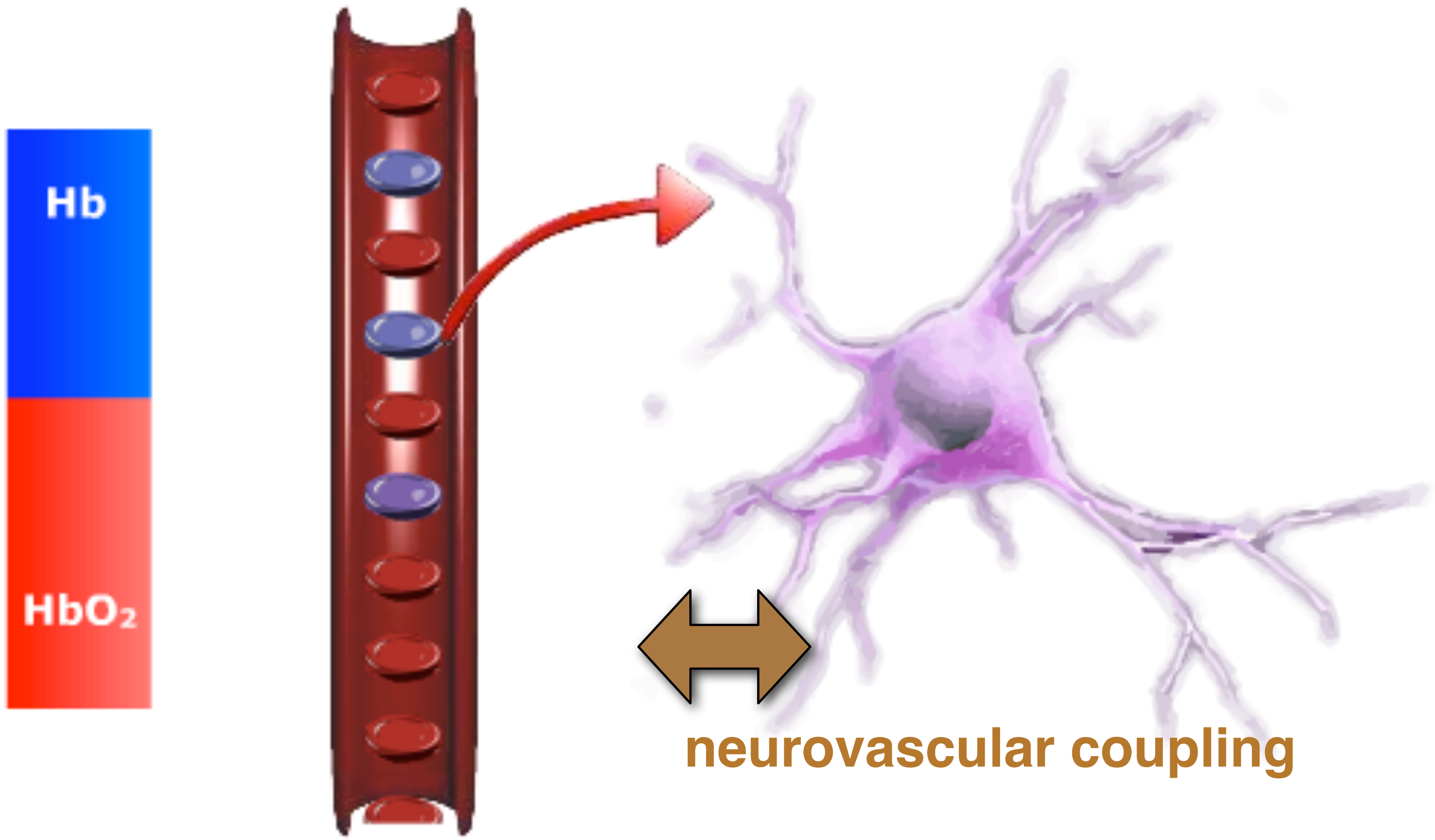
EEG

[Movie from e-mri.org] 

# Measuring brain activity

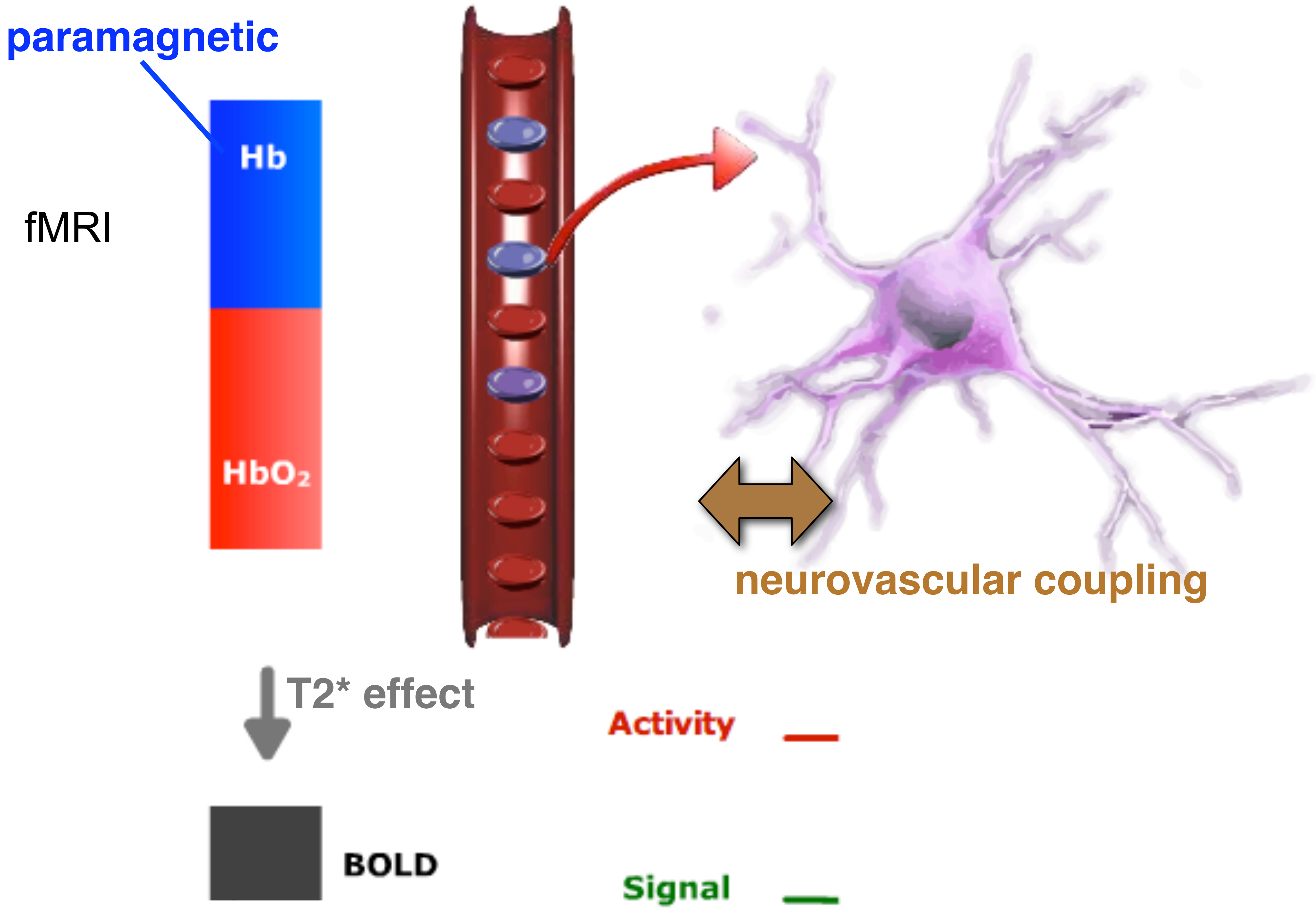
- Electrical vs. blood oxygenation dependent signal

fMRI



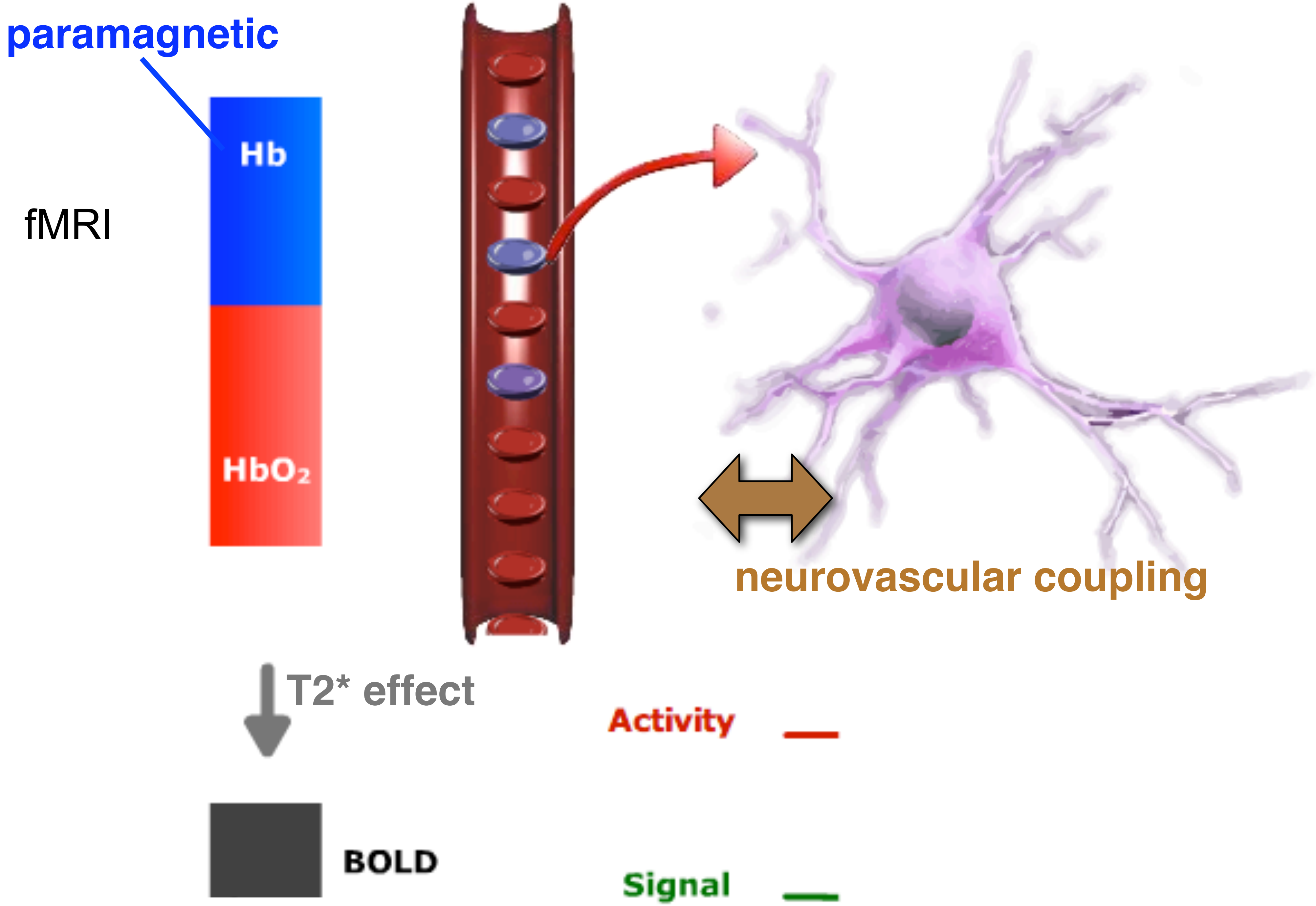
# Measuring brain activity

- Electrical vs. blood oxygenation dependent signal



# Measuring brain activity

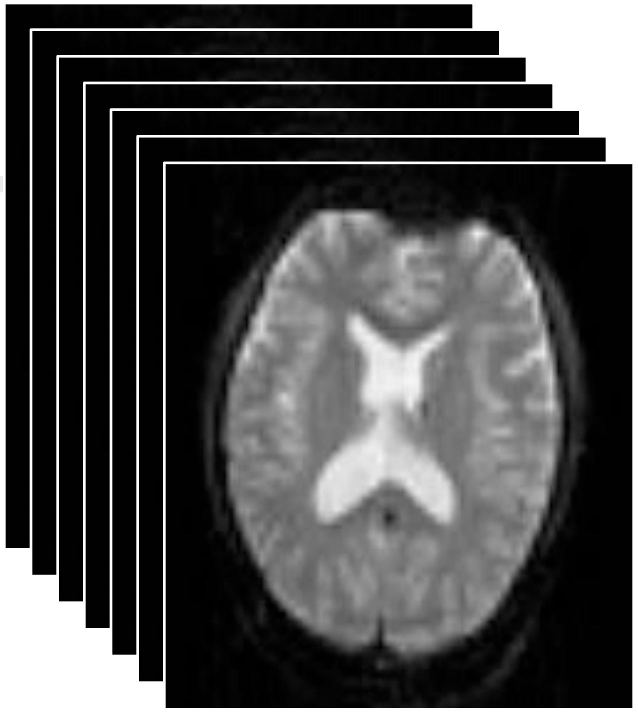
- Electrical vs. blood oxygenation dependent signal



# Functional MRI

- But, what do we really measure?
  - imaging artefacts
    - scanner noise and drifts
    - reconstruction (from k-space to image space)
    - (residual) movements
  - physiological fluctuations
    - aliased cardiac and respitoratory cycle (typical TR=1-3s)
    - regulatory processes
  - intrinsic brain activity
    - “background” activity
      - non-modeled events: thoughts, imagination
      - attention network, default model network
  - evoked brain activity
    - neurovascular coupling;
      link with intensity and duration of neural activation
    - hemodynamic response; interplay of flow, volume, $O_2$ consumption
    - habituation and anticipation effects

# Functional MRI

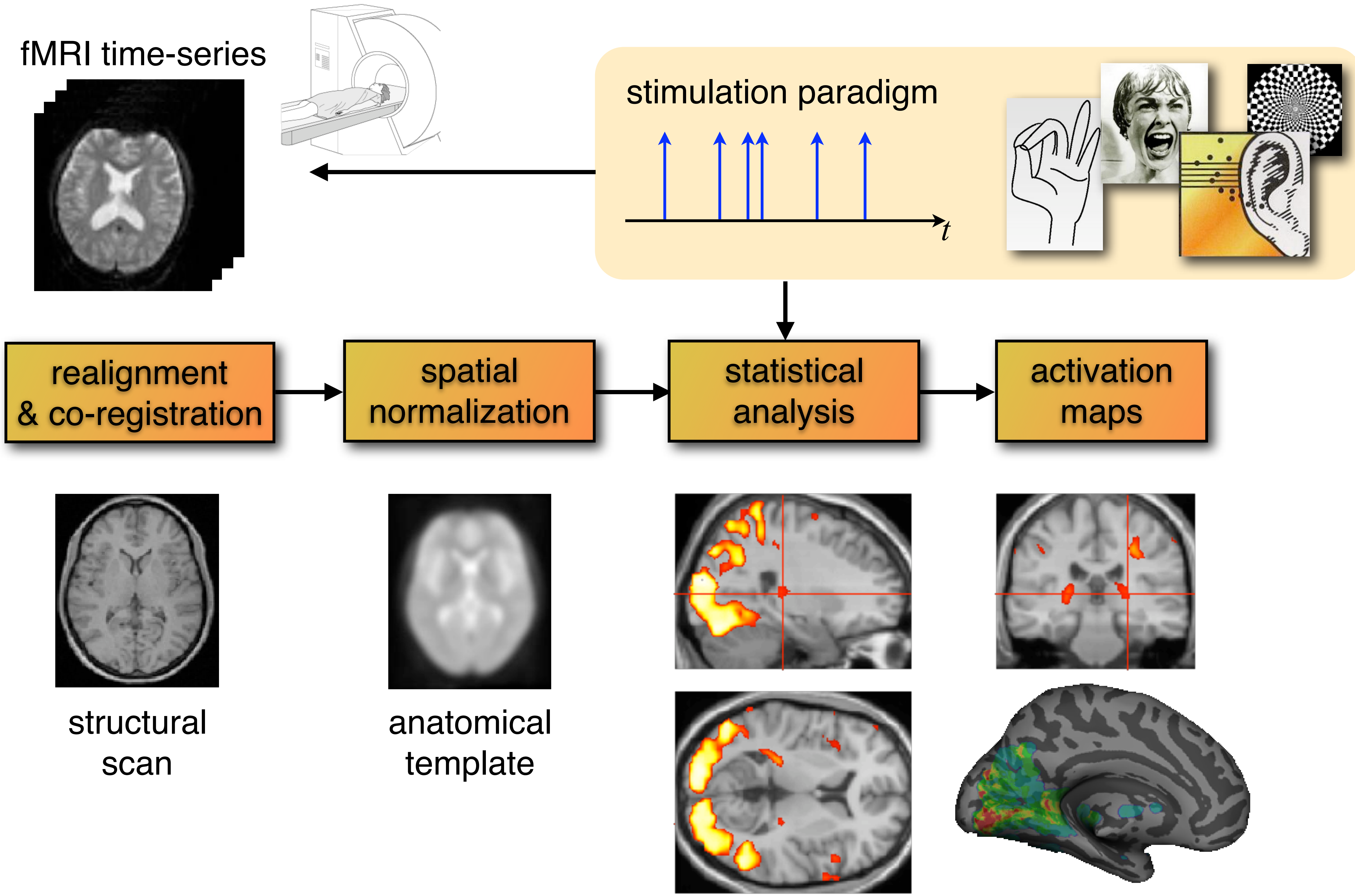


- But, what do we really measure?
  - imaging artefacts
    - scanner noise and drifts
    - reconstruction (from k-space to image space)
    - (residual) movements
  - physiological fluctuations
    - aliased cardiac and respitoratory cycle (typical TR=1-3s)
    - regulatory processes
  - intrinsic brain activity
    - “background” activity
      - non-modeled events: thoughts, imagination
      - attention network, default model network
  - evoked brain activity
    - neurovascular coupling;
      link with intensity and duration of neural activation
    - hemodynamic response; interplay of flow, volume, $O_2$ consumption
    - habituation and anticipation effects

# Confirmatory analysis

- Workhorse of fMRI studies!
  - hypothetize the BOLD response for a given task
  - look for evidence in the measured data
- Parametric model-based hypothesis testing
  - **model-based** vs. blind
  - **univariate** vs. multivariate
    - voxel-wise, fast, easy to interpret
  - **parametric** vs. non-parametric
- Central element of every self-respecting fMRI software!

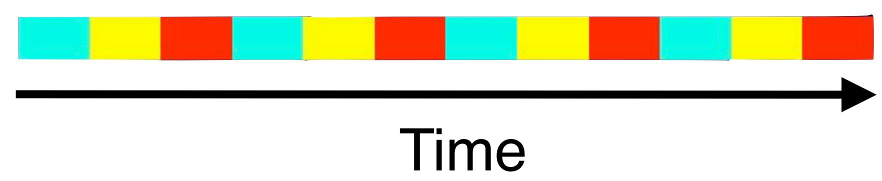
# FMRI data processing pipeline



# Example fMRI experiment

- Three conditions

*Is there a face-selective brain region?*

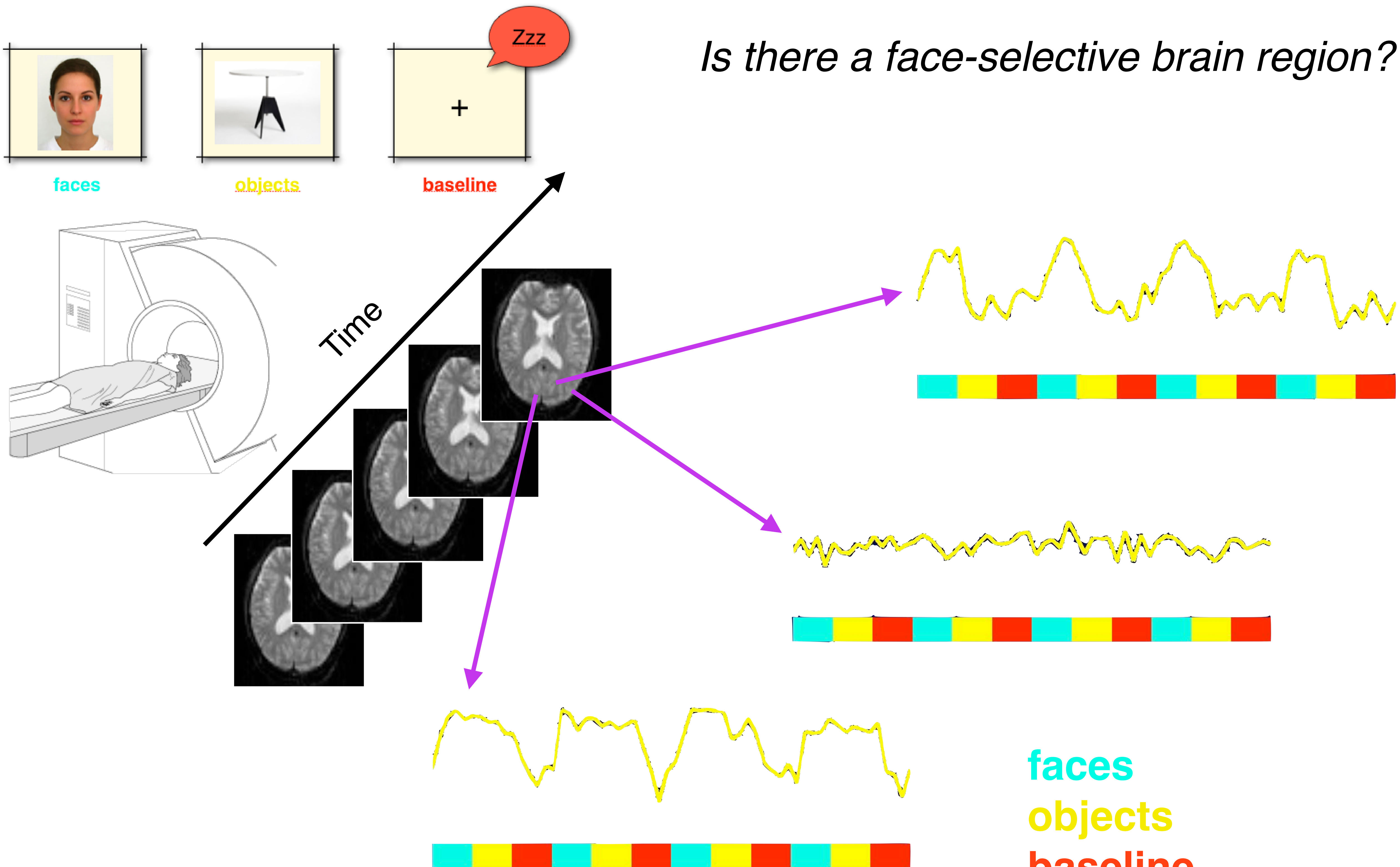




**faces** **objects** **baseline**

# Typical fMRI experiment

*Is there a face-selective brain region?*




faces
objects
baseline

# FMRI statistical analysis

- Voxel-wise General Linear Model (GLM)

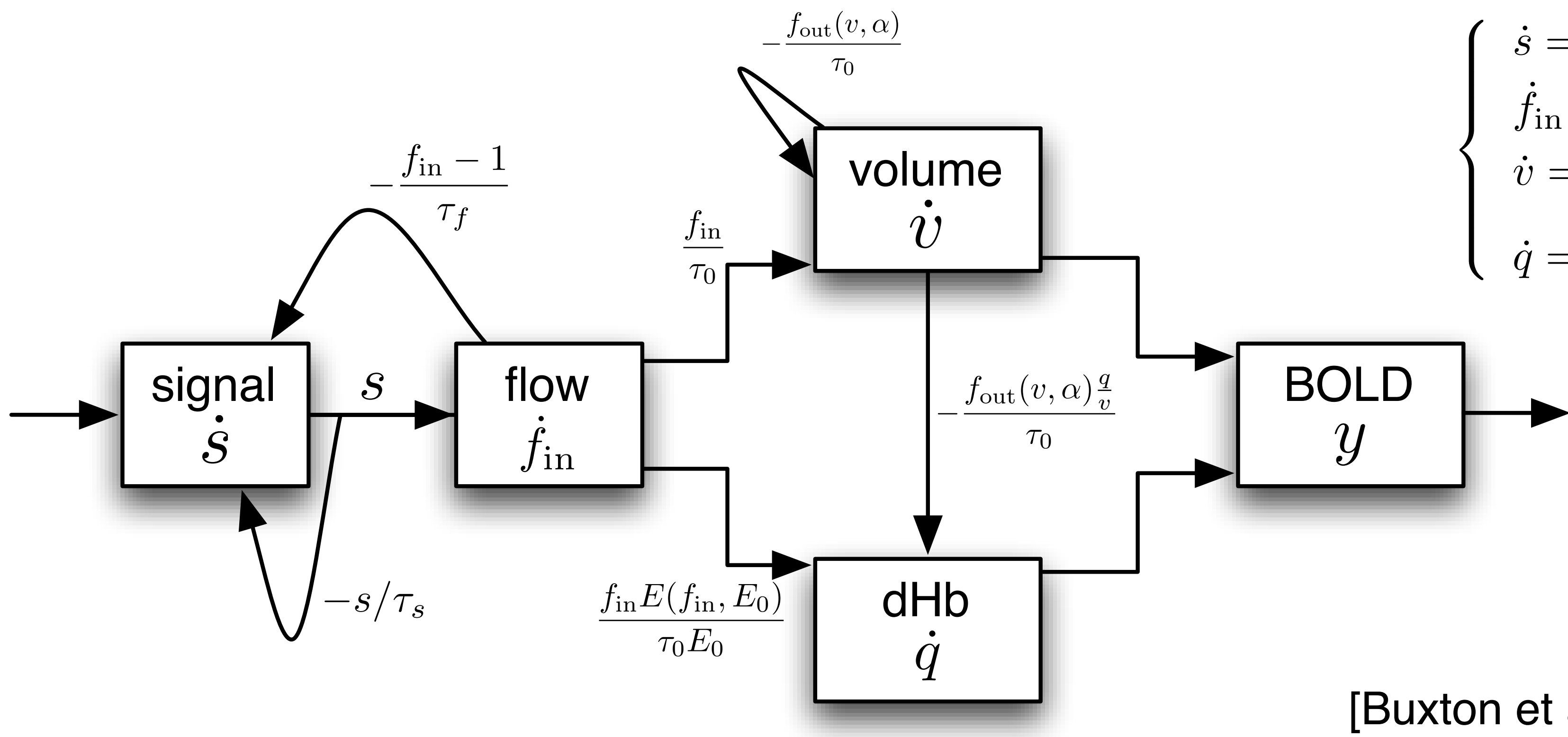
$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$$

  - $\boldsymbol{y} : N \times 1$ — measured timecourse
  - $\boldsymbol{X} : N \times L$ — columns contain $L$ regressors
  - $\boldsymbol{\beta} : L \times 1$ — unknown parameters
  - $\boldsymbol{e} : N \times 1$ — estimation error

- Model specification: setup "design matrix" $\boldsymbol{X}$
- Parameter estimation: find "best" $\hat{\boldsymbol{\beta}}$
- Hypothesis testing and inference

# Model specification

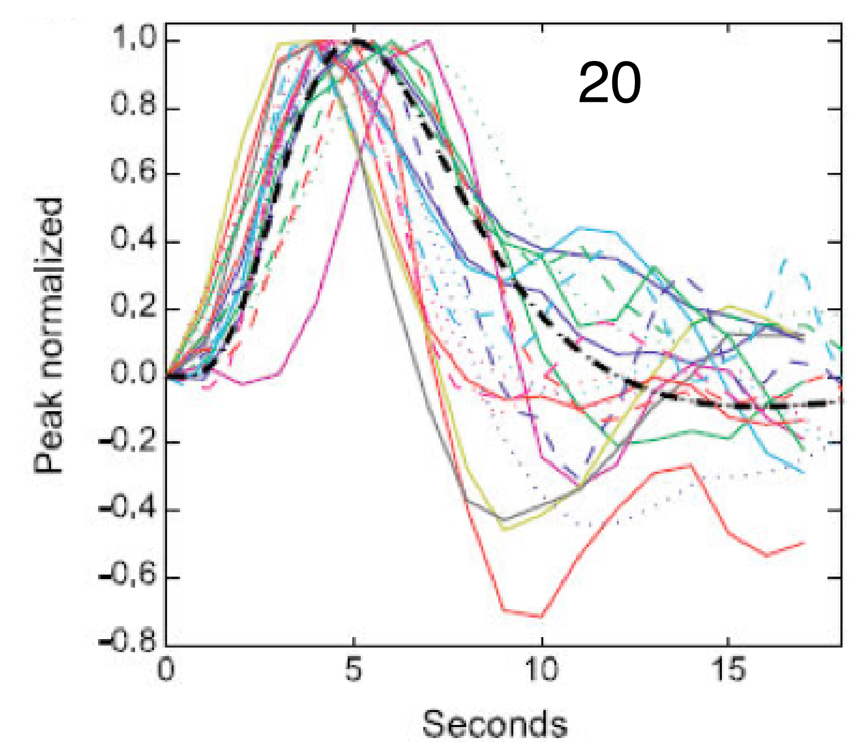
- Hemodynamic modeling




$$\begin{cases} \dot{s} = \epsilon u - \frac{s}{\tau_s} - \frac{f_{\mathrm{in}}-1}{\tau_f} \\ \dot{f}_{\mathrm{in}} = s \\ \dot{v} = \frac{1}{\tau_0}\left(f_{\mathrm{in}} - v^{\frac{1}{\alpha}}\right) \\ \dot{q} = \frac{1}{\tau_0}\left(f_{\mathrm{in}}\frac{1-(1-E_0)^{\frac{1}{f_{\mathrm{in}}}}}{E_0} - v^{\frac{1}{\alpha}-1}q\right) \end{cases}$$

[Buxton et al., 1999; Friston et al., 2000]

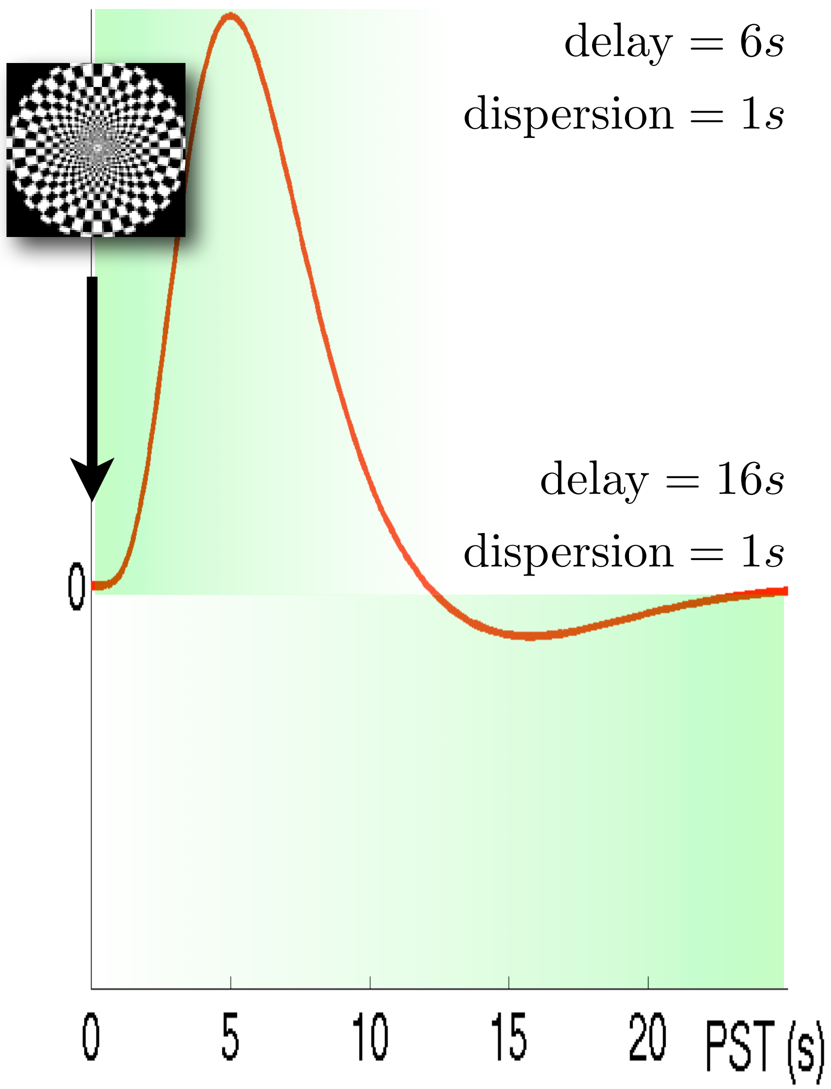
- inter-subject variability
- intra-subject variability
  - same region
    slightly different responses
    - habituation, anticipation
  - different regions
    time-to-peak differs up to 1s
    - eg., visual cortex vs. prefrontal




[Handwerker, 2004; Buckner, 1999]

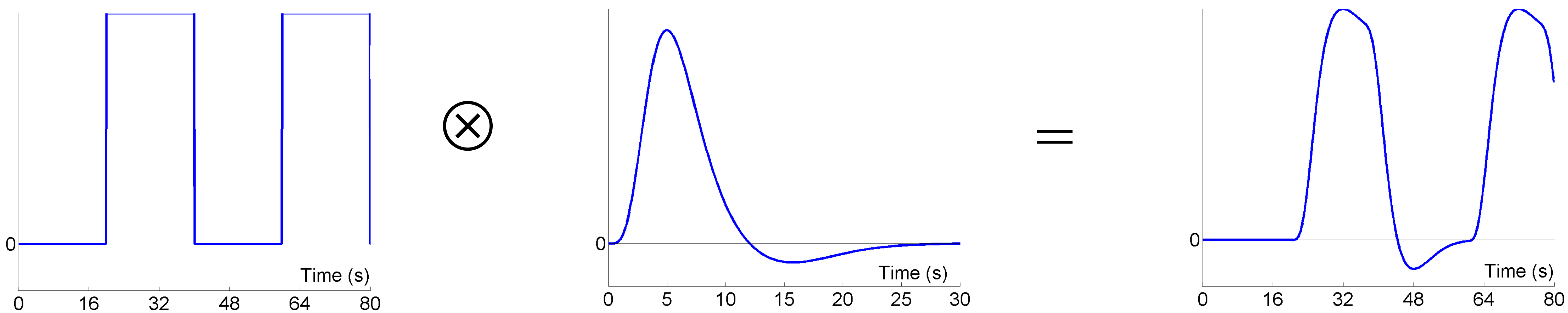
# Model specification

- Hemodynamic response modeling for GLM
  - neglecting non-linear effects
  - mostly a problem for event-related paradigms
- Canonical hemodynamic response function
  - fit of impulse response by difference of Gamma functions

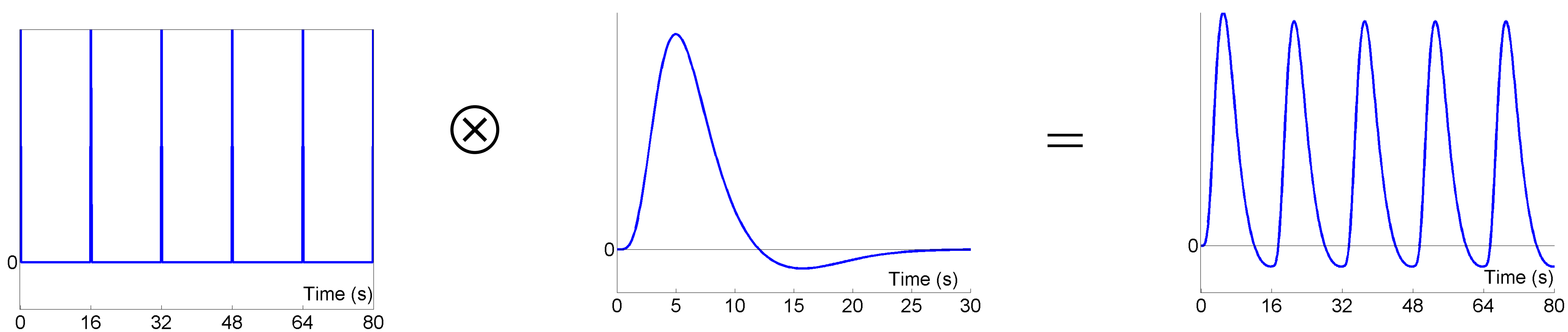


# Model specification

- From stimuli to modeled BOLD response
  - blocks (epochs)

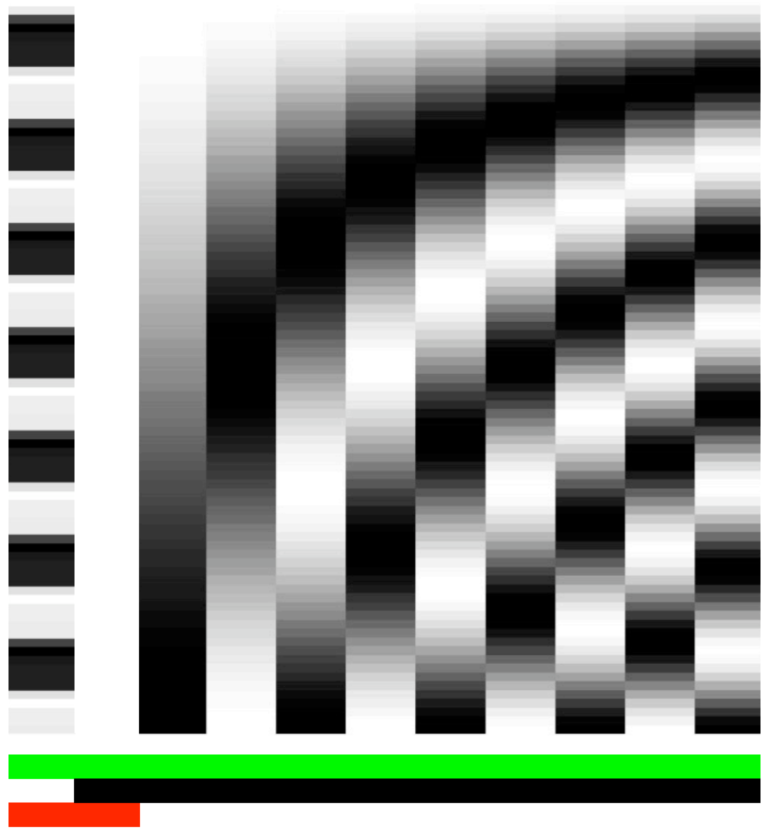


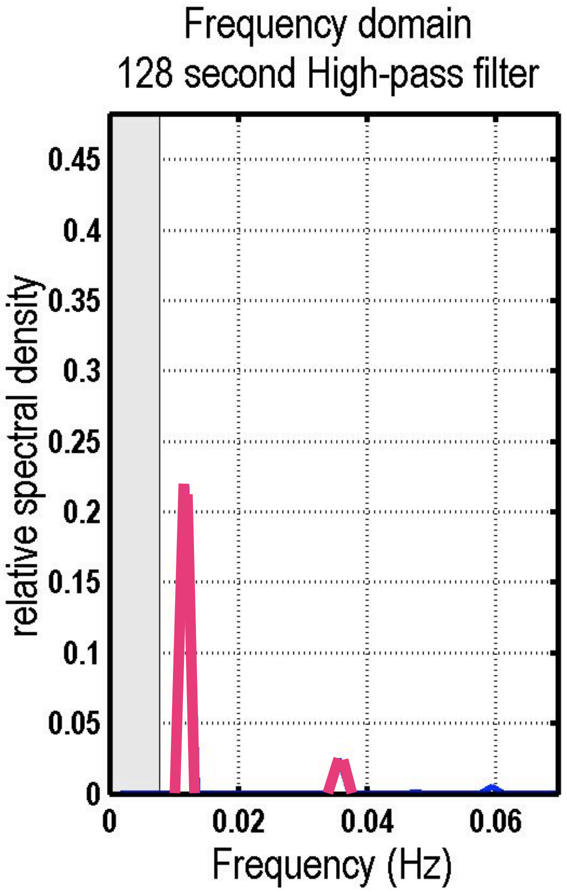

  - events

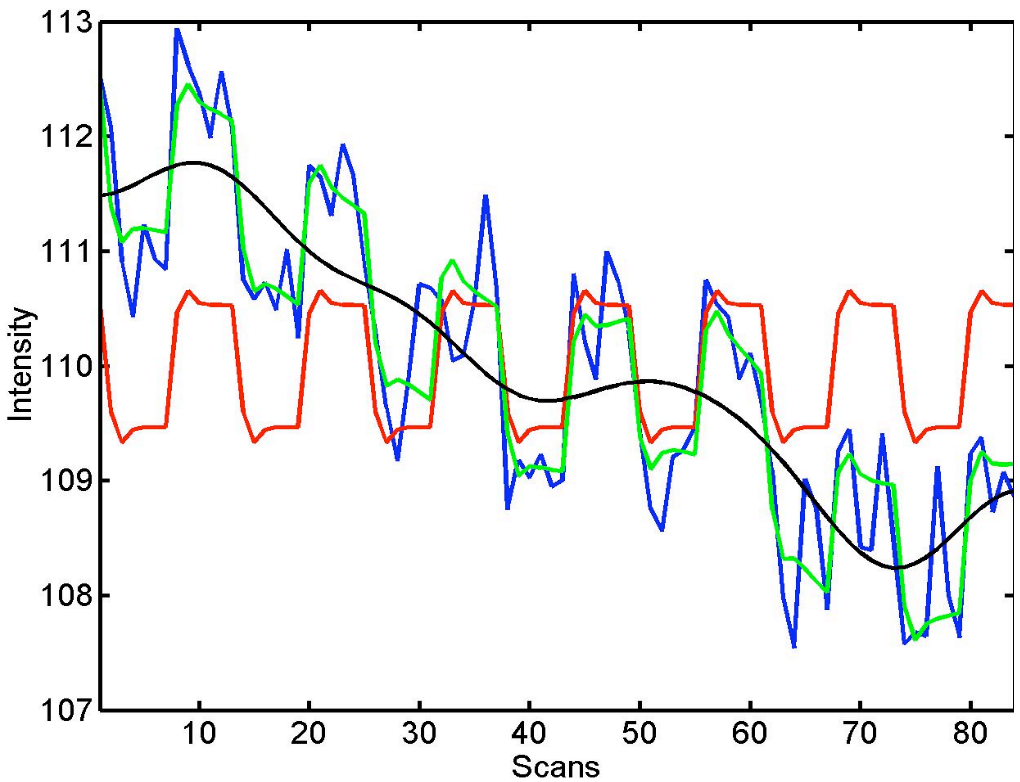


# Model specification

- But, there’s much more going on in the data...
- Add nuissance regressors
- Low-frequency components
  - truncated DCT-basis
  - act like a high-pass filter
  - scanner drifts
  - physiological fluctuations
  - intrinsic brain activity

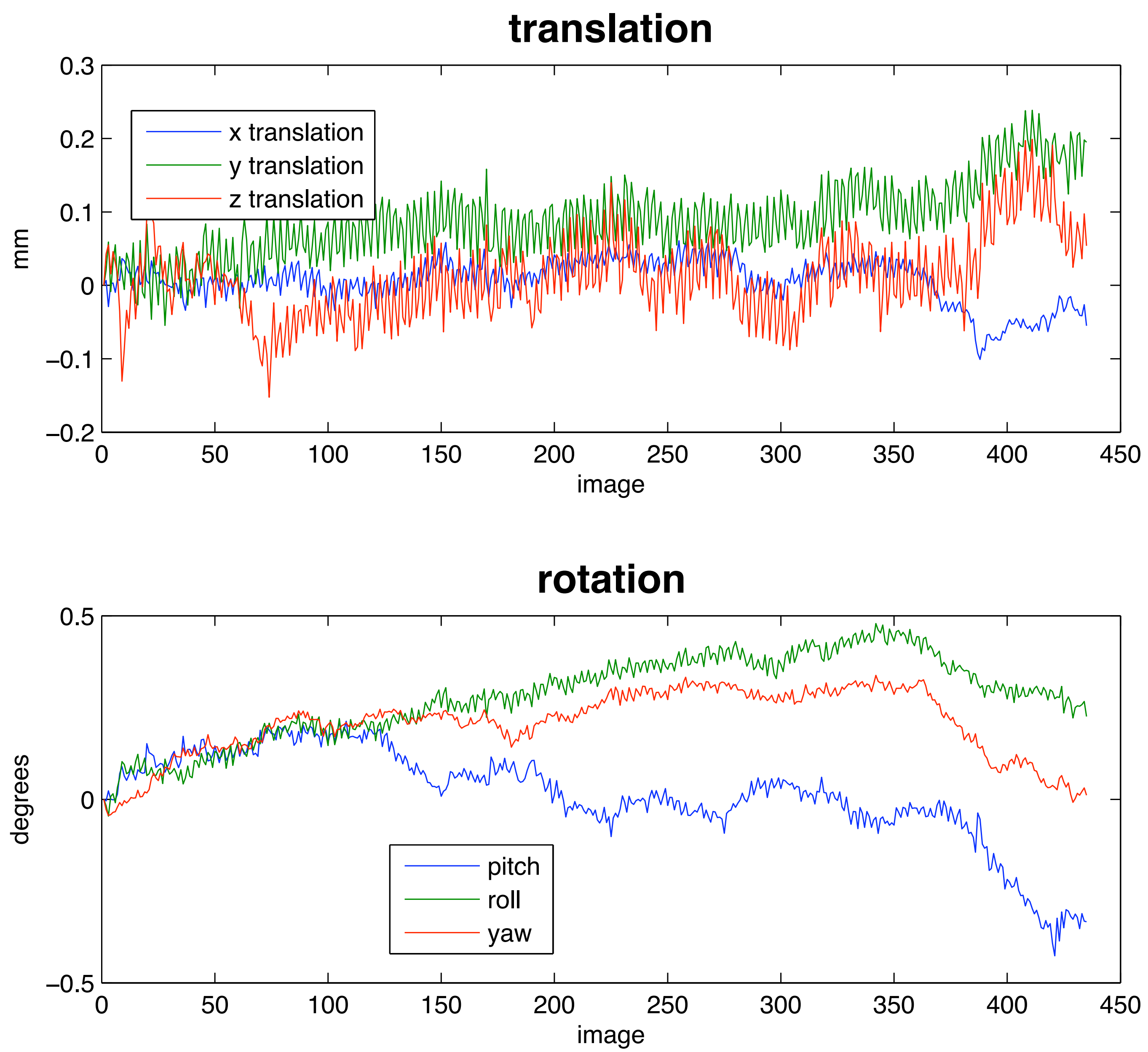


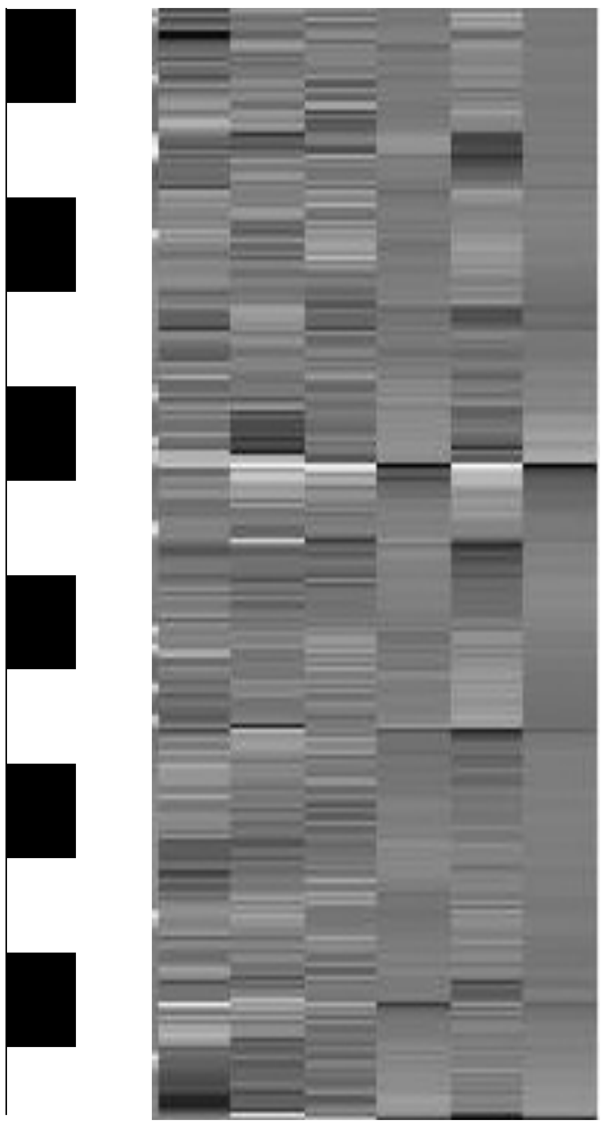


# Model specification

- Realignment parameters
  - capture signal due to residual movement
  - 6 regressors for rigid-body realignment
  - caution: can be correlated with paradigm

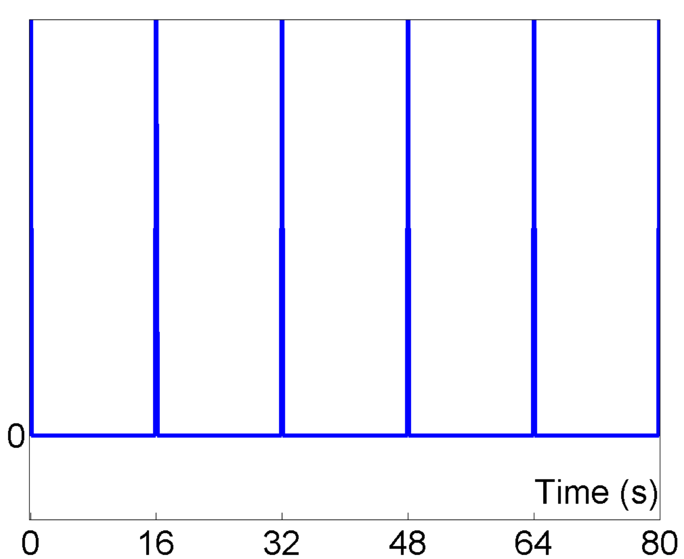


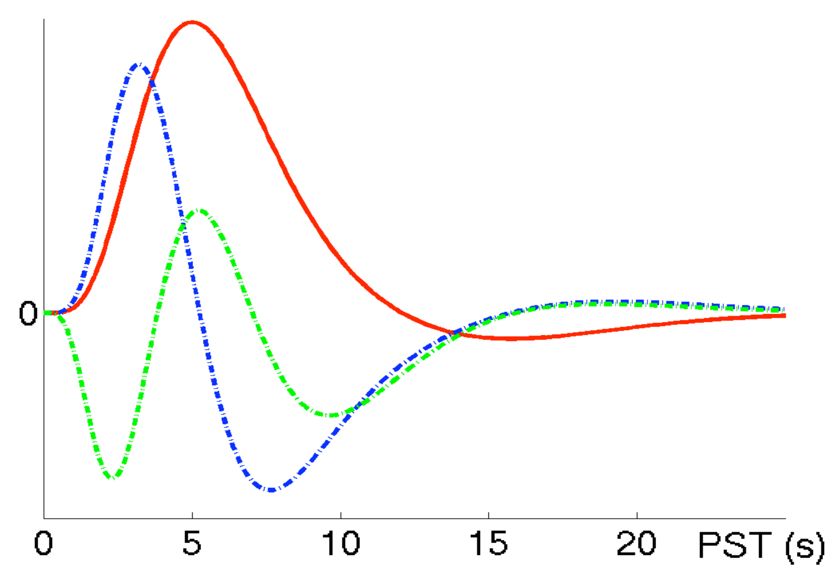
# Model specification

- Hemodynamic variations
  - subject-dependent, regional changes, habituation and anticipation effects
  - approximate $h(t + \Delta t; w + \Delta w)$, where $w$ is dispersion

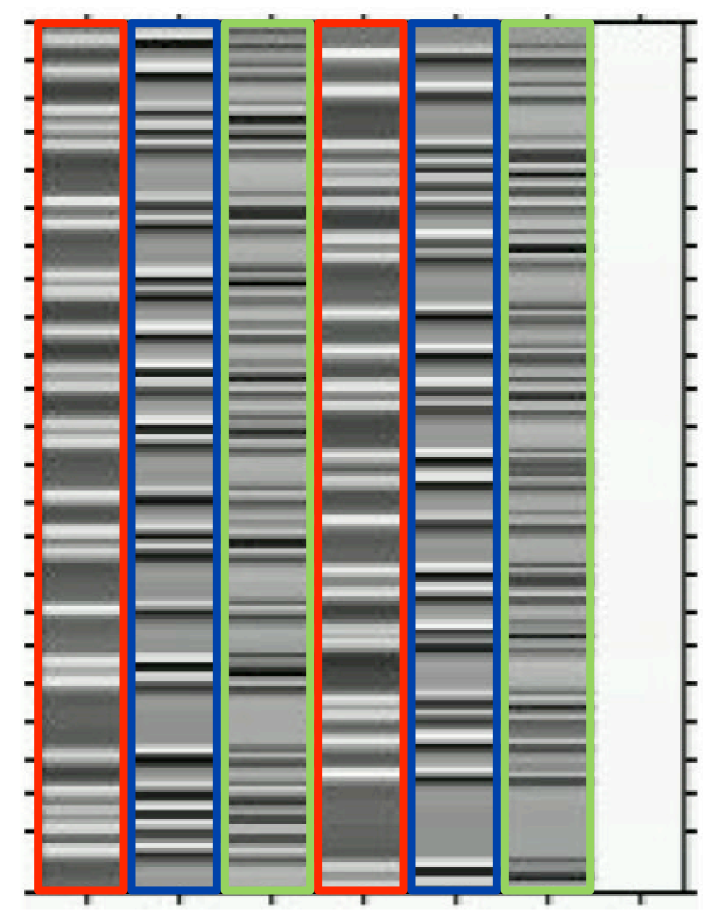
$$h(t + \Delta t; w + \Delta w) \approx \boxed{h(t; w)} + \Delta t \boxed{\frac{\partial h}{\partial \Delta t}} + \Delta w \boxed{\frac{\partial h}{\partial \Delta w}}$$




$\otimes$

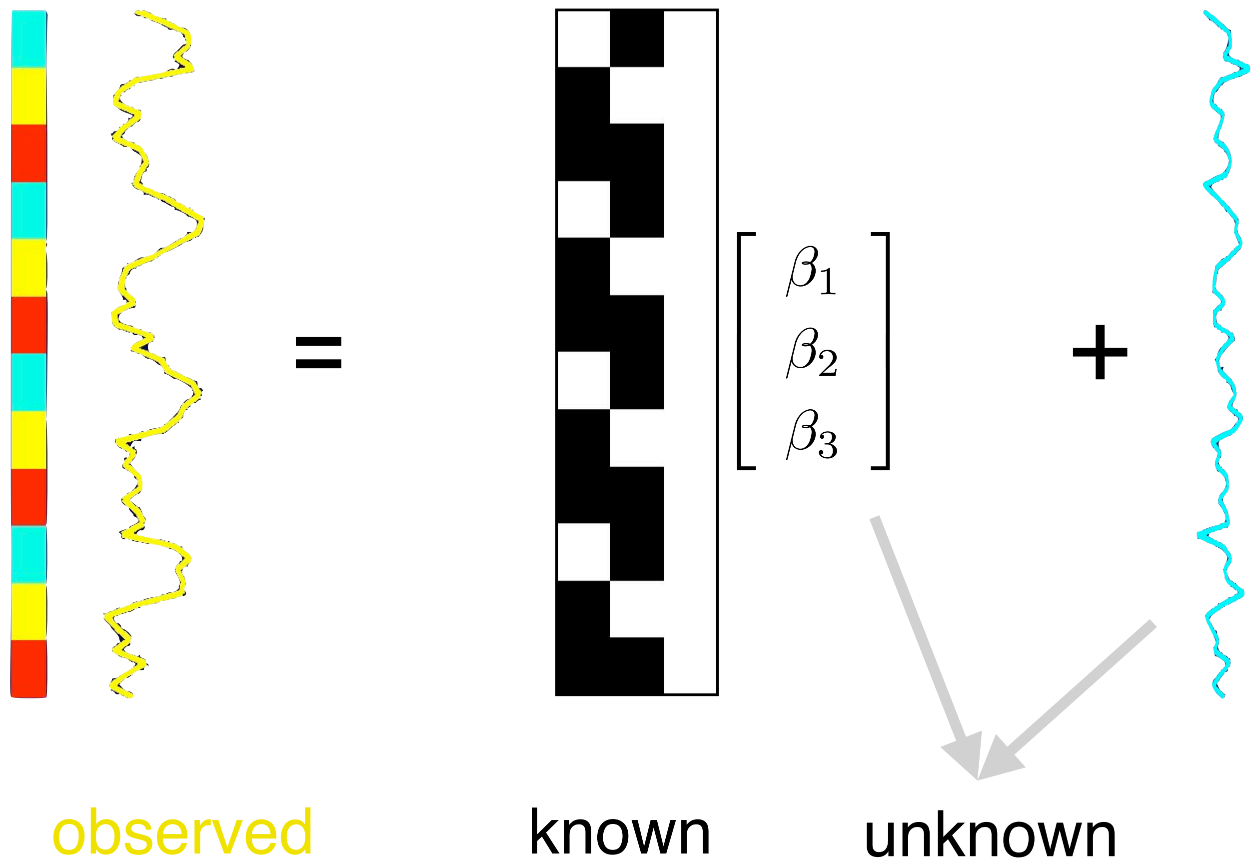



$=$



- More involved techniques using Volterra kernels

# Parameter estimation




- General Linear Model (GLM)

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$$

- Ordinary least-squares (OLS) estimator

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \arg\min_{\boldsymbol{\beta}} \sum_k \hat{e}_k^2 = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \\ \text{cov}(\hat{\boldsymbol{\beta}}) &= \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1} \end{aligned}$$

# Parameter estimation

- Gauss-Markov conditions:
  - "sphericity": $\boldsymbol{e}$ is independently and identically distributed as $N(0, \sigma^2 I)$
  - effects of $\boldsymbol{X}$ are independent of $\boldsymbol{e}$; i.e., deterministic and known
  - effects of $\boldsymbol{X}$ are linearly independent; $L = \mathrm{rank}(\boldsymbol{X})$

then the OLS estimator is

- minimum variance linear unbiased estimator (BLUE)
- $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ is *asymptotically* normal $N(0, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$
- any linear combination $\boldsymbol{c}^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/s$, where $s = \sqrt{\frac{\hat{\mathbf{e}}^T\hat{\mathbf{e}}\ \boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{c}}{N-L}}$, follows Student t-distribution with $N - L$ degrees of freedom

# Parameter estimation

- Gauss-Markov conditions:
  - "sphericity": $\boldsymbol{e}$ is independently and identically distributed as $N(0, \sigma^2 I)$
  - effects of $\boldsymbol{X}$ are independent of $\boldsymbol{e}$; i.e., deterministic and known
  - effects of $\boldsymbol{X}$ are linearly independent; $L = \mathrm{rank}(\boldsymbol{X})$

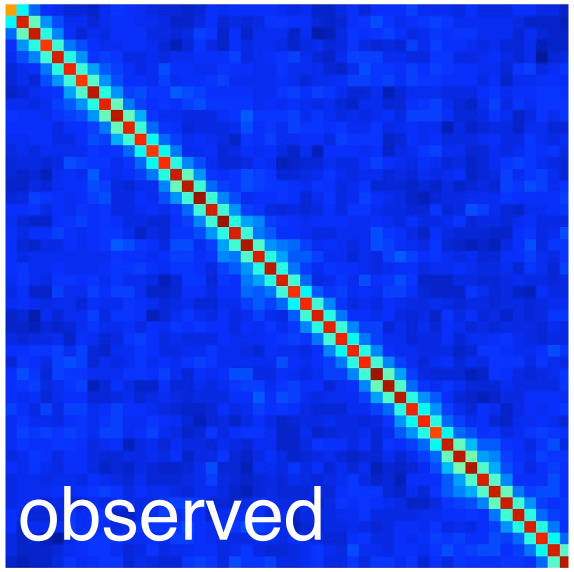
then the OLS estimator is

- minimum variance linear unbiased estimator (BLUE)
- $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ is *asymptotically* normal $N(0, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$
- any linear combination $\boldsymbol{c}^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/s$, where $s = \sqrt{\frac{\hat{\mathbf{e}}^T\hat{\mathbf{e}}\ \boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{c}}{N-L}}$, follows Student t-distribution with $N - L$ degrees of freedom

[Worsley et al., 1996]

# Parameter estimation

- Noise in fMRI is serially correlated
  - scanner (low-frequency) drifts
  - (aliased) cardiac and respiratory pulsations
  - residual movement artifacts




consequently, OLS remains unbiased, but variance will be biased

- Common solution: AR(1) + white noise

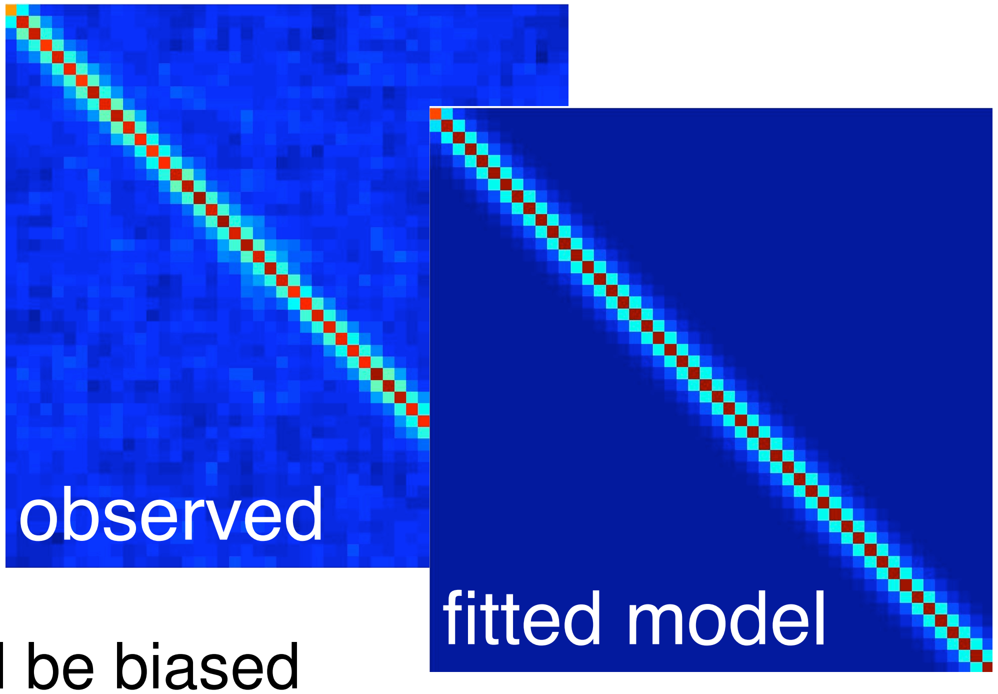
$$\begin{aligned} e_k &= z_k + n_k^{(1)}, \quad \text{where } n_k^{(1)} \text{ follows } N(0, \sigma_1^2) \\ z_k &= a z_{k-1} + n_k^{(2)}, \quad \text{where } n_k^{(2)} \text{ follows } N(0, \sigma_2^2) \end{aligned}$$

  - “prewhitening”: ReML integrates estimation (parameters and $\boldsymbol{K}$, where $\boldsymbol{K}\boldsymbol{K}^T = \hat{\boldsymbol{\Sigma}}$)
  - $v$: adjusted degrees of freedom (Satterthwaite)

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\boldsymbol{X}^T \boldsymbol{K}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{K}^{-1} \boldsymbol{y} \\ \text{cov}(\hat{\boldsymbol{\beta}}) &= \sigma^2 (\boldsymbol{X}^T \boldsymbol{K}^{-1} \boldsymbol{X})^{-1} \end{aligned}$$

# Parameter estimation

- Noise in fMRI is serially correlated
  - scanner (low-frequency) drifts
  - (aliased) cardiac and respiratory pulsations
  - residual movement artifacts




consequently, OLS remains unbiased, but variance will be biased

- Common solution: AR(1) + white noise

$$\begin{aligned} e_k &= z_k + n_k^{(1)}, \quad \text{where } n_k^{(1)} \text{ follows } N(0,\sigma_1^2) \\ z_k &= a z_{k-1} + n_k^{(2)}, \quad \text{where } n_k^{(2)} \text{ follows } N(0,\sigma_2^2) \end{aligned}$$

  - “prewhitening”: ReML integrates estimation (parameters and $\boldsymbol{K}$, where $\boldsymbol{K}\boldsymbol{K}^T = \hat{\boldsymbol{\Sigma}}$)
  - $v$: adjusted degrees of freedom (Satterthwaite)

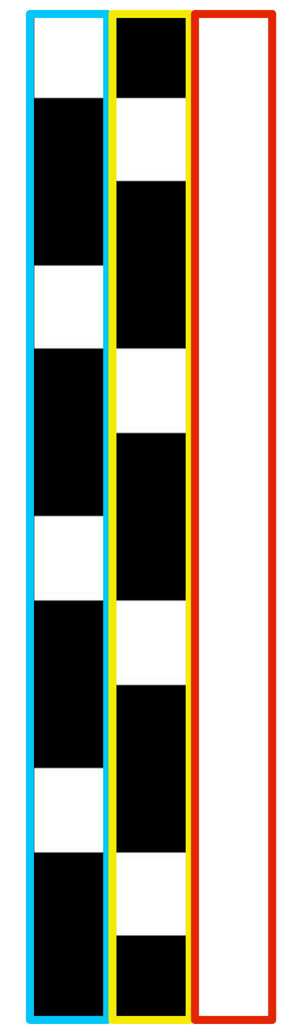
$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\boldsymbol{X}^T\boldsymbol{K}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{K}^{-1}\boldsymbol{y} \\ \text{cov}(\hat{\boldsymbol{\beta}}) &= \sigma^2(\boldsymbol{X}^T\boldsymbol{K}^{-1}\boldsymbol{X})^{-1} \end{aligned}$$

[Bullmore et al., 1996]

# Contrasts and design efficiency

- Questioning the fitted model—extracting “contrast” $\boldsymbol{c}^T\hat{\boldsymbol{\beta}}$
  - response to “faces”: $\boldsymbol{c}^T = [1\ 0\ 0]$
  - response to “faces vs objects”: $\boldsymbol{c}^T = [1\ \ -1\ 0]$

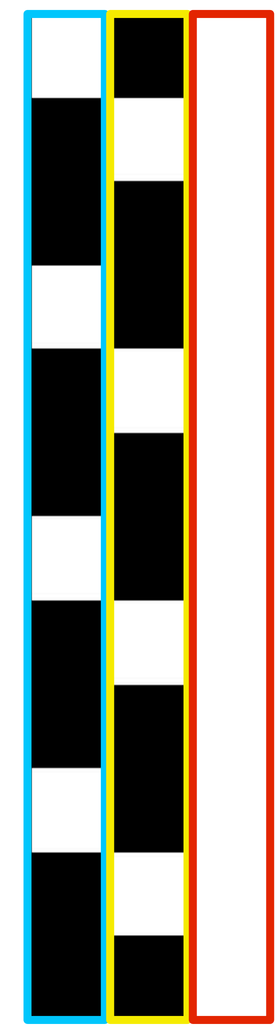

- Design efficiency (Gauss-Markov assumptions)
  - covariance matrix of parameters: $\mathrm{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$
  - covariance matrix of selected parameters: $\mathrm{cov}(\boldsymbol{C}^T\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{C}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{C}$
  - efficiency $\propto \mathrm{trace}\left((\boldsymbol{C}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{C})^{-1}\right)$
  - optimize timing, number of events, stochastic design, ...



# Contrasts and design efficiency

- Questioning the fitted model—extracting "contrast" $\boldsymbol{c}^T\hat{\boldsymbol{\beta}}$
  - response to "faces": $\boldsymbol{c}^T = [1\ 0\ 0]$
  - response to "faces vs objects": $\boldsymbol{c}^T = [1\ \ -1\ 0]$


- Design efficiency (Gauss-Markov assumptions)
  - covariance matrix of parameters:
    $\mathrm{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$
  - covariance matrix of selected parameters:
    $\mathrm{cov}(\boldsymbol{C}^T\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{C}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{C}$
  - efficiency $\propto \mathrm{trace}\left((\boldsymbol{C}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{C})^{-1}\right)$
  - optimize timing, number of events, stochastic design, ...



[Henson et al., 1999]

# Hypothesis testing

- Null hypothesis expresses "no effect" (i.e., true $\boldsymbol{c}^T\boldsymbol{\beta}$ is $\mathbf{0}$)

$$\mathcal{H}_0 : E[\boldsymbol{c}^T\hat{\boldsymbol{\beta}}] = 0$$

  - $\hat{t} = \boldsymbol{c}^T\hat{\boldsymbol{\beta}} / \sqrt{\frac{\hat{\mathbf{e}}^T\hat{\mathbf{e}}\ \boldsymbol{c}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{c}}{v}}$ follows Student t-distribution assuming $\mathcal{H}_0$
  - reject $\mathcal{H}_0$ if $\hat{t} \geq T$, where the $\alpha$-level is the acceptable false positive rate:
    $\alpha = P(t \geq T)$ (one-sided t-test)
  - $p$-value indicates the assessment of $\hat{t}$ assuming $\mathcal{H}_0$:
    $p = P(t \geq \hat{t})$
  - specificity: risk of false positives (type I errors)
    sensitivity: risk of false negatives (type II errors)

- Null hypothesis acceptation/rejection controls specificity only
  - useful as "evidence of presence", not "evidence of absence" (neurosurgeon!)

# Hypothesis testing

- Mass univariate testing ($V$=10K-100K intracranial voxels)
  - $E[\text{FP}] = \alpha V$, so false positives should be controlled adequately!

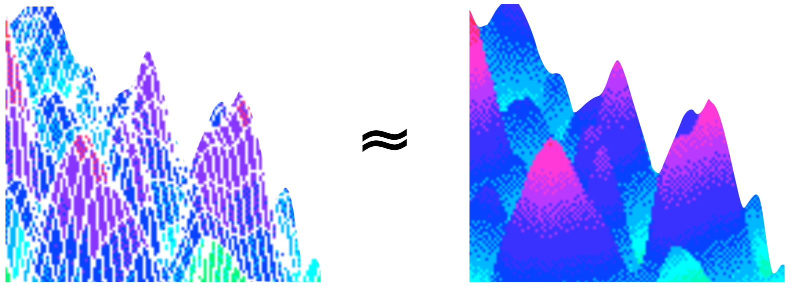
- Family-wise error rate: $\alpha_{\text{FWE}} = P(\cup_{k=1}^{V} t_k \geq T)$
- **Bonferroni correction**: assuming independent observations

$$\alpha_{\text{FWE}} = 1 - (1 - \alpha)^V \approx \alpha V$$

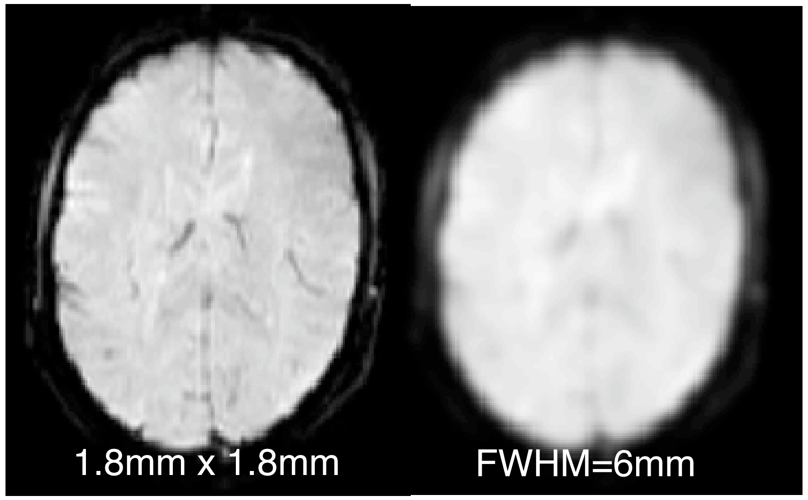
  - to obtain $\alpha_{\text{FWE}}$, use $\alpha_{\text{FWE}}/V$ at the individual tests
  - high specificity, low sensitivity since neglecting spatial correlation therefore, too conversative
  - can be applied locally (if ROI is chosen a priori)

# Gaussian random field theory

- Consider contrast as lattice representation of continuous Gaussian random field




  - spatially smooth data with 3D Gaussian kernel, typical full-width-at-half-maximum (FWHM) about 6–12 mm




- Euler characteristic:

$\xi_T$=topological measure $\#$blobs $-$ $\#$holes

- Assuming $\mathcal{H}_0$ and high $T$, we have

$$
\begin{aligned}
P(\cup_{k=1}^{V} t_k > T) &= P(\max_k(t_k) > T) \\
&= P(\text{one or more blobs}) \\
&\approx P(\xi_T \geq 1) && \text{(no holes)} \\
&\approx E[\xi_T] && \text{(one blob)}
\end{aligned}
$$

# Gaussian random field theory

- Expected Euler characteristic can be further approximated

$$
\begin{aligned}
P(\cup_{k=1}^{V} t_k > T) &\approx E[\xi_T] \\
&\approx \frac{\lambda\sqrt{|\mathbf{\Lambda}|}\ (T^2-1)\exp(-T^2/2)}{(2\pi)^2}
\end{aligned}
$$

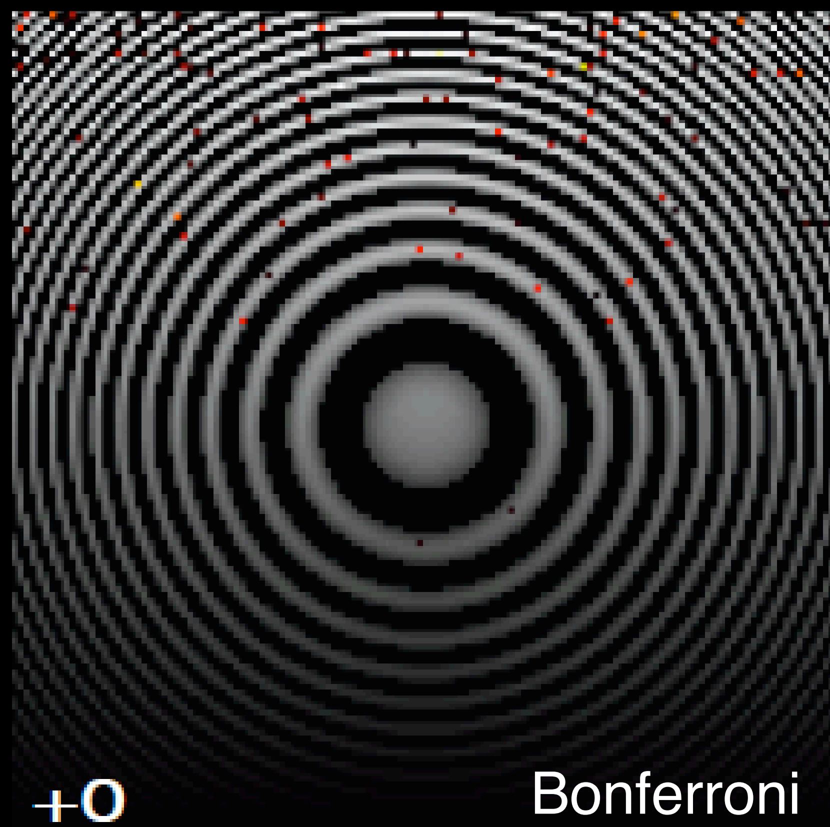
  - $\lambda$: volume of search region
  - $\mathbf{\Lambda}$: spatial roughness matrix = covariance matrix of $\partial I/\partial\ x, y, z$
    $\sqrt{|\mathbf{\Lambda}|} = \frac{(4\log 2)^{3/2}}{\mathrm{FWHM}_x\mathrm{FWHM}_y\mathrm{FWHM}_z}$: estimate of local smoothness
    where $\mathrm{FWHM}$ characterizes Gaussian kernel that makes white noise like $\mathbf{\Lambda}$
    "RESolution ELement", $1\ \mathrm{RESEL} = \mathrm{FWHM}_x\mathrm{FWHM}_y\mathrm{FWHM}_z$

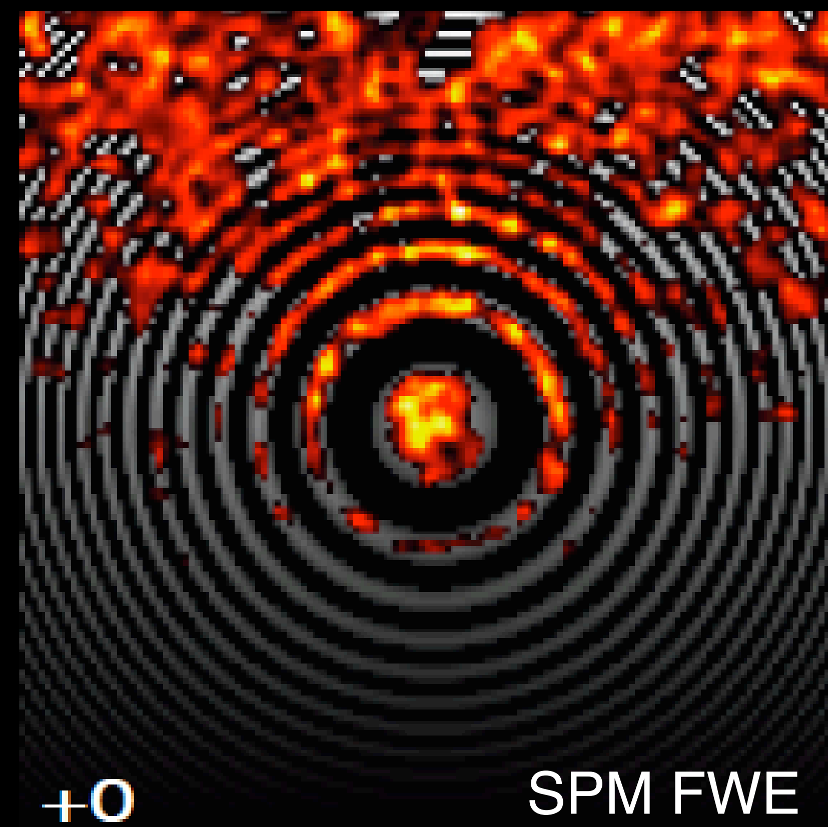
- Finding threshold $T$ for desired $\alpha$-level involves LambertW-function

# Gaussian random field theory

- Advantages:
  - increased sensitivity
  - decreased inter-subject variability (group studies!)
- Limitations:
  - requires sufficient smoothness
    - high $v$: FWHM like $3 - 4\times$ voxel size
    - low $v$: FWHM rather $10\times$ voxel size
  - smoothness needs to be estimated
    - bias if not sufficiently smooth
  - several approximations in cascade (high $T$)

# 5% corrected



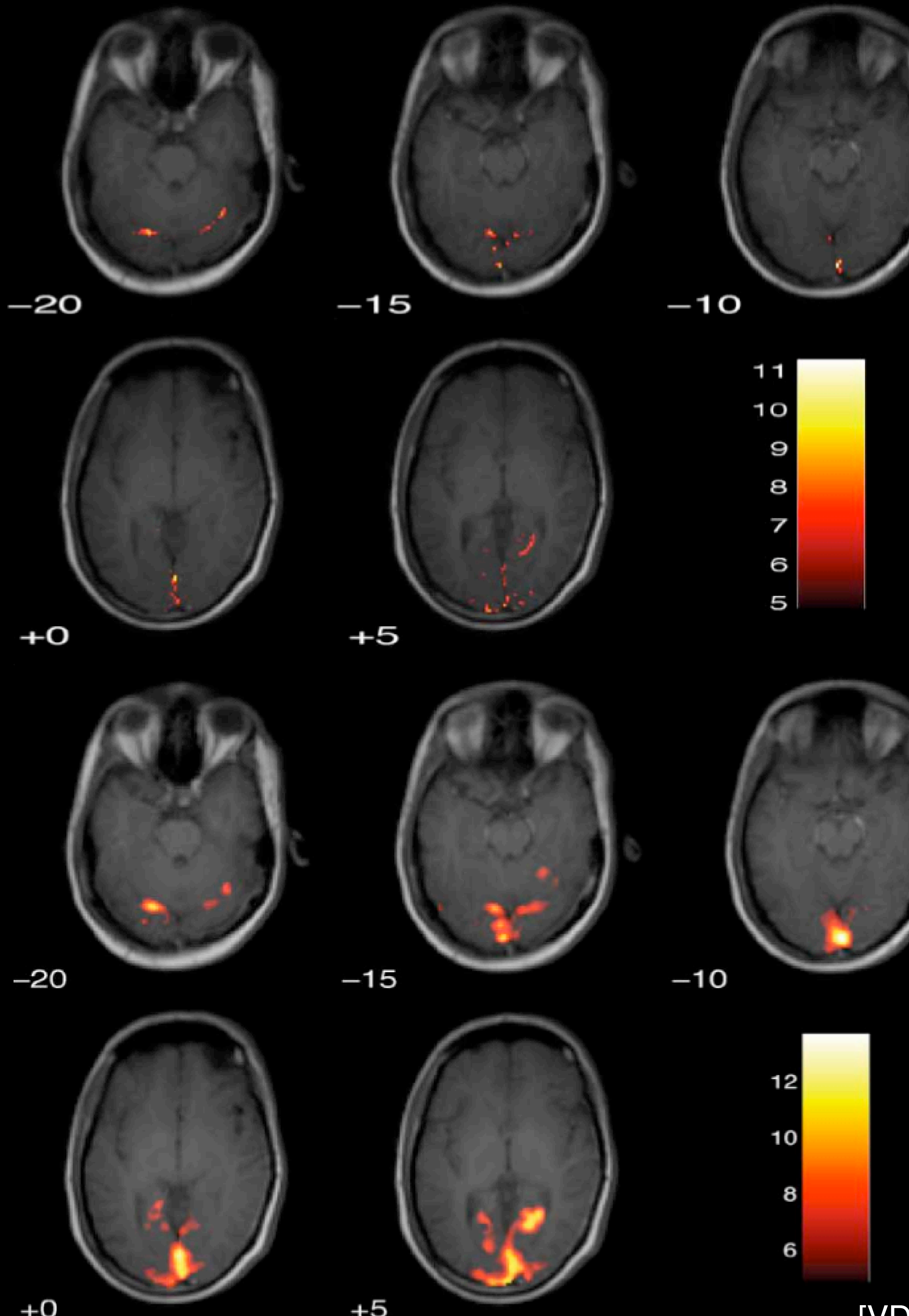




[VDV et al., IEEE JSTSP, 2008]

# 5% corrected

Bonferroni

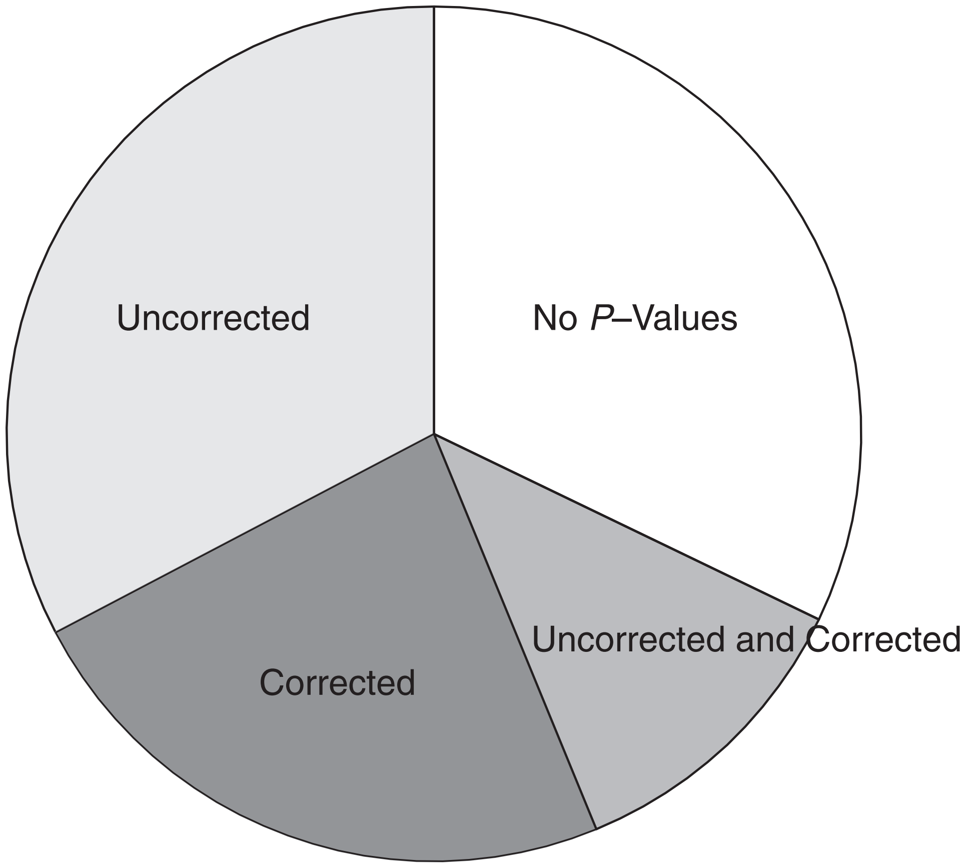
SPM FWE




[VDV et al., IEEE JSTSP, 2008]

# To correct or not to correct...

- Dealing with multiple hypothesis testing
  - weak type I error control (no correction)
    - only for omnibus-test (“something *anywhere*?”)
  - for localizing power, strong type I error control
    - Bonferroni
      - too conservative
    - GRF theory
      - requires smoothing
  - alternative error rate (FDR,...)
    - be more permissive as #detections increases
- If left uncorrected...
  - significance of results?
  - recent controversy (independent tests)
    “*Voodoo correlations in social neurosciences*” (Vul et al.)




from [Nielsen et al, IEEE Engineering in Medicine and Biology Magazine, 2006]

# Hypothesis testing - F test

- Partitioning into two blocks of regressors
- Consider reduced model by design matrix $\boldsymbol{X}_0$

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \\ \boldsymbol{y}_0 &= \boldsymbol{X}_0\boldsymbol{\beta}_0 + \boldsymbol{e}_0 \end{aligned}$$

- Null hypothesis $\mathcal{H}_0$ expresses “no improvement of $\boldsymbol{X}$ over $\boldsymbol{X}_0$”

$$\hat{F} = \frac{\frac{\hat{\mathbf{e}}^T\hat{\mathbf{e}} - \hat{\boldsymbol{e}_0}^T\hat{\boldsymbol{e}_0}}{L-L_0}}{\frac{\hat{\mathbf{e}}^T\hat{\mathbf{e}}}{N-L}}$$

  - follows F-distribution $(L - L_0, N - L)$ assuming $\mathcal{H}_0$
  - reject $\mathcal{H}_0$ if $\hat{F} \geq T$, where the $\alpha$-level is the acceptable false positive rate: $\alpha = P(F \geq T)$
  - two-sided test, one-sided extension [Calhoun, 2004; Worsley, 2006]

# Hypothesis testing - F test

- More flexible F-test by contrast matrix
  - reduced model can be made up by linear combinations of regressors
  - avoid reparametrization of model

$$\hat{F} = \frac{\boldsymbol{y}^T \boldsymbol{M} \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{R} \boldsymbol{y}} \frac{N-L}{L_a} = \frac{\hat{\boldsymbol{\beta}}^T \boldsymbol{X}^T \boldsymbol{M} \boldsymbol{X} \hat{\boldsymbol{\beta}}}{\boldsymbol{y}^T \boldsymbol{R} \boldsymbol{y}} \frac{N-L}{L_a} \sim F(L_a, N-L)$$

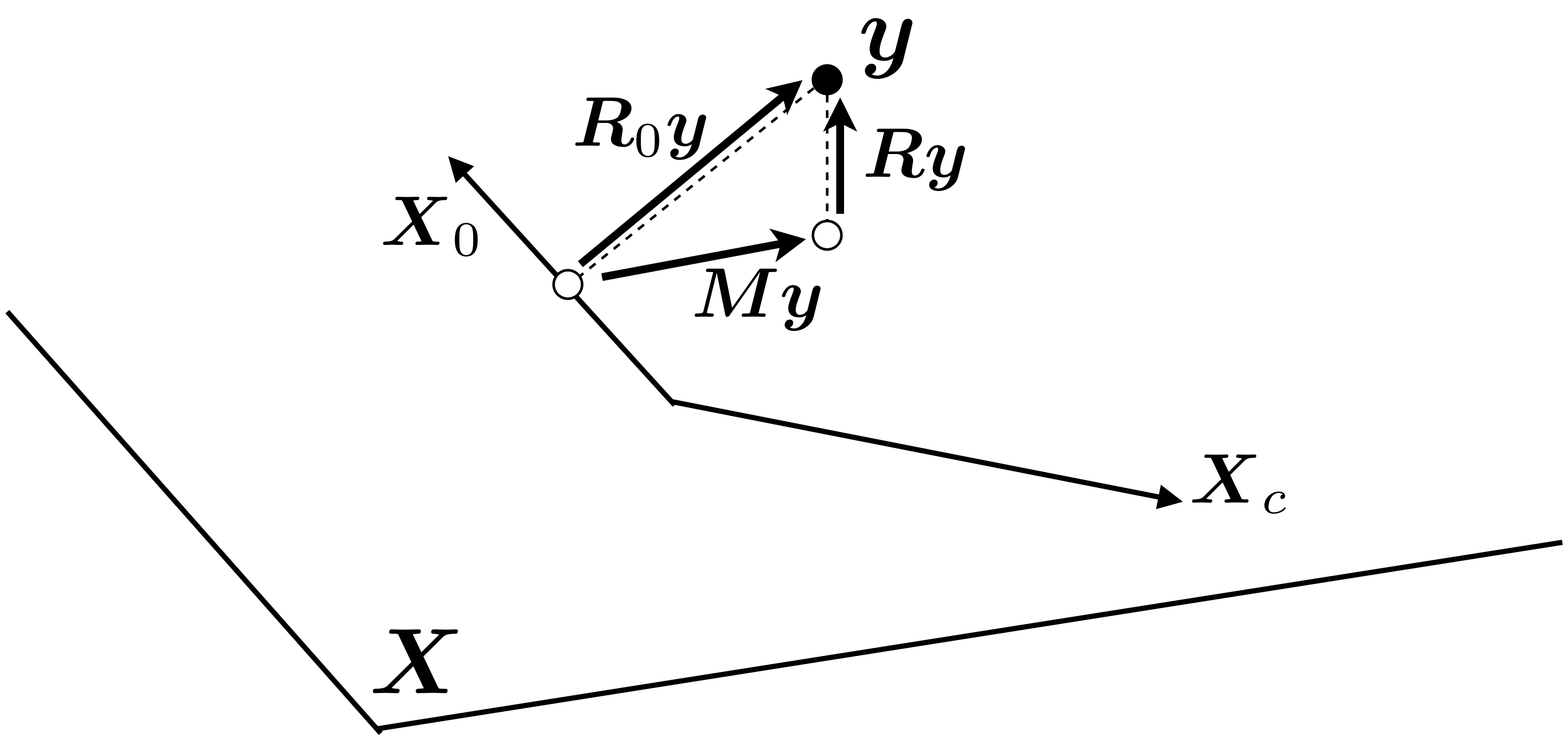
$\boldsymbol{X}_c = \boldsymbol{X}\boldsymbol{C}$

$\boldsymbol{X}_0 = \boldsymbol{X}\boldsymbol{C}_0$ where $\boldsymbol{C}_0 = \boldsymbol{I}_L - \boldsymbol{C}\boldsymbol{C}^-$

$\boldsymbol{R} = \boldsymbol{I}_N - \boldsymbol{X}\boldsymbol{X}^-$

$\boldsymbol{R}_0 = \boldsymbol{I}_N - \boldsymbol{X}_0\boldsymbol{X}_0^-$

$\boldsymbol{M} = \boldsymbol{R}_0 - \boldsymbol{R}$



# Hypothesis testing - F test

- Use of F-test
  - interested in activation for “faces or objects”; some arrogant voxel is activating during “faces”, deactivating during “objects”

    ✘ – $t$-contrast: $\boldsymbol{c}^T = [1\,0\,0\,1\,0\,0\,0]$

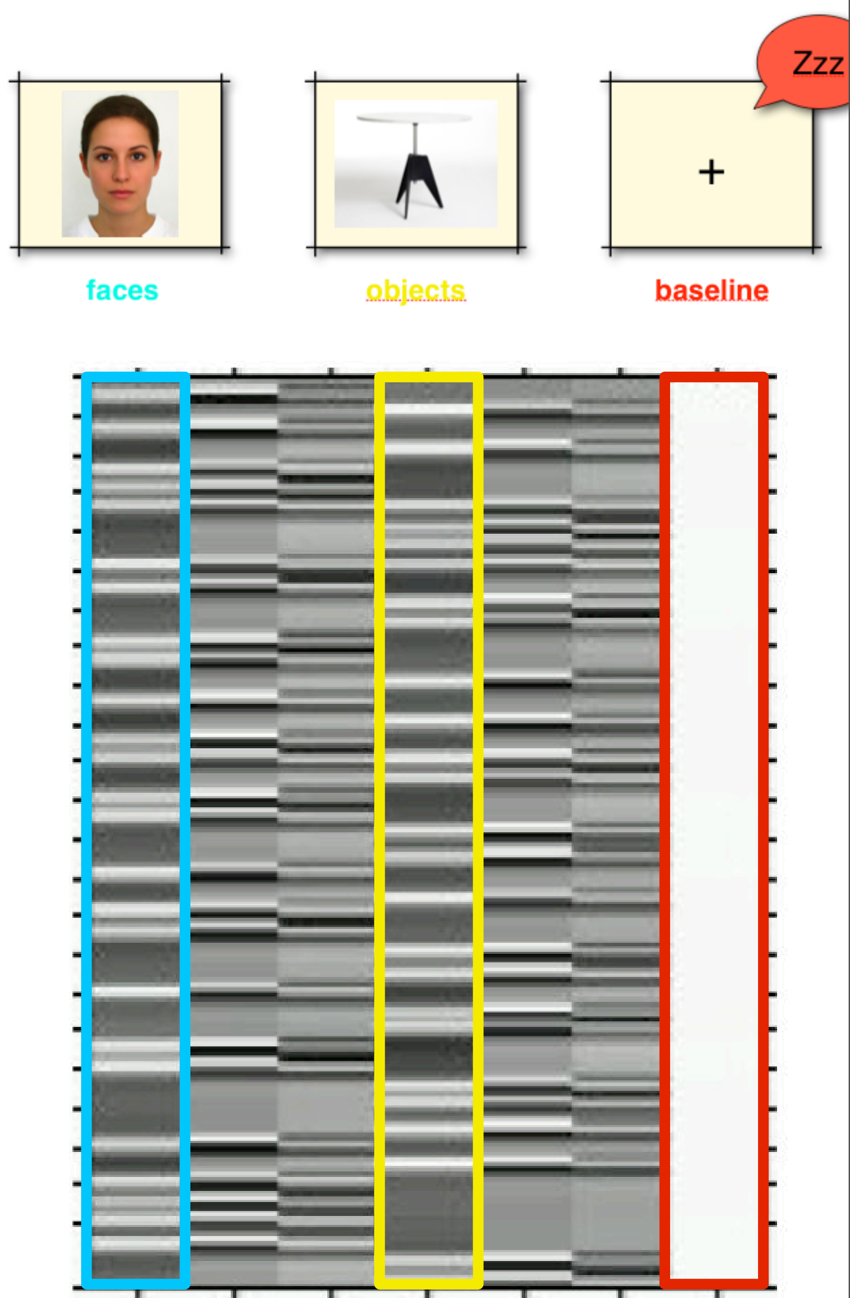
    ✔ – $F$-contrast: $\boldsymbol{C}^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$

  - use of canonical hemodynamic response function with derivatives any activation for “objects”

    ✔ – $F$-contrast: $\boldsymbol{C}^T = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$

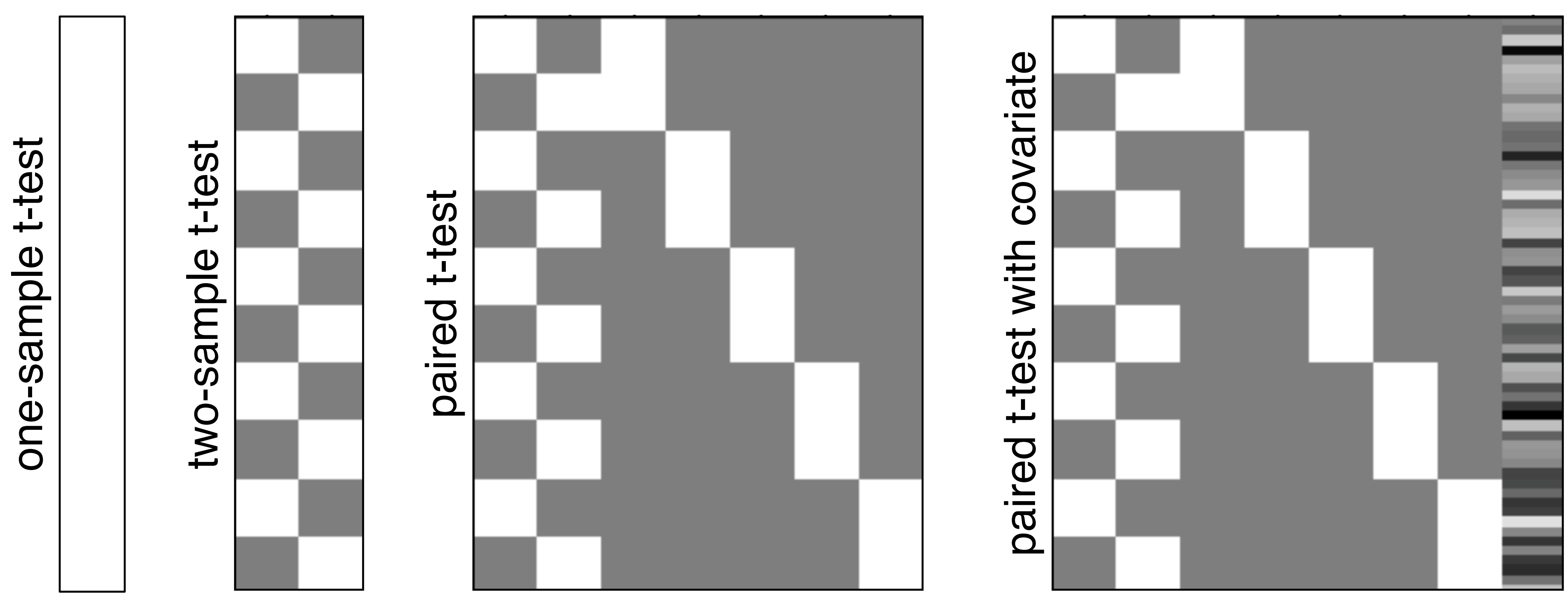
  - any difference between three conditions (like ANOVA)

    ✔ – $F$-contrast: $\boldsymbol{C}^T = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$



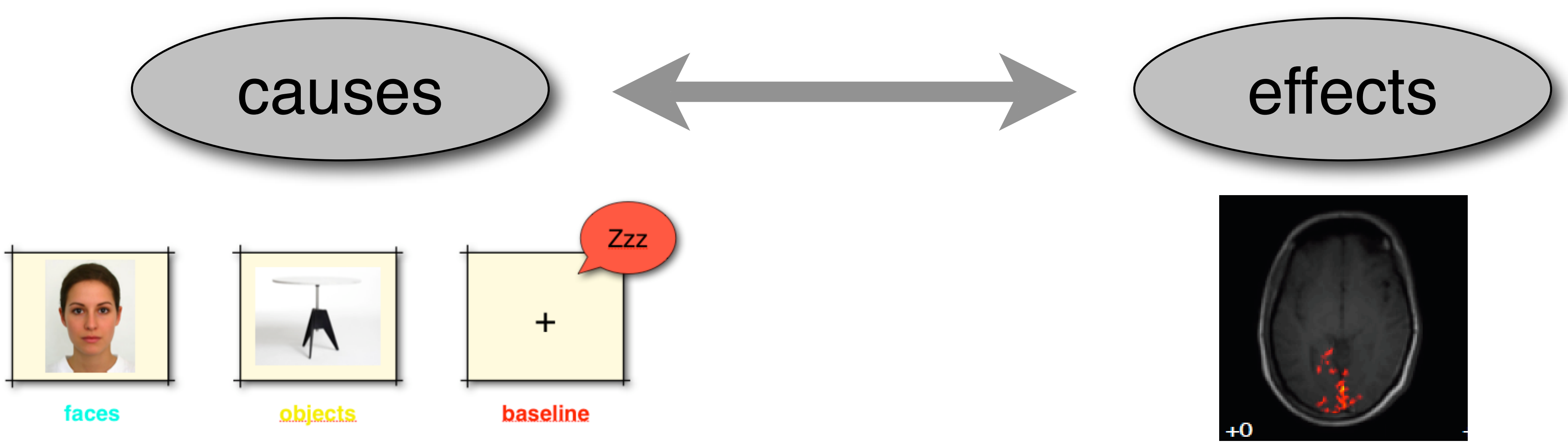
# Multi-subjects analysis

- Fixed effects analysis
  - concatenate data and design matrices of subjects
  - inference on the observed group
- Random effects analysis
  - estimate contrast of interest for individual subjects (1st level)
  - enter contrast in “basic model” and re-estimate (2nd level)
  - inference on the population from which group is sample

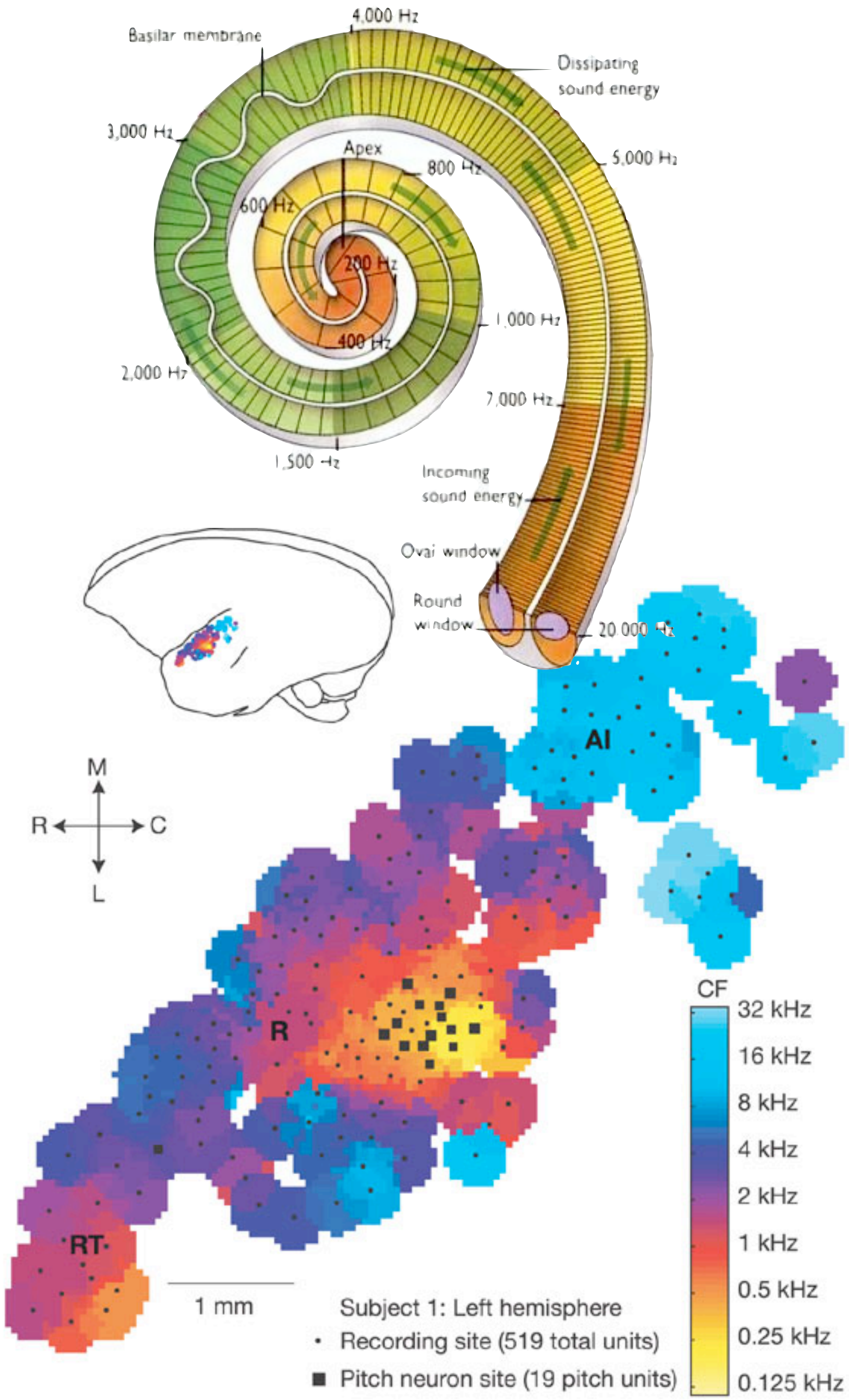


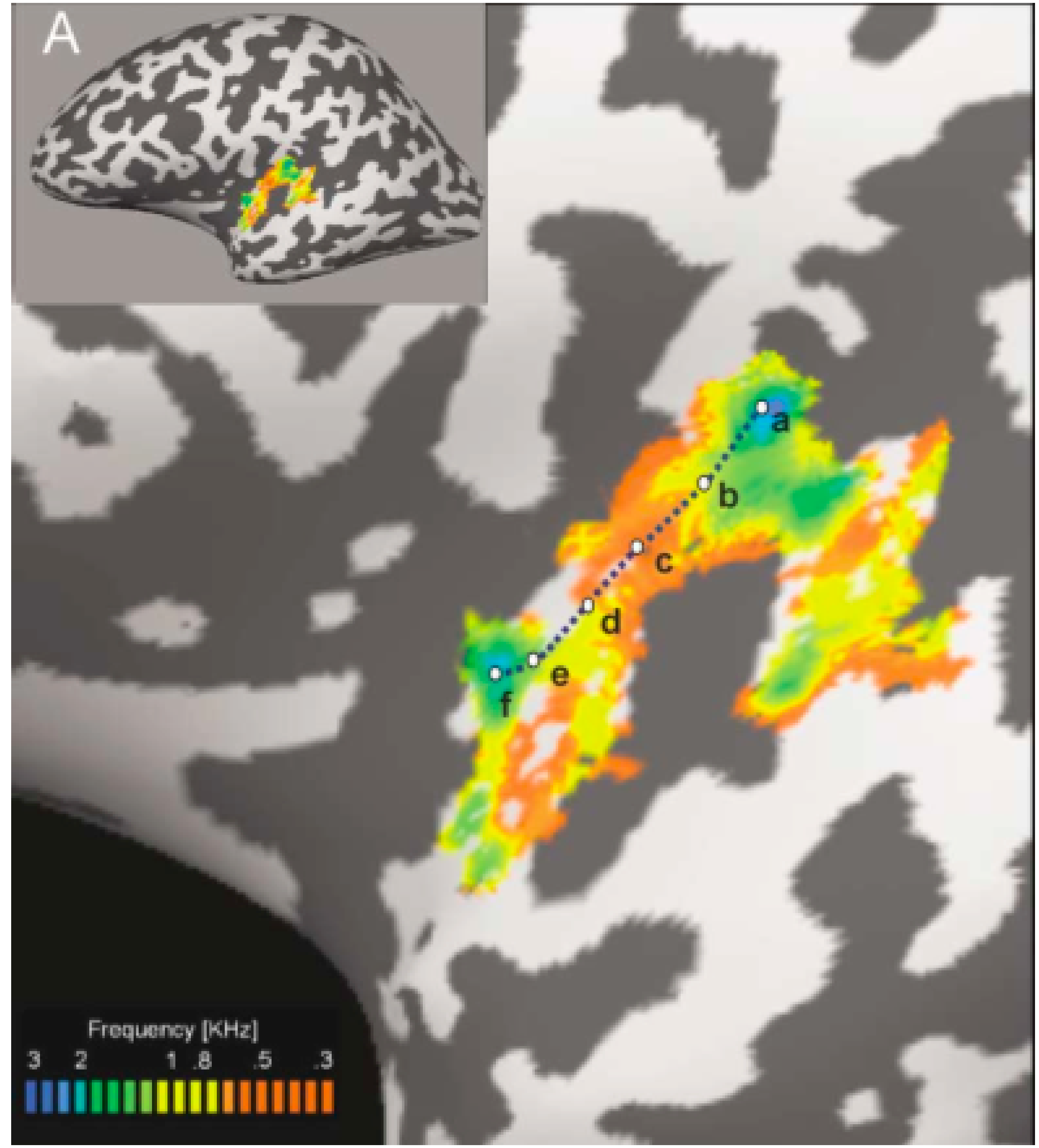
# Forward and backward inference

- Generative model
  - causes explain effects
  - GLM, forward inference
- Recognition model
  - effects identify causes (“mind reading”)
  - machine learning, backward inference
    - [Haxby, 2001; Haynes, 2005]
  - avenue to (local) multivariate analysis (exploit voxels’ interplay)
    - searchlight framework [Kriegeskorte, 2006]



# Tonotopic organization

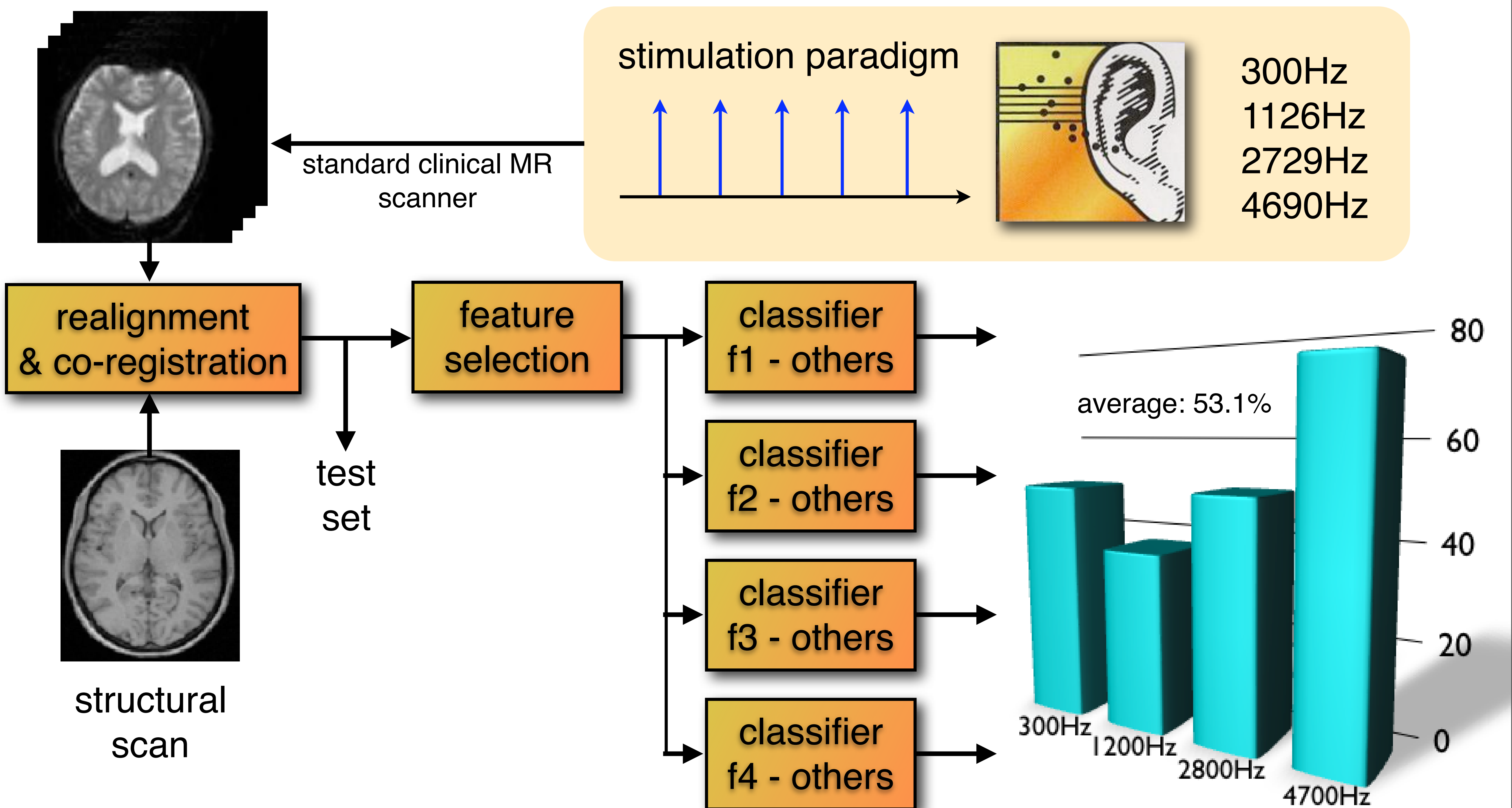


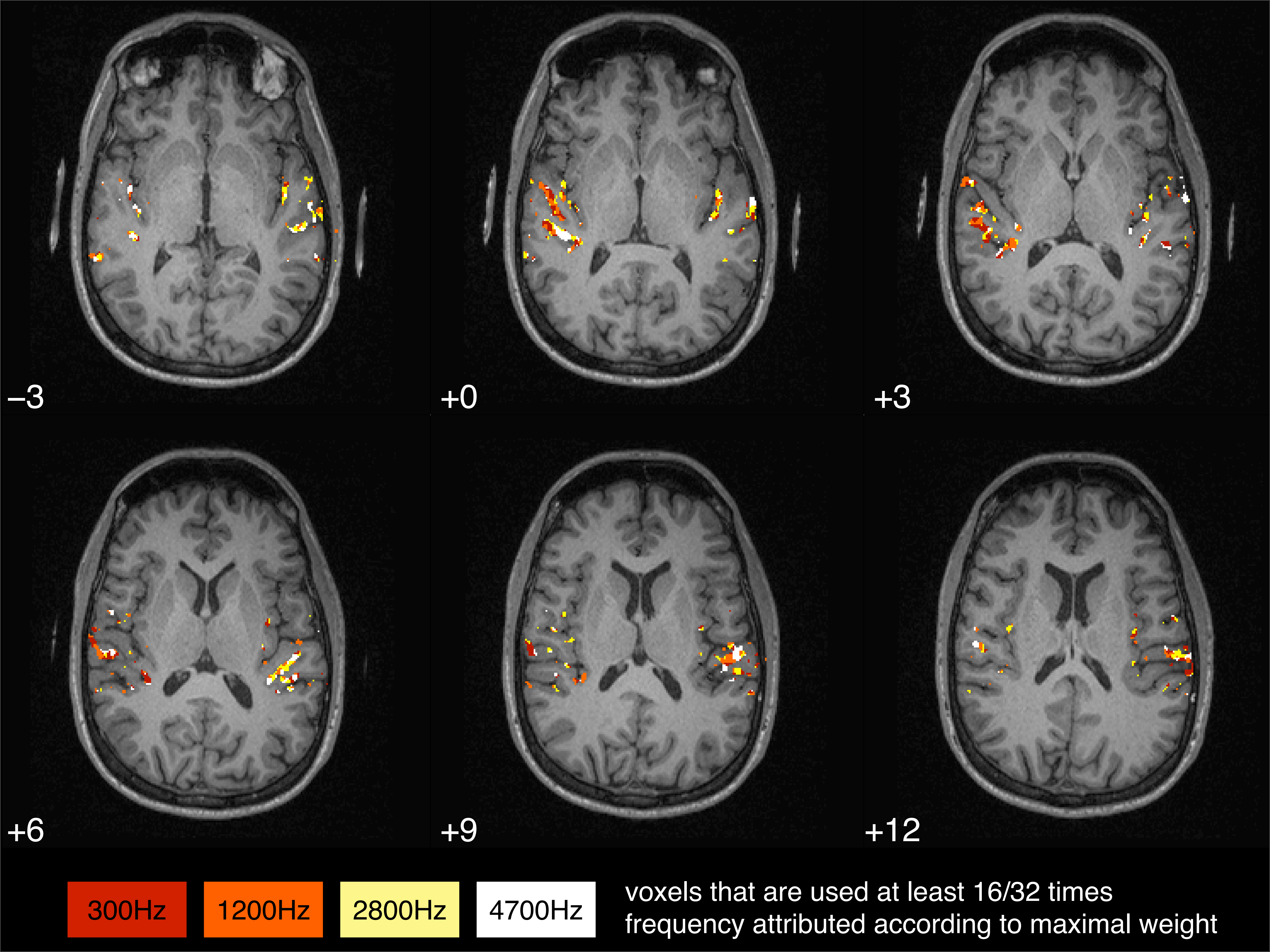

[Bendor and Wang, Nature, 2005]

[Formisano et al., Neuron, 2003]

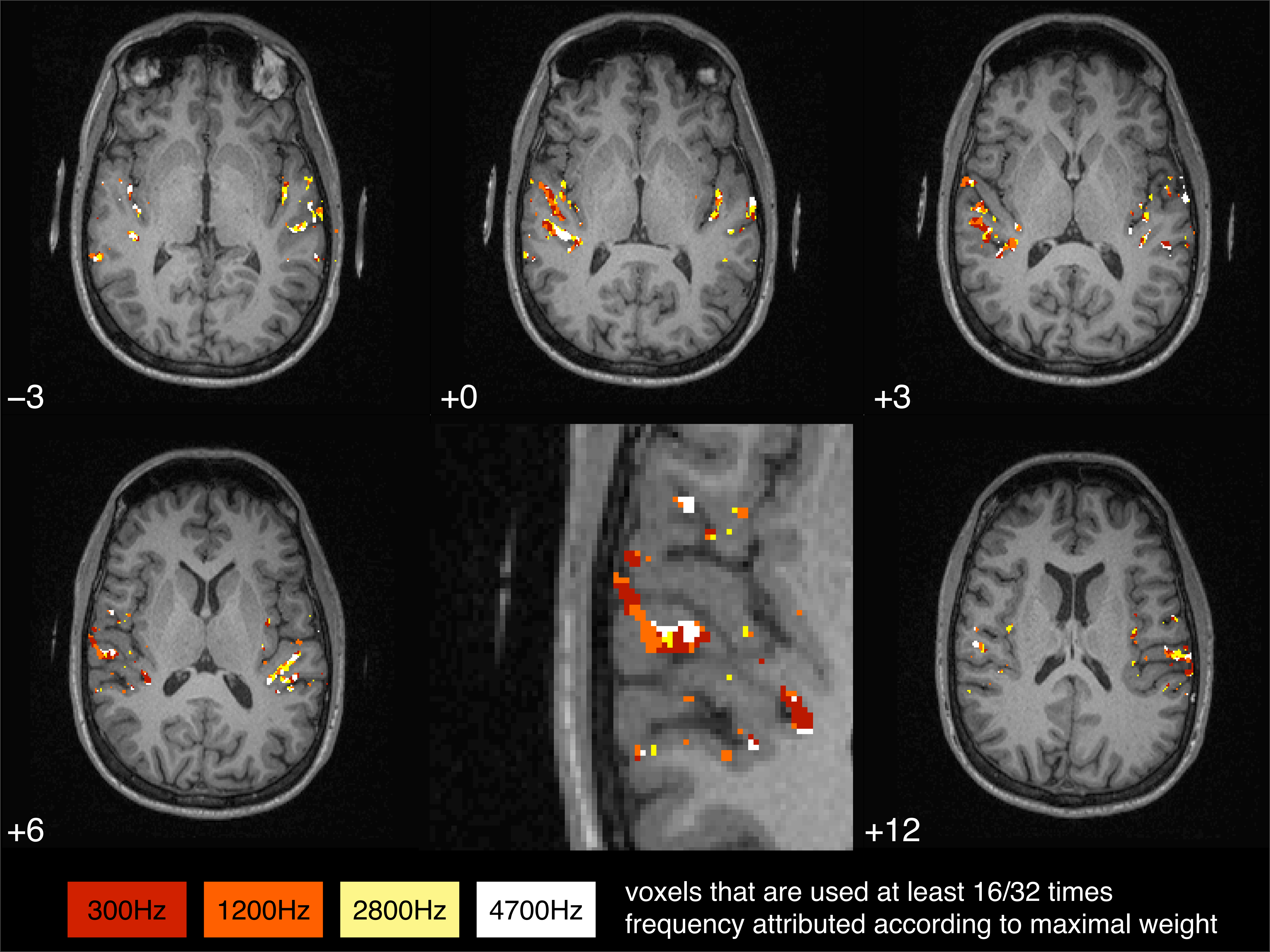
# Tonotopic organization - fMRI



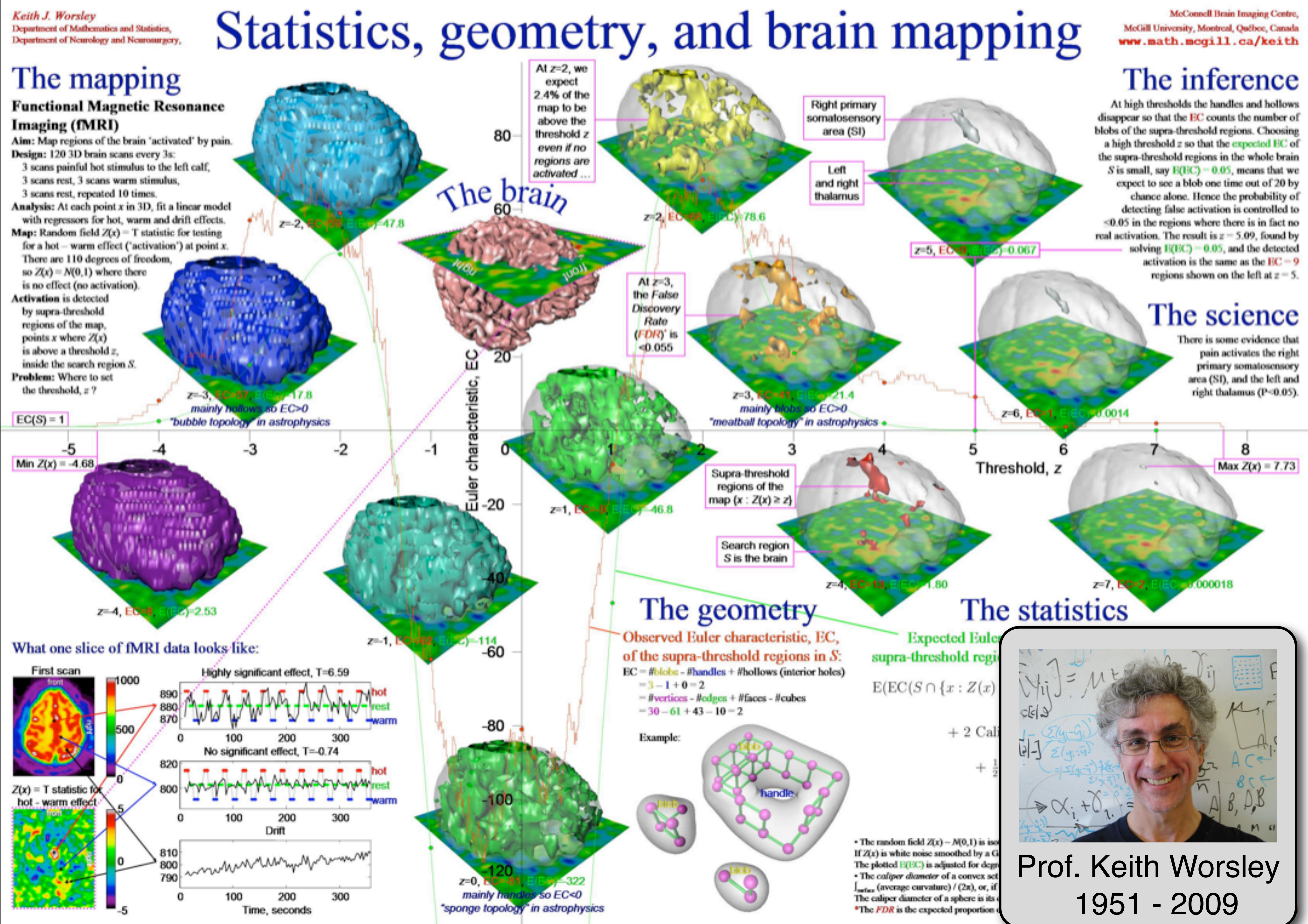

[Richiardi, VDV et al., in preparation]



−3
+0
+3
+6
+9
+12
300Hz
1200Hz
2800Hz
4700Hz
voxels that are used at least 16/32 times
frequency attributed according to maximal weight



−3
+0
+3
+6
+12
300Hz
1200Hz
2800Hz
4700Hz
voxels that are used at least 16/32 times
frequency attributed according to maximal weight




Prof. Keith Worsley
1951 - 2009

# Conclusions

- General linear model
  - conceptually simple, yet powerful and flexible
  - many tricks to “enrich” the model
  - many ways of testing the fitted parameters
    - t-test, F-test, ANOVA
  - also deals with “basic” models (2nd level analysis)
- Gaussian smoothing
  - improves sensitivity, reduces inter-subject variability
  - alternatives
    - Bayesian modeling
    - wavelet-based SPM (coming up)
- Backward inference as (local) multivariate analysis
- Exploratory analyses
  - PCA, ICA, CCA,...

# Thanks!