

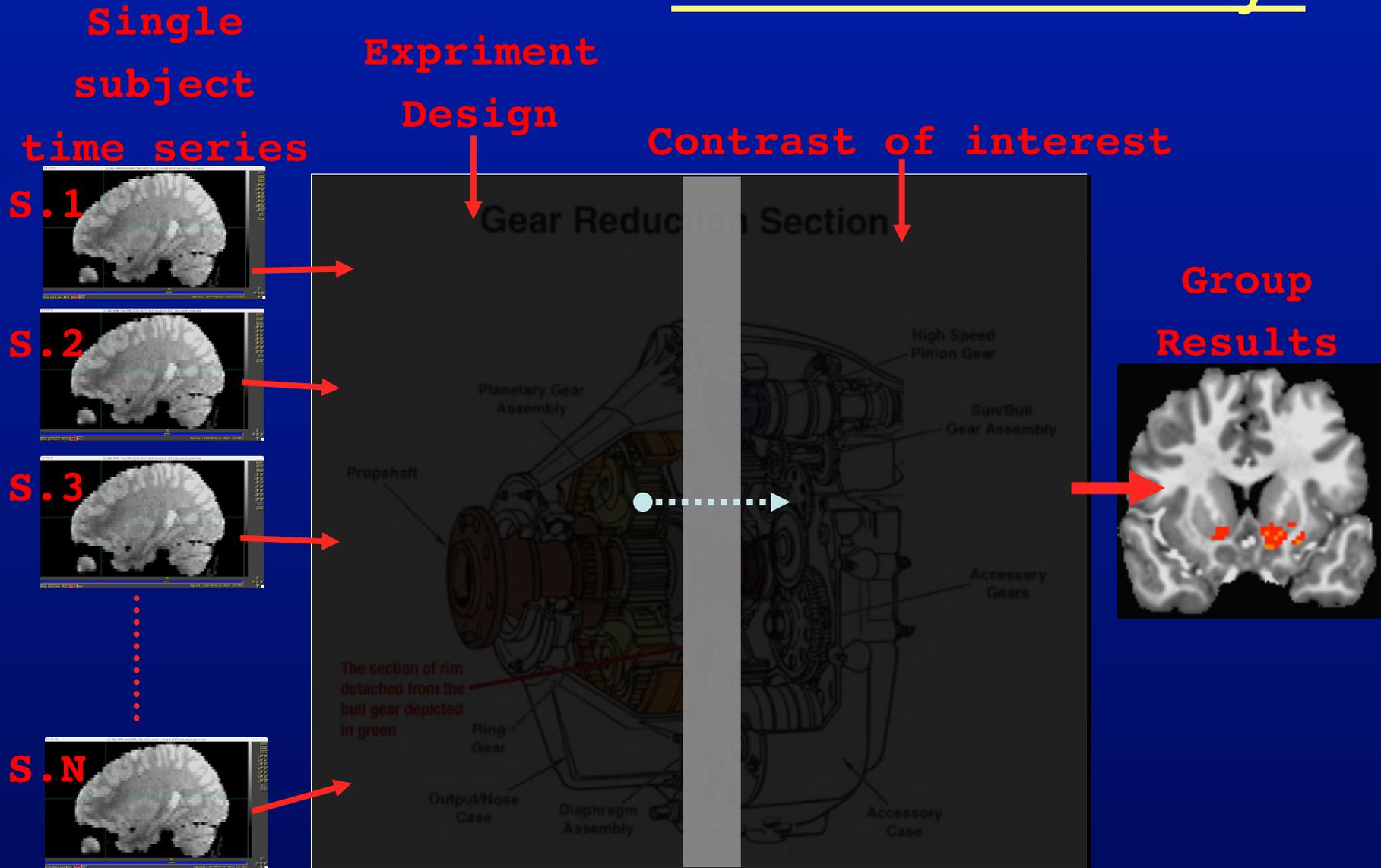
From Image-Space To Blob-Space: the processing pipeline of fMRI data

Ziad S Saad, PhD

SSCC / NIMH & NINDS / NIH / DHHS / USA /
EARTH



FMRI? It's easy!

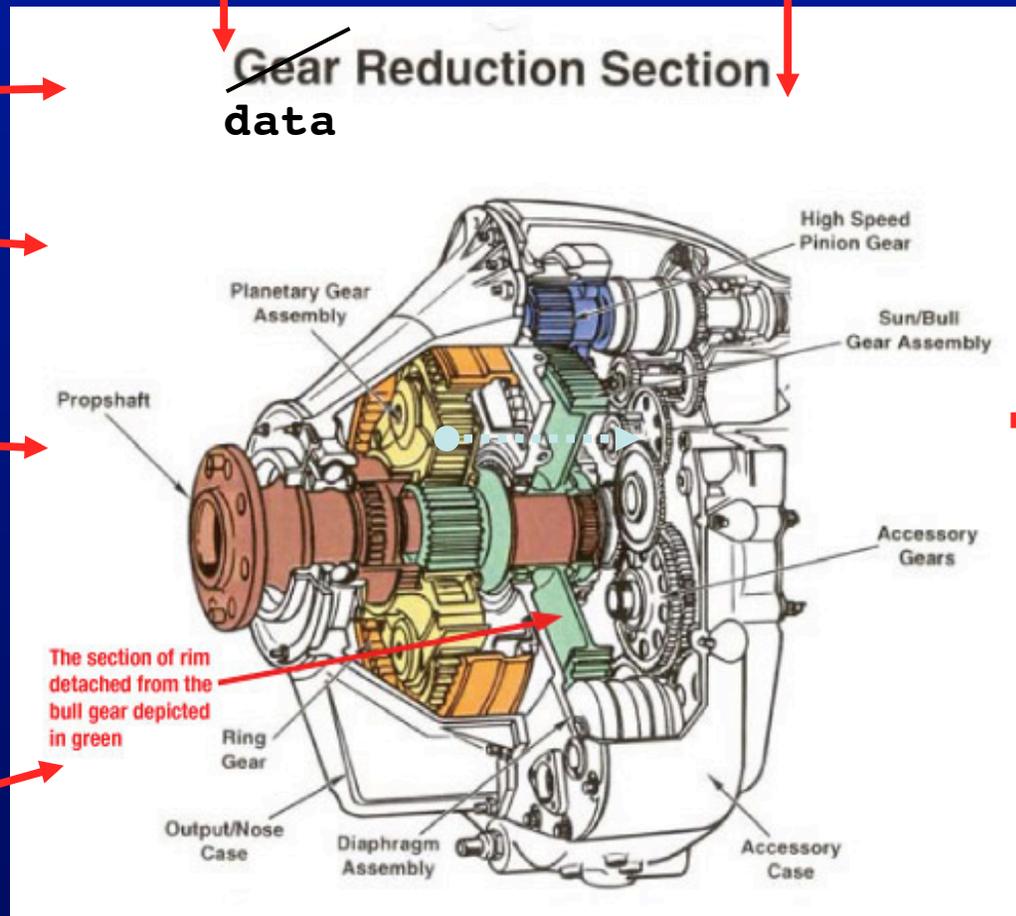
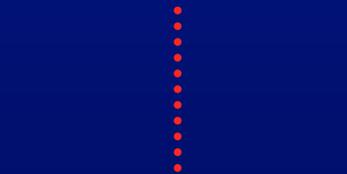
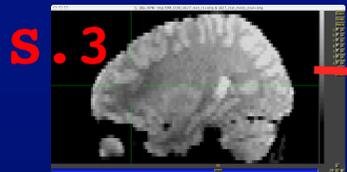
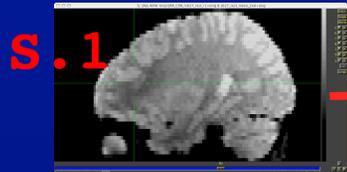


FMRI? It's easy!

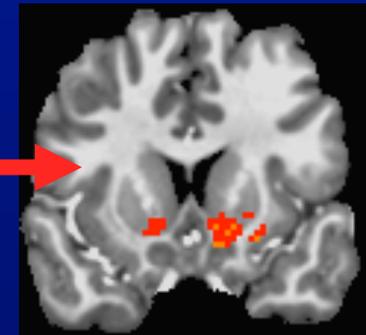
Single
subject
time series

Experiment
Design

Contrast of interest



Group
Results



FMRI Analysis

Experiment Design

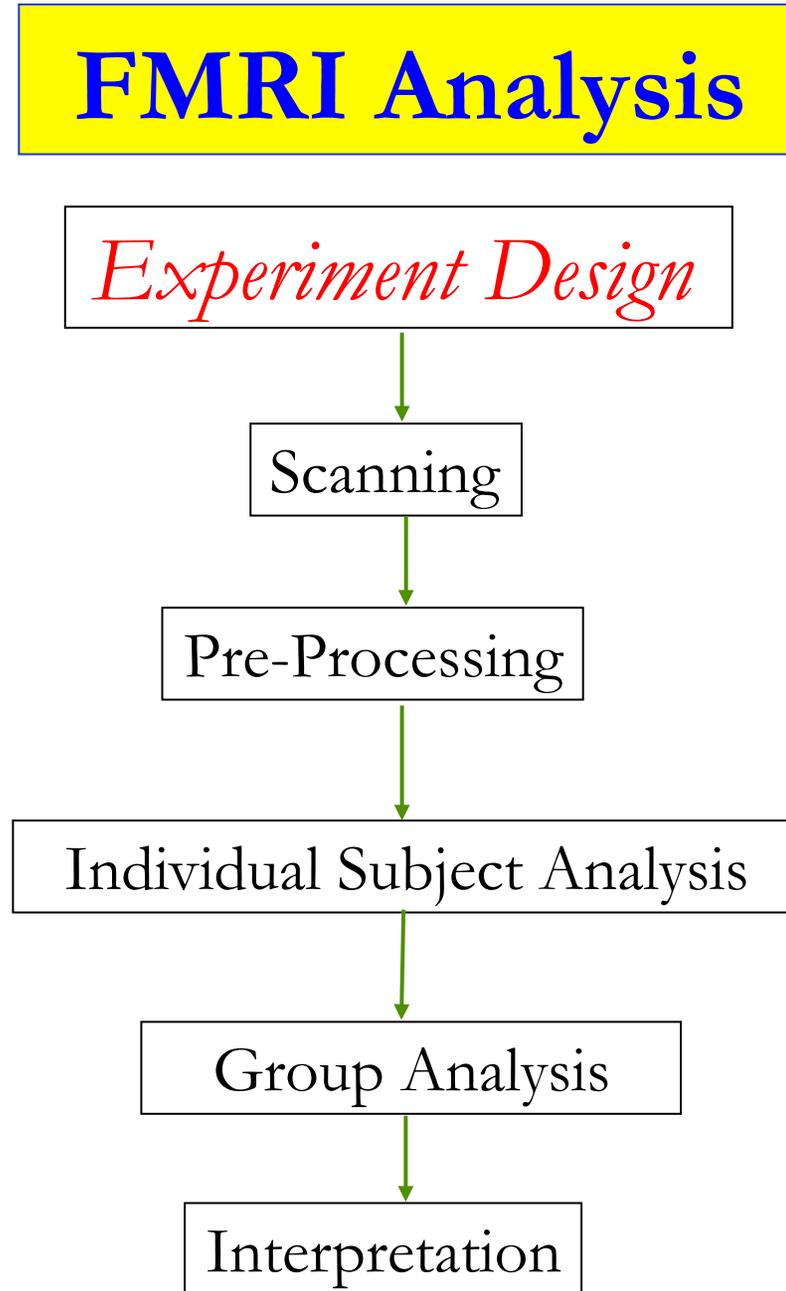
Scanning

Pre-Processing

Individual Subject Analysis

Group Analysis

Interpretation



General suggestions

- ❖ picture this experiment as your own
 - decisions on processing were made by you (and your colleagues)
 - hopefully before acquiring any data
 - there is no single "correct" way to analyze data, just reasonable ways
- ❖ focus on understanding the processing steps
 - in light of your having chosen which steps to perform
- ❖ practice the good habit of reviewing results
 - do the initial images look good?
 - review each processing step along with data
 - are the EPI and anat well aligned by the end?
 - do the statistical results look reasonable?
- ❖ create scripts for any processing step
 - they are a record of how data was processed
 - easy to apply to any new subjects
 - easy to repeat
 - **expect to re-analyze everything (mistake, new decision, etc.)**
 - keep original data and all processing scripts

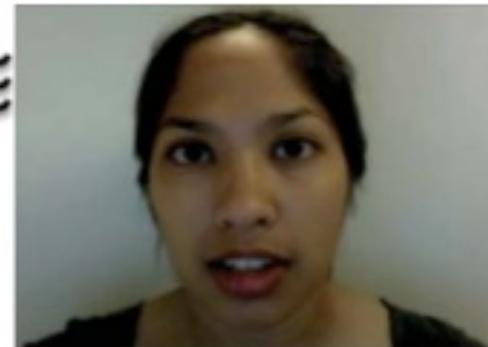
A sample Study

- ◆ Speech Perception Task: Subjects were presented with audiovisual speech that was presented in a predominantly auditory or predominantly visual modality.
- ◆ A digital video system was used to capture auditory and visual speech from a female speaker.
- ◆ There were 2 types of stimulus conditions:



(1) Auditory-Reliable

Example: Subjects can clearly *hear* the word “cat,” but the video of a woman mouthing the word is degraded.

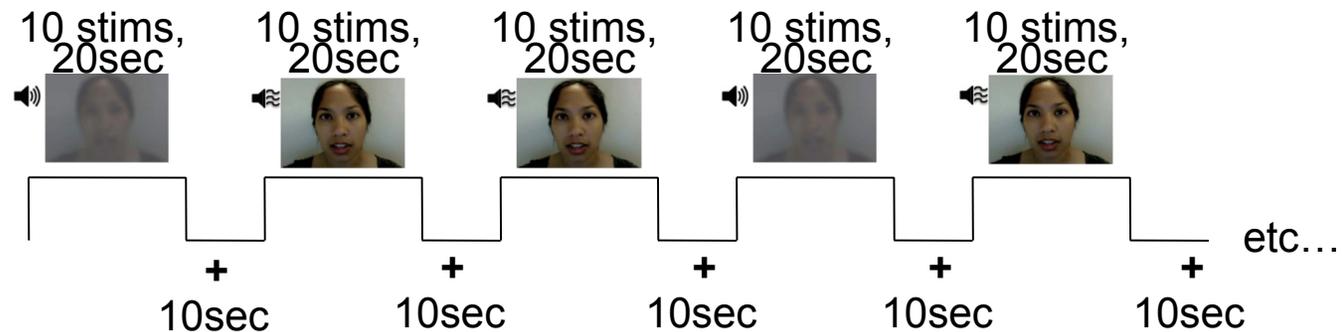


(2) Visual-Reliable

Example: Subjects can clearly *see* the video of a woman mouthing the word “cat,” but the audio of the word is degraded.

❖ Experiment Design:

- ◆ There were 3 runs in a scanning session.
- ◆ Each run consisted of 10 blocked trials:
 - 5 blocks contained Auditory-Reliable (*AreI*) stimuli, and
 - 5 blocks contained Visual-Reliable (*Vrel*) stimuli.
- ◆ Each block contained 10 trials of *AreI* stimuli OR 10 trials of *Vrel* stimuli.
 - Each block lasted for 20 seconds (1 second for stimulus presentation, followed by a 1-second inter-stimulus interval).
- ◆ Each baseline block consisted of a 10-second fixation point.



Players in Experiment Design

- Design of the study
 - Complexity: factors, levels, covariate, contrasts of interest, ...
 - Design choices may limit statistical analysis options
- Number of events per class (sample size for a regressor)
 - The more the better (20+), but no magic number
- Number of condition classes (regressors)
 - Be parsimonious
- HRF modeling
 - Fixed shape, whatever fits the data, or other basis functions?
- Event arrangement
 - How to design? How to define the 'best' design?
 - Efficiency: achieve highest statistical power within fixed scanning time
- Inter-Stimulus Interval (ISI) and Stimulus Onset Asynchrony (SOA)
 - ISI: from the end (offset) of an event to the beginning (onset) of the next
 - $SOA = \text{stimulus duration} + \text{ISI}$

Players in Experiment Design

- Number of subjects (n)
 - Important for group analysis: inter-subject vs. intra-subject variation
 - Power (success to detect signal if present) roughly proportional to \sqrt{n}
 - Design type: block vs. event-related
 - Recommended: 25+ for event-related
- Number of time points
 - Important for individual subject analysis, but also group analysis **when estimate variance is considered**
 - Power proportional to \sqrt{DF}
 - Limited by subject's tolerance in scanner: 30-90 min per session
- TR length
 - Shorter TR yields more time points (and potentially more power), but
 - Power improvement limited by weaker MR signal
 - Usually limited by hardware considerations

Design Types

- Event-related design
 - Modeling options
 - Rigid - Prefixed shape: **GAM**(**p**, **q**) (instantaneous duration), **BLOCK**(**d**,**p**)
 - Reliable and low cost if the HRF is very close to the model
 - Flexible - Whatever fits the data: deconvolution: **TENT**(**b**,**c**,**n**), **CSPLIN**(**b**,**c**,**n**)
 - Sensitive to HRF subtle changes across regions/conditions
 - High statistical cost; over-fitting; difficulty in group analysis
 - Middle ground - Various basis functions: **SPMG1/2/3**, **SIN**, **POLY**
- Block design
 - Conditions with lasting durations of more than one TR
 - Other terminologies: epoch, box-car
 - Usually modeled with prefixed-shape HRF (**BLOCK**), but
 - basis function (**TENT**) approach for flexible shapes
 - multiple events for each block: can model amplitude attenuation
- Mixed design

Statistical Theory Of Level 1 Tests

- Regression Model (GLM)

- $Y = X\beta + \varepsilon$, X : design matrix with regressors as columns

- General Linear testing

- Hypothesis $H_0: c' \beta = 0$ with $c = \text{vector } (c_0, c_1, \dots, c_p)$ or matrix

- $t = c' \beta / \sqrt{[c' (X' X)^{-1} c] MSE}$ (MSE : unknown but same across tests)

- Signal-to-noise ratio

- Effect vs. uncertainty

- $\sqrt{c' (X' X)^{-1} c}$: **normalized standard deviation** of contrast $c' b$

- Scaling factor for uncertainty/unreliability/imprecision, and totally under our control

- Efficiency = $1 / \sqrt{c' (X' X)^{-1} c}$: Smaller norm. std. dev. → more efficient

- $X' X$ measures co-variation among regressors: Less correlated regressors → more efficient and easier to tease apart regressors

- Goal: find a design (X) that renders low norm. std. dev. or less correlated regressors

- Assuming no temporal correlations in the residuals: real power might be slightly lower

Find an efficient design

- Efficient design search used event-related type
- Block or mixed type is typically designed manually
- Most parameters (TR, number of subjects/conditions/runs/sessions/time points, ...) are preset usually through other considerations before design search
- **There are many good designs**
 - Infinite possibilities
 - Used to avoid undesirable designs (collinearity problem) more than optimal one(s)
 - A manual design might be approximately (if not equally) optimal

Multiple Stimuli - Experiment Design

- How many distinct stimuli do you need in each class? Our rough recommendations:
 - Short event-related designs: at least 25 events in each stimulus class (spread across multiple imaging runs) — and more is better
 - Block designs: at least 5 blocks in each stimulus class — 10 would be better

- While we're on the subject: **How many subjects?**
 - Several independent studies agree that 20-25 subjects in each category are needed for highly reliable results
 - This number is more than has usually been the custom in fMRI-based studies!!

Data Analysis Philosophy

- **Signal** = Measurable response to stimulus
- **Noise** = Components of measurement that interfere with detection of signal
- Statistical detection theory:
 - **Understand** relationship between stimulus & signal
 - Characterize noise statistically
 - Can then devise methods to distinguish noise-only measurements from signal+noise measurements, and assess the methods' reliability
 - Methods and usefulness depend strongly on the assumptions
 - Some methods are more “robust” against erroneous assumptions than others, but may be less sensitive

Time Series Analysis on Voxel Data

- Most common forms of fMRI analysis involve fitting an activation+BOLD model to each voxel's time series *separately* (“massively univariate” analysis)
 - Some pre-processing steps do include inter-voxel computations; e.g.,
 - spatial smoothing to reduce noise
 - spatial registration to correct for subject motion
- Result of model fits is a set of parameters at each voxel, estimated from that voxel's data
 - e.g., activation amplitude (β), delay, shape
 - “**SPM**” = statistical parametric map; e.g., β or t or F
- Further analysis steps operate on individual SPMs
 - ★ e.g., combining/contrasting data among subjects
 - sometimes called “second level” or “meta” analysis

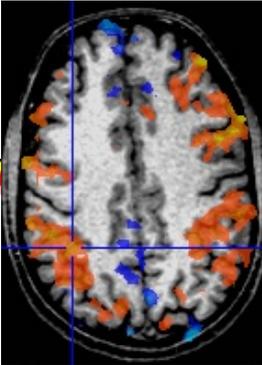
Some Features of fMRI Voxel Time Series

- fMRI only measures changes due to neural “activity”
 - Baseline level of signal in a voxel means little or nothing about neural activity
 - Also, baseline level tends to drift around slowly (100 s time scale or so; mostly from small subject motions)
- Therefore, an fMRI experiment must have at least 2 different neural conditions (“tasks” and/or “stimuli”)
 - Then statistically test for differences in the MRI signal level between conditions
 - Many experiments: one condition is “rest/control”
- Baseline is modeled separately from activation signals, and baseline model includes “rest” periods
 - In AFNI, that is; in SPM, “rest” is modeled explicitly

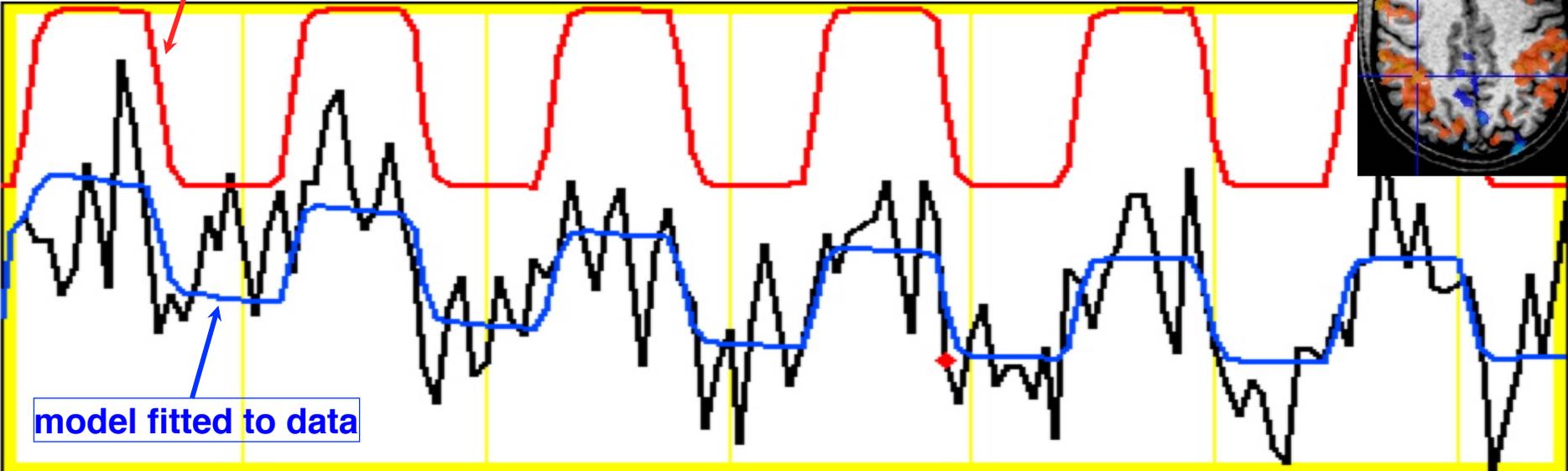
Some Sample fMRI Data Time Series

- First sample: Block-trial fMRI data
 - “Activation” occurs over a sustained period of time (say, 10 s or longer), usually from more than one stimulation event, in rapid succession
 - BOLD (hemodynamic) response accumulates from multiple close-in-time neural activations and is large
 - BOLD response is often visible in time series
 - Noise magnitude about same as BOLD response
- Next 2 slides: same brain voxel in 3 (of 9) EPI runs
 - **black curve** (noisy) = data
 - **red curve** (above data) = ideal model response
 - **blue curve** (within data) = model fitted to data
 - somatosensory task (finger being rubbed)

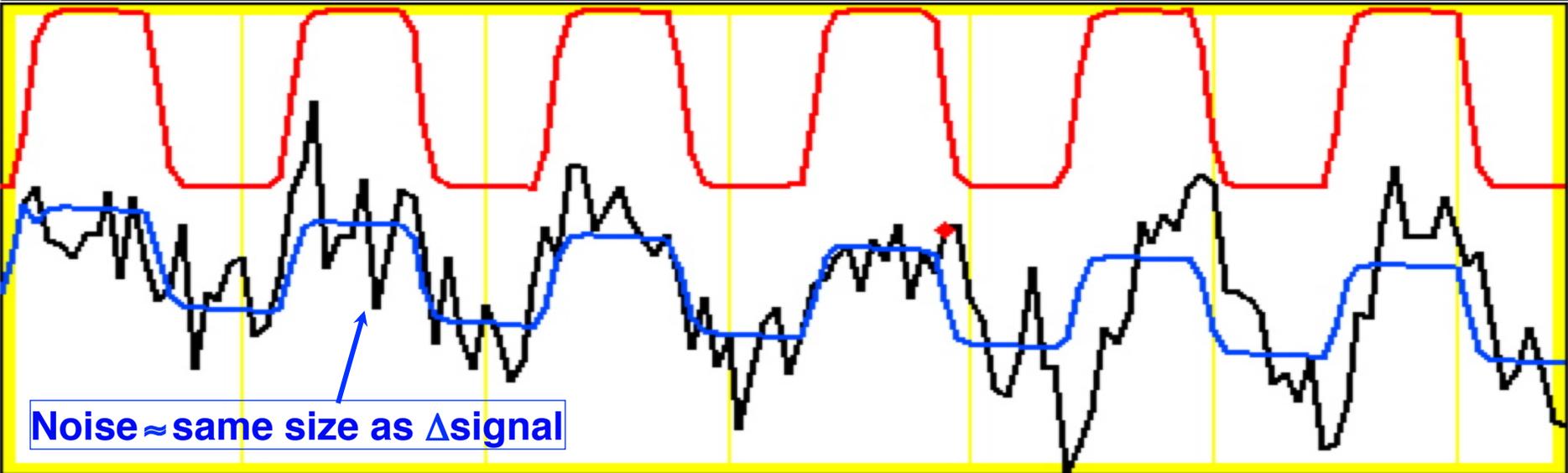
Same Voxel: Runs 1 and 2



model regressor

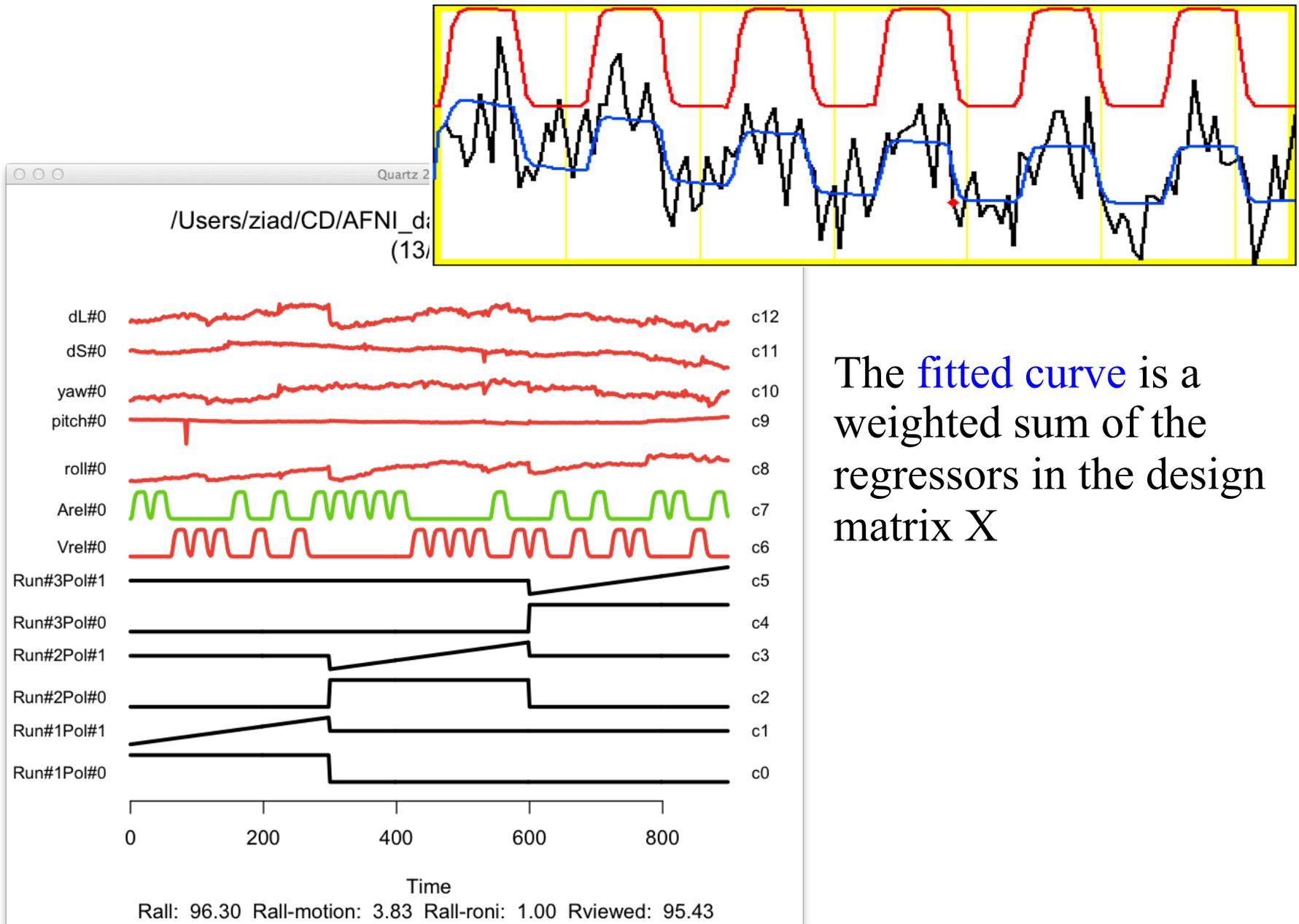


model fitted to data



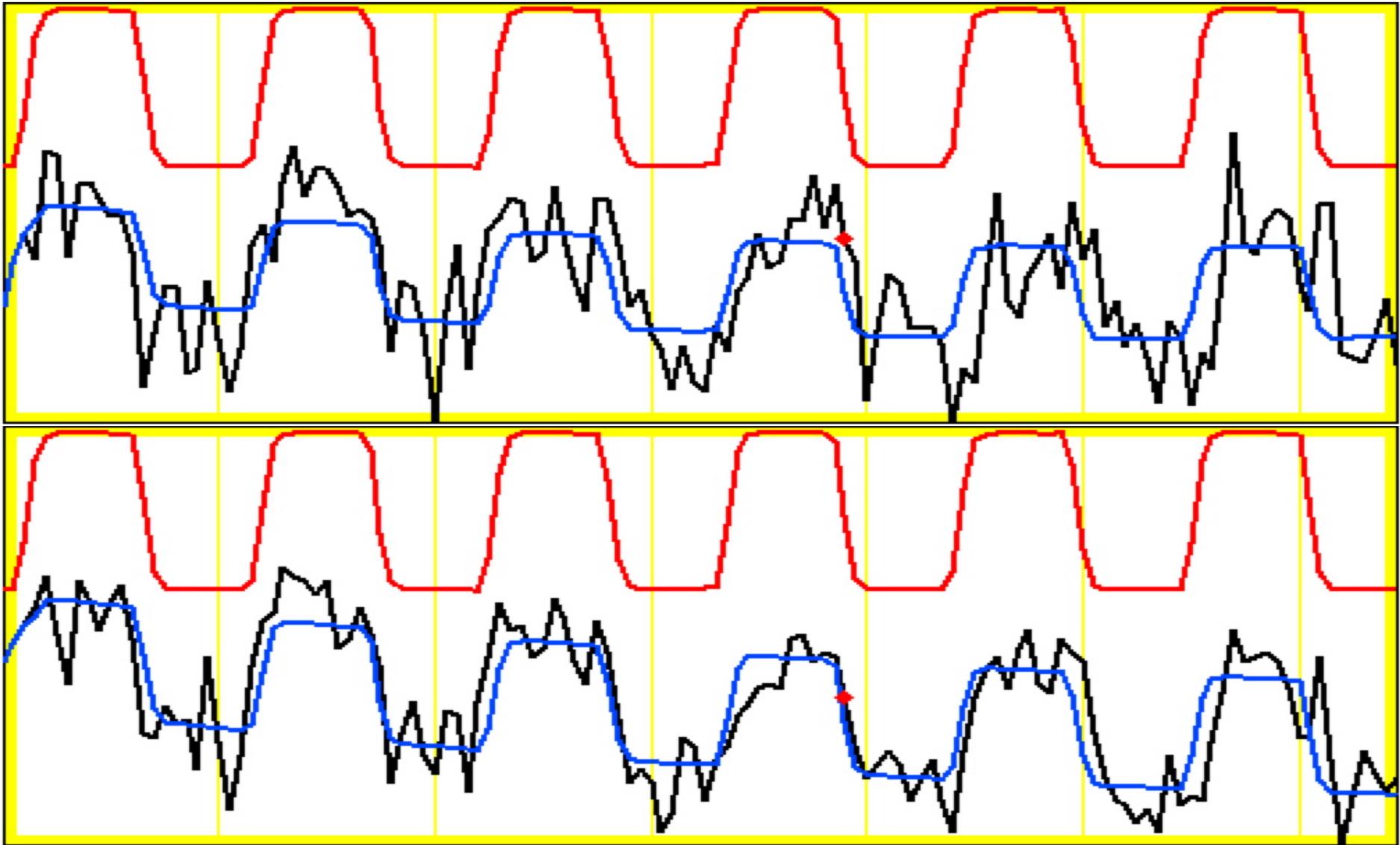
Noise ~ same size as Δ signal

Block-trials: 27 s “on” / 27 s “off”; TR=2.5 s; 130 time points/run



The **fitted curve** is a weighted sum of the regressors in the design matrix X

Same Voxel: Run 3 and Average of all 9



Activation amplitude & shape vary among blocks! Why???

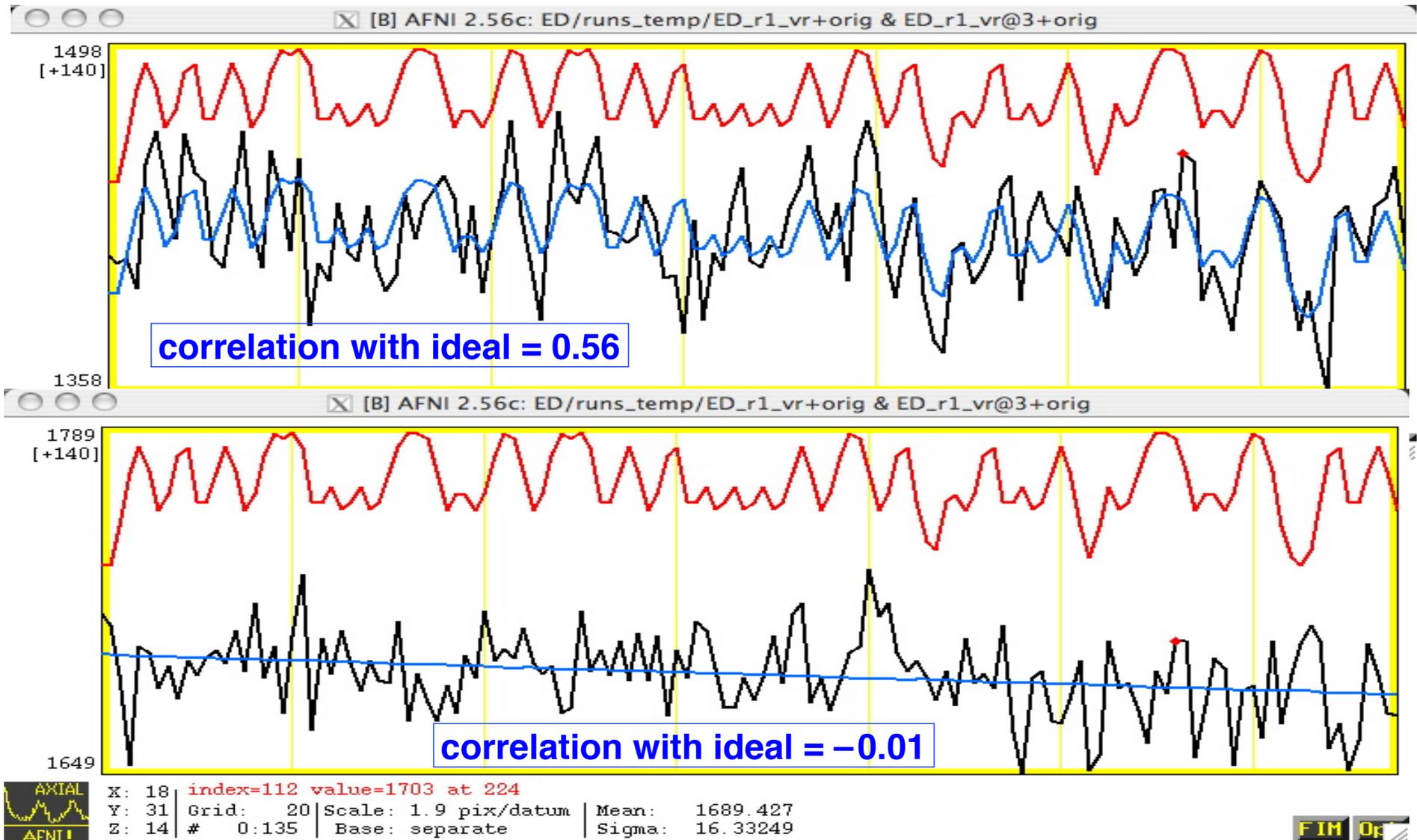
More Sample FMRI Data Time Series

- Second sample: Event-Related FMRI
 - “Activation” occurs in single relatively brief intervals
 - “Events” can be randomly or regularly spaced in time
 - If events are randomly spaced in time, signal model itself looks noise-like (to the pitiful human eye)
 - BOLD response to stimulus tends to be weaker, since fewer nearby-in-time “activations” have overlapping signal changes (hemodynamic responses)
- Next slide: Visual stimulation experiment

“Active” voxel shown in next slide



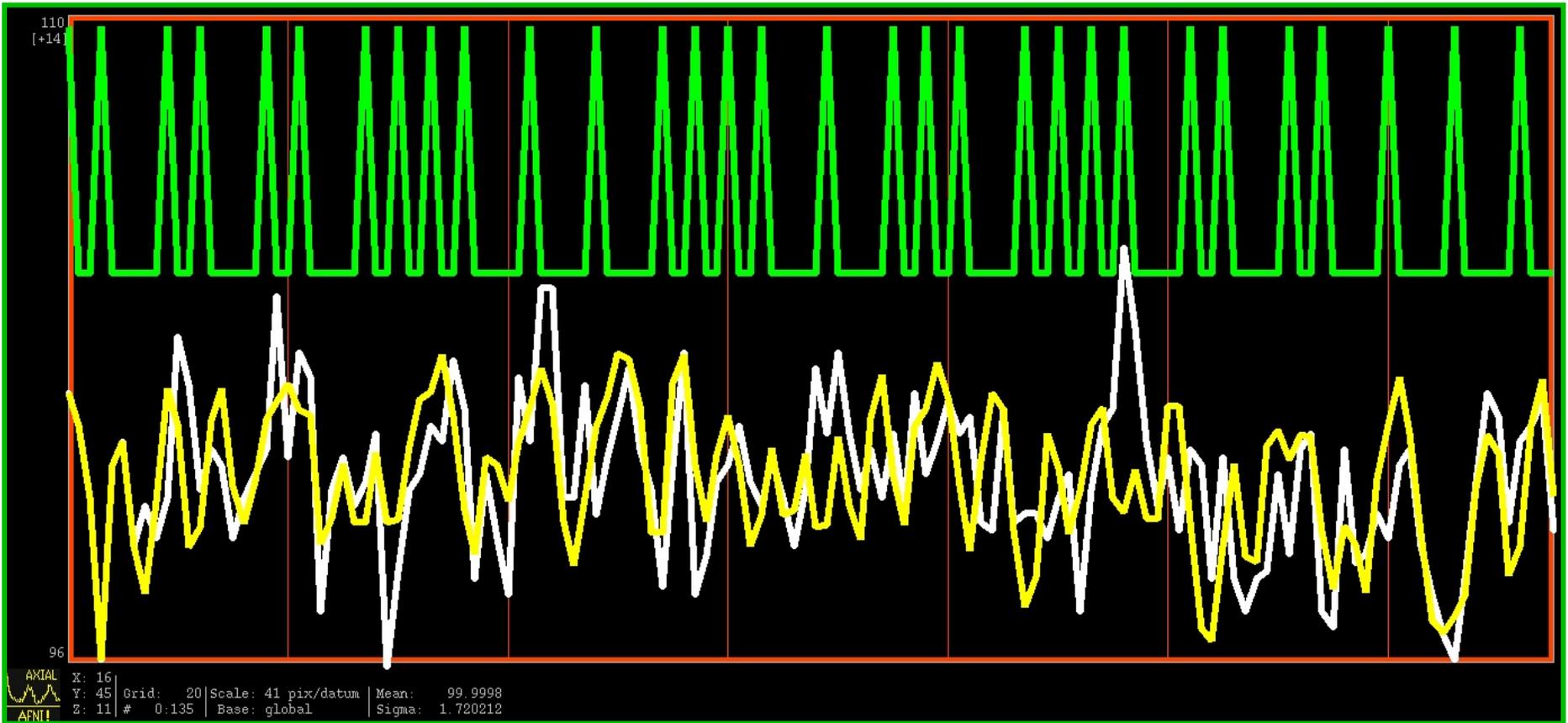
Two Voxel Time Series from Same Run



Lesson: ER-FMRI activation is not obvious via casual inspection

More Event-Related Data

Four different visual stimuli

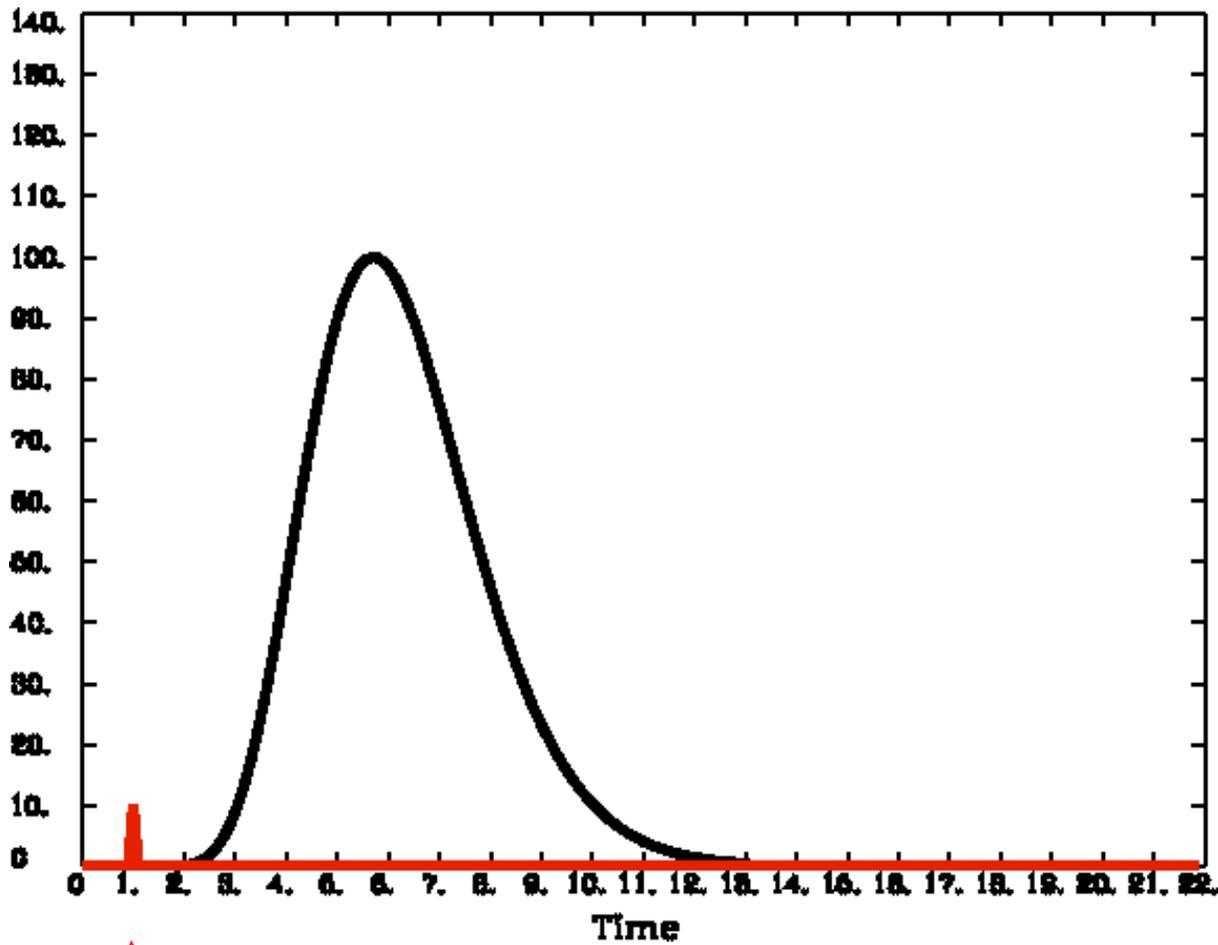


- White curve = Data (first 136 TRs)
- Yellow curve = Model fit ($R^2=50\%$)
- Green = Stimulus timing

Very good fit for ER data ($R^2=10-20\%$ more usual).
Noise is as big as BOLD!

Hemodynamic Response Function (HRF)

- **HRF** is the idealization of measurable fMRI signal change responding to a single activation cycle (up and down) from a stimulus in a voxel



Response to brief activation (< 1 s):

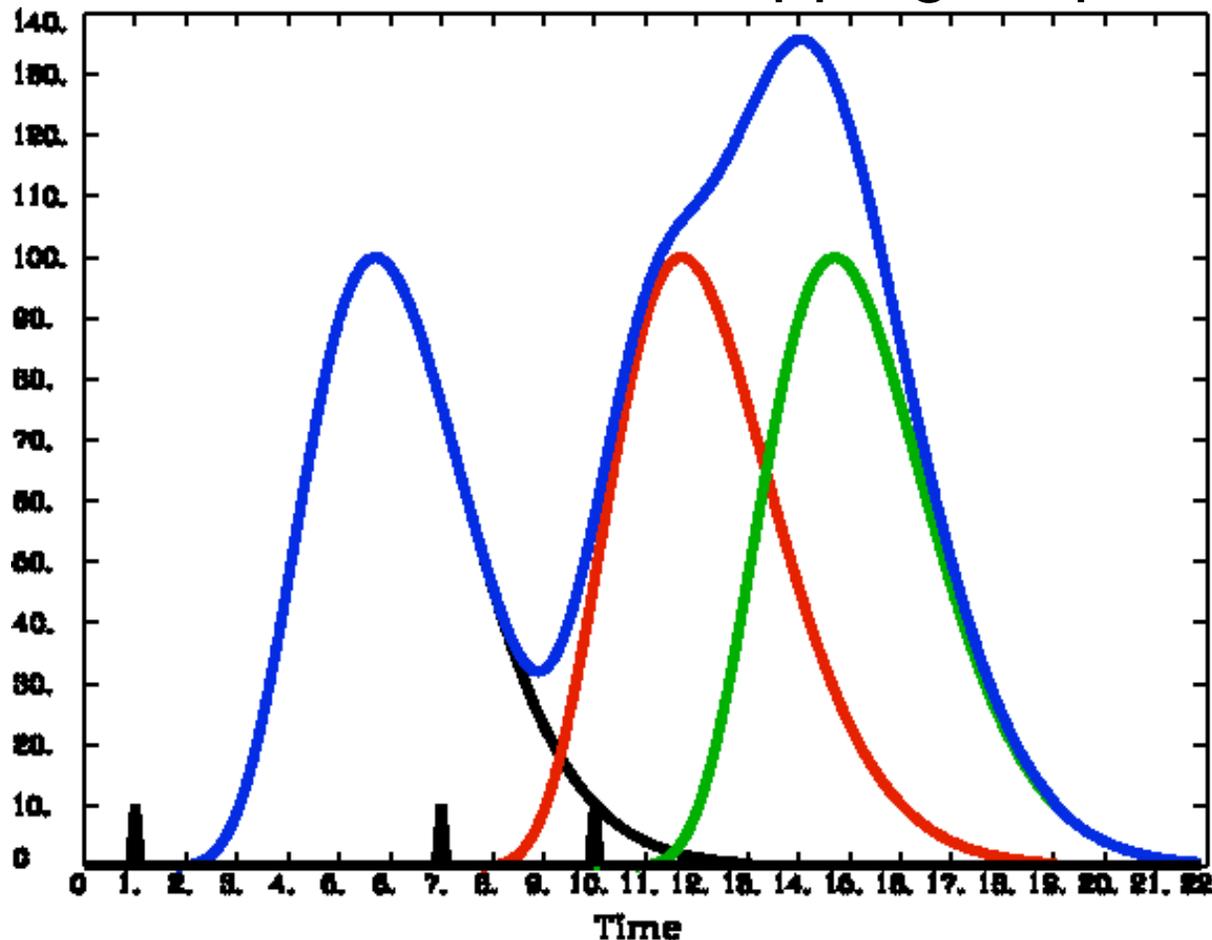
- delay of 1-2 s
- rise time of 4-5 s
- fall time of 4-6 s
- model equation:
$$h(t) \propto t^b e^{-t/c}$$
- $h(t)$ is signal change t seconds **after** activation

1 Brief Activation (Event)



Linearity (Additivity) of HRF

- Multiple activation cycles in a voxel, closer in time than duration of HRF:
 - Assume that overlapping responses add

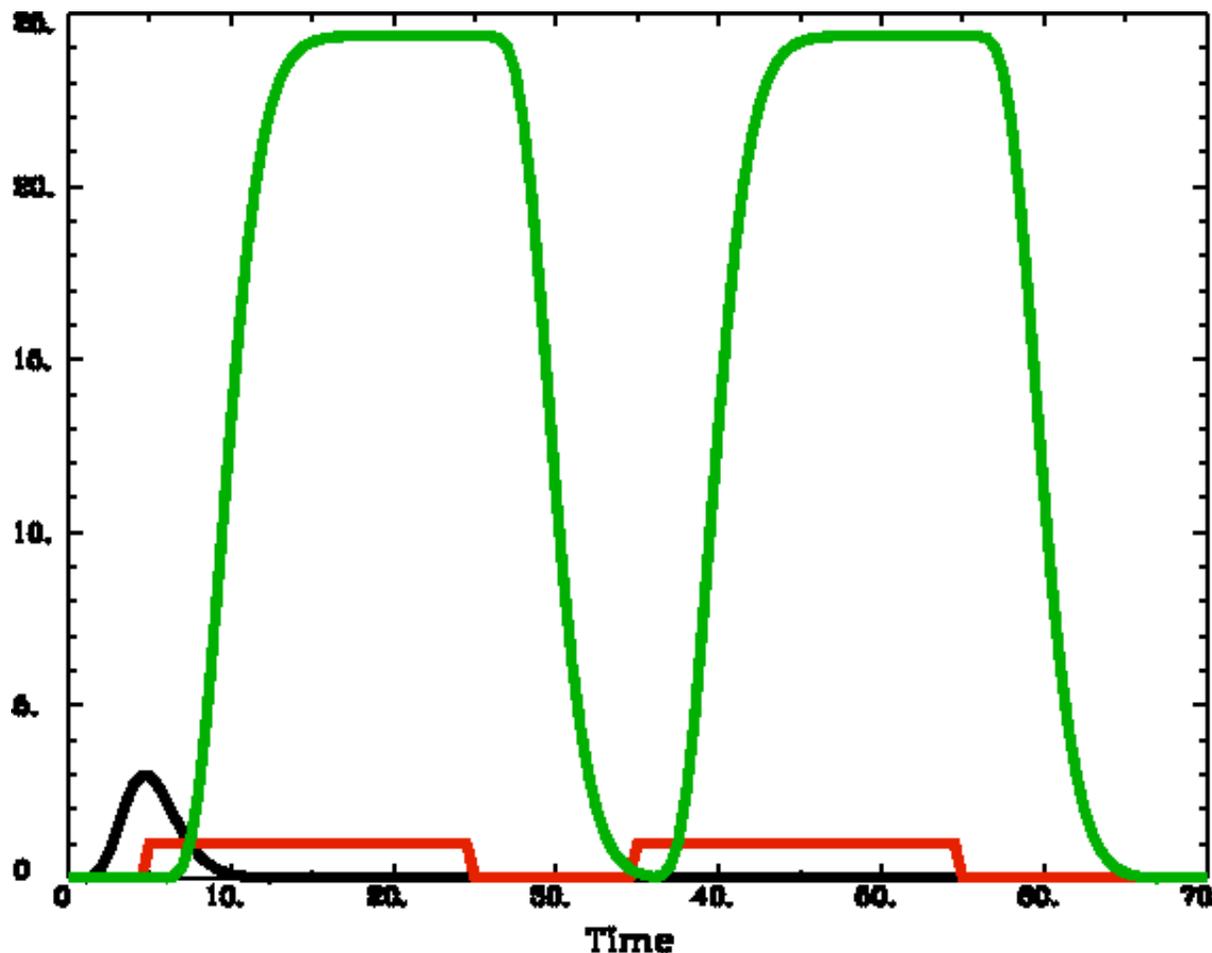


- Linearity is a pretty good assumption
- But not apparently perfect — about 90% correct
- Nevertheless, is widely taken to be true and is the basis for the “general linear model” (GLM) in FMRI analysis

3 Brief Activations

Linearity and Extended Activation

- Extended activation, as in a block-trial experiment:
 - HRF accumulates over its duration ($\approx 10-12$ s)

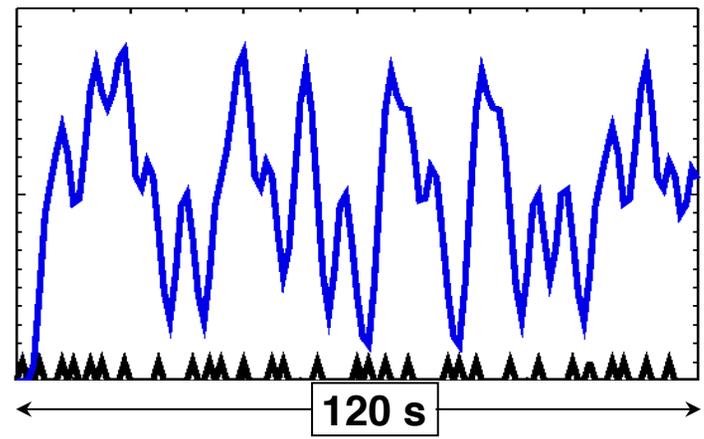
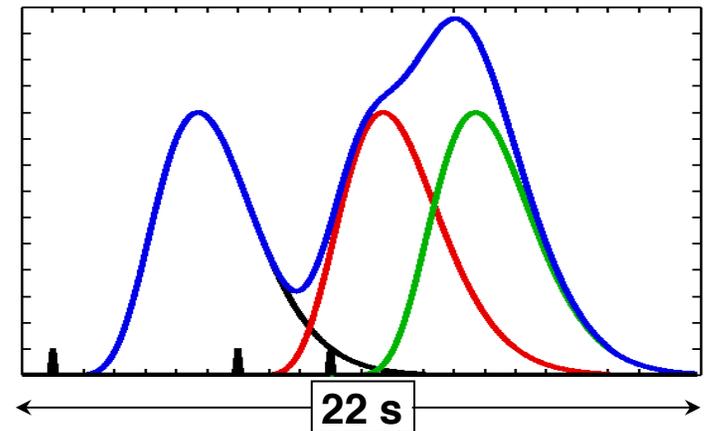


- **Black** curve = response to a single brief stimulus
- **Red** curve = activation intervals
- **Green** curve = summed up HRFs from activations
- Block-trials have larger BOLD signal changes than event-related experiments

2 Long Activations (Blocks)

Convolution Signal Model

- FMRI signal model (in each voxel) is taken as sum of the individual trial HRFs (assumed equal)
 - Stimulus timing is assumed known (or measured)
 - Resulting time series (in **blue**) are called the **convolution** of the HRF with stimulus timing
 - Finding HRF = “deconvolution”
 - AFNI code = 3dDeconvolve (or its daughter 3dREMLfit)
 - Convolution models only the FMRI signal **changes** →



• Real data starts at and returns to a nonzero, slowly drifting baseline

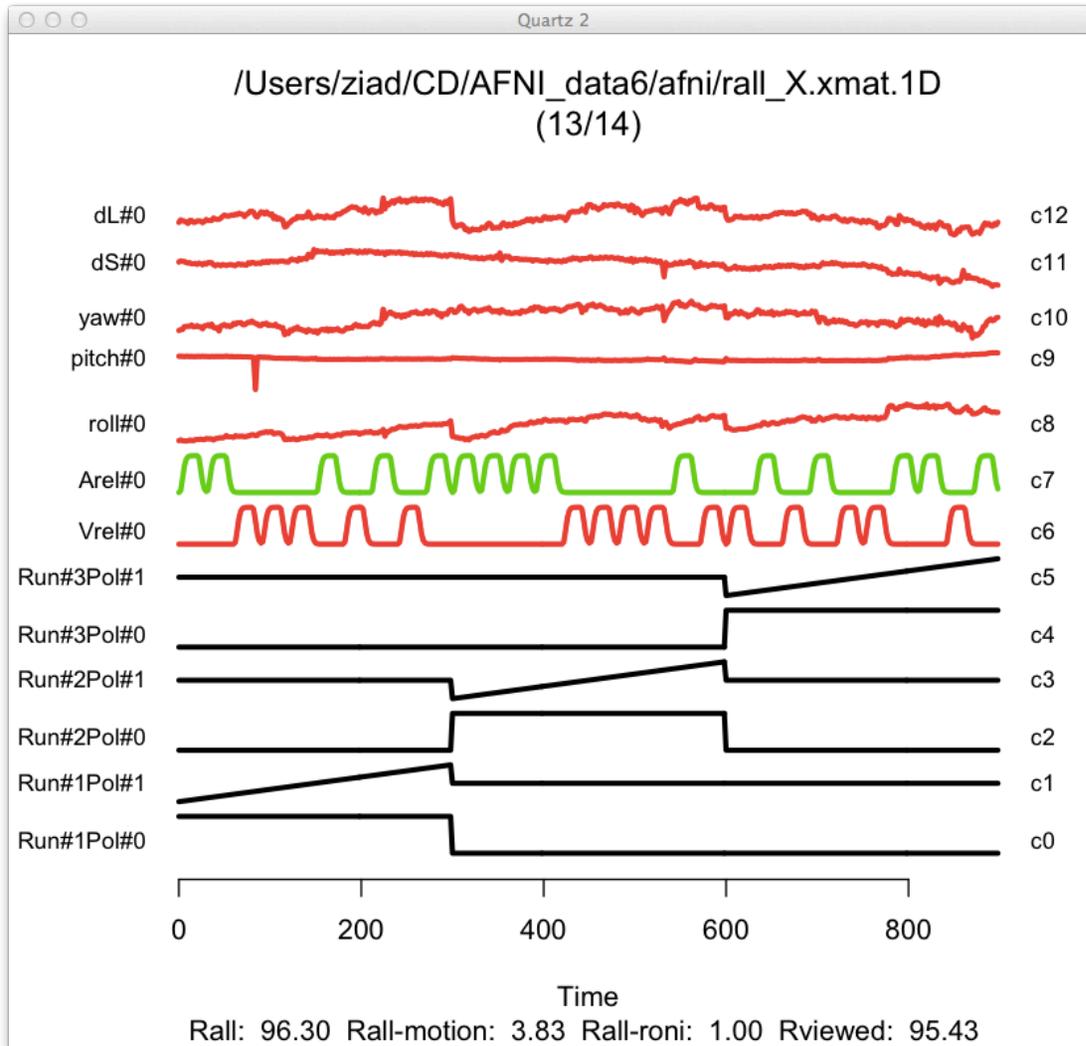
Simple Regression Models

- Assume a fixed shape $h(t)$ for the HRF
 - e.g., $h(t) = t^{8.6} \exp(-t/0.547)$ [MS Cohen, 1997]
 - Convolve with stimulus timing to get ideal response (temporal pattern) $r(t) = \sum_{k=1}^K h(t - \tau_k) =$ sum of HRF copies
- Assume a form for the baseline (data without activation)
 - e.g., $a + b \cdot t$ for a constant plus a linear trend
- In each voxel, fit data $Z(t)$ to a curve of the form
$$Z(t) \approx a + b \cdot t + \beta \cdot r(t)$$

← The signal model!

 - a, b, β are unknown values, in each voxel
 - a, b are “nuisance” parameters
 - β is amplitude of $r(t)$ in data = “how much” BOLD
 - In this model, each stimulus assumed to get same BOLD response — in shape and in amplitude

Signal models: $r(t)$



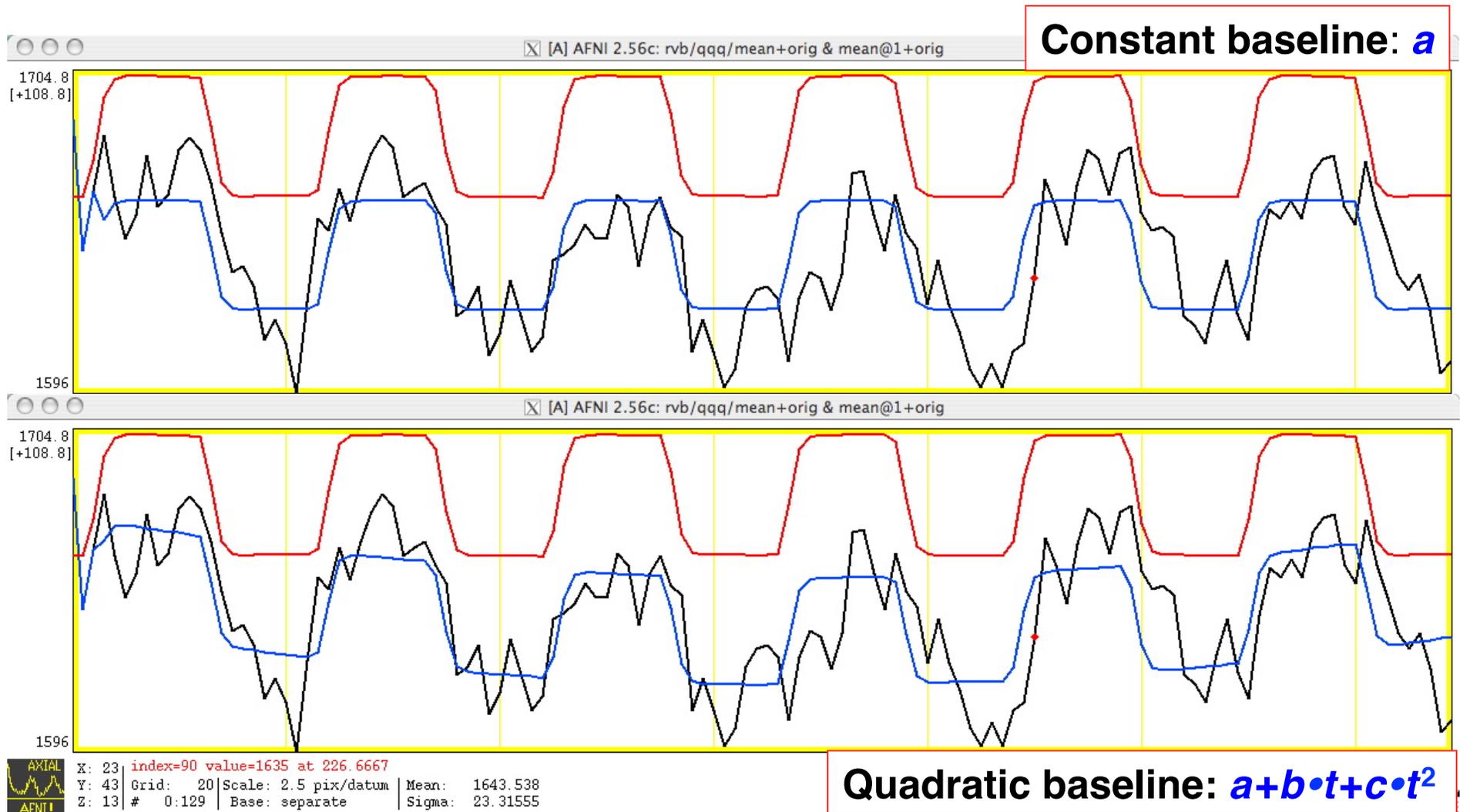
Motion

Task

Baseline

**Motion and Baseline are
*Nuisance Regressors***

Simple Regression: Sample Fits



- Necessary baseline model complexity depends on duration of **continuous** imaging — e.g., 1 parameter per ≈ 150 seconds

Duration of Stimuli - Important Caveats

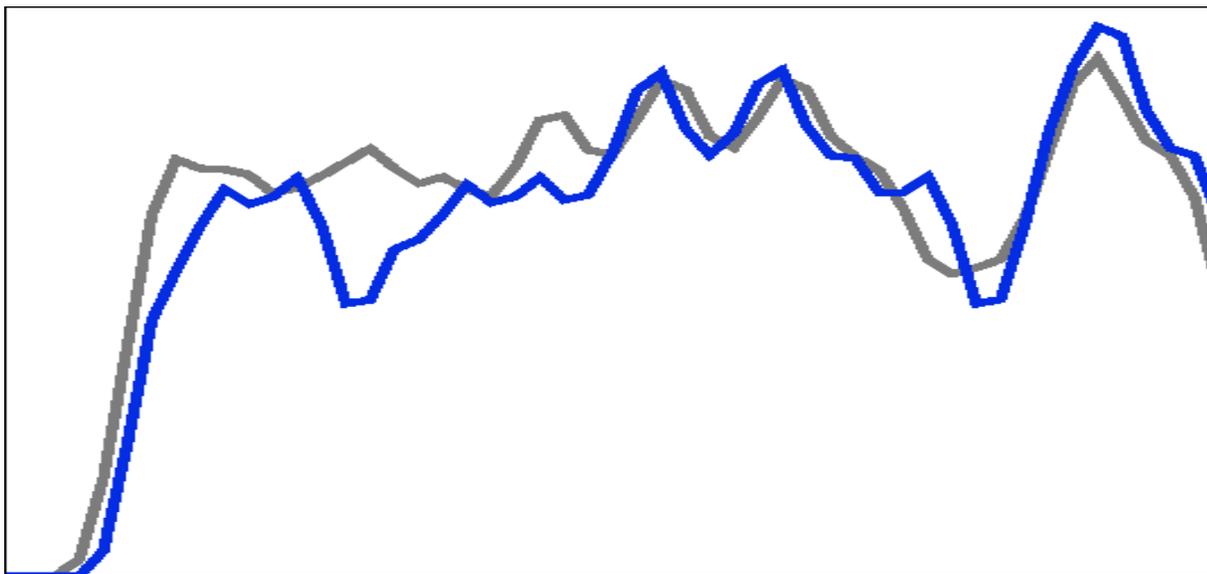
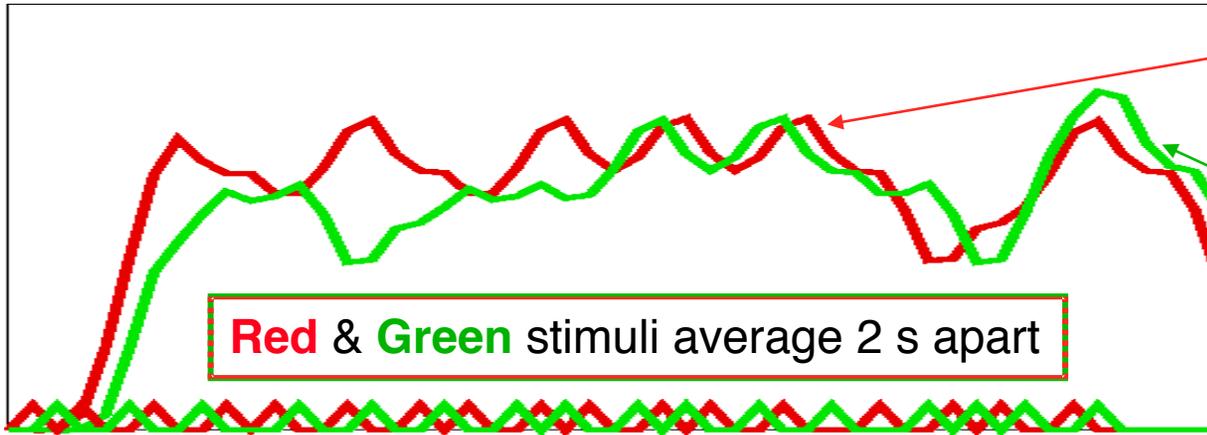
- Slow baseline drift (time scale 100 s and longer) makes doing fMRI with long duration stimuli difficult
 - Learning experiment: where the task is done continuously for ≈ 15 minutes and the subject is scanned to find parts of the brain that adapt during this time interval
 - Pharmaceutical challenge: where the subject is given some psychoactive drug whose action plays out over 10+ minutes (e.g., cocaine, ethanol)

- Multiple very short duration stimuli that are also very close in time to each other are very hard to tell apart, since their HRFs will have 90-95% overlap
 - Binocular rivalry, where percept switches ≈ 0.5 s

Multiple Stimuli = Multiple Regressors

- Usually have more than one class of stimulus or activation in an experiment
 - e.g., want to see size of “**face activation**” vis-à-vis “**house activation**”; or, “**what**” vs. “**where**” activity
- Need to model each separate class of stimulus with a separate response function $r_1(t)$, $r_2(t)$, $r_3(t)$,
 - Each $r_j(t)$ is based on the stimulus timing for activity in class number j
 - Calculate a β_j amplitude = amount of $r_j(t)$ in voxel data time series $Z(t)$ = average BOLD for stim class # j
 - **Contrast** β s to see which voxels have differential activation levels under different stimulus conditions
 - e.g., statistical test on the question $\beta_1 - \beta_2 = 0$?

Multiple Regressors: Near Collinearity



- **Red** curve = signal model for class #1
- **Green** curve = signal model for #2

• **Blue** curve = $\beta_1 \cdot \#1 + (1 - \beta_1) \cdot \#2$
Where β_1 varies randomly from 0.0 to 1.0 in animation

• **Gray** curve = $0.66 \cdot \#1 + 0.33 \cdot \#2$
= simulated data *with no noise*

- Lots of different combinations of **#1** and **#2** are decent fits to gray curve

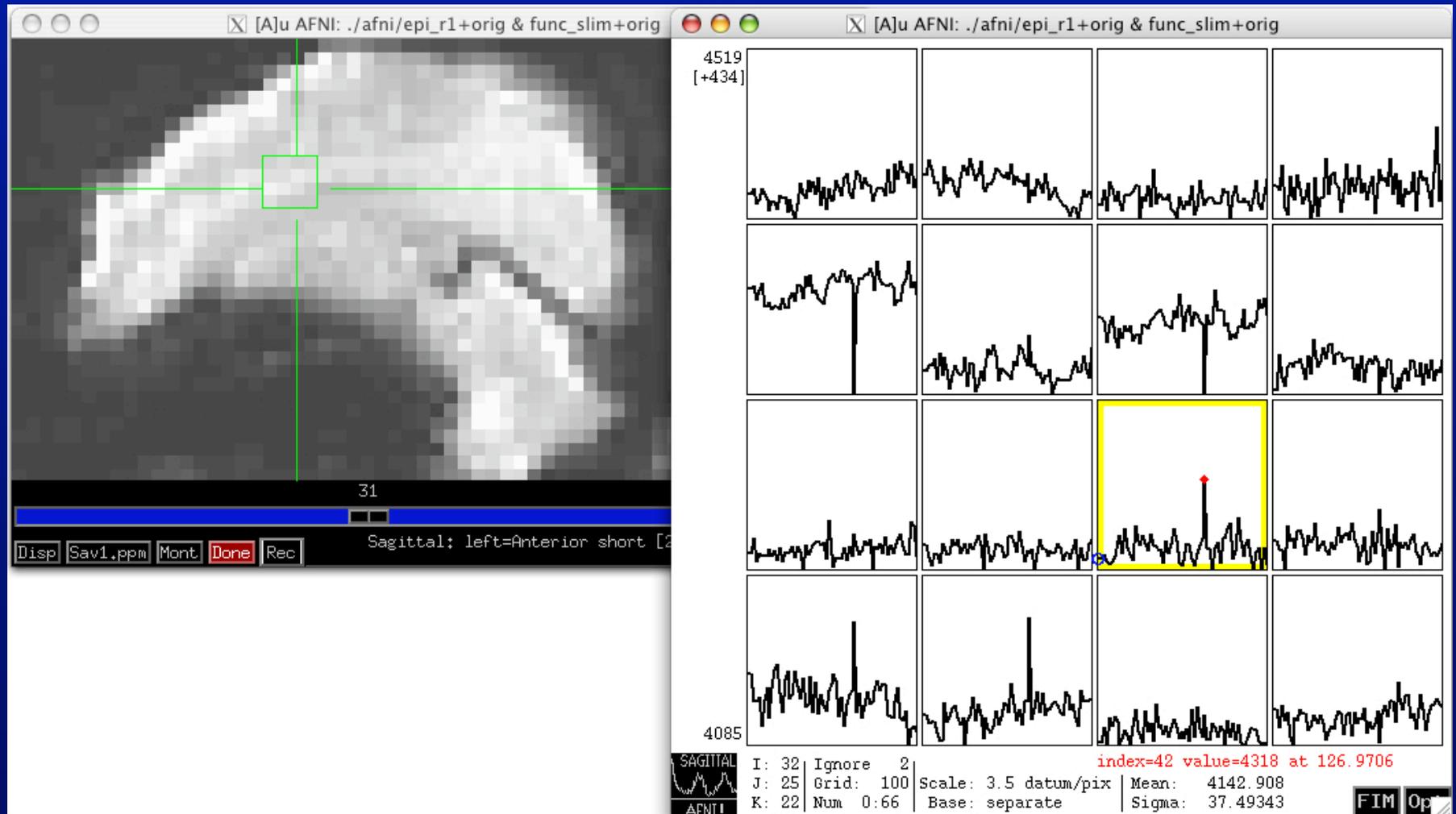
Stimuli are too close in time to distinguish response **#1** from **#2**, considering noise

Simple Regression: Recapitulation

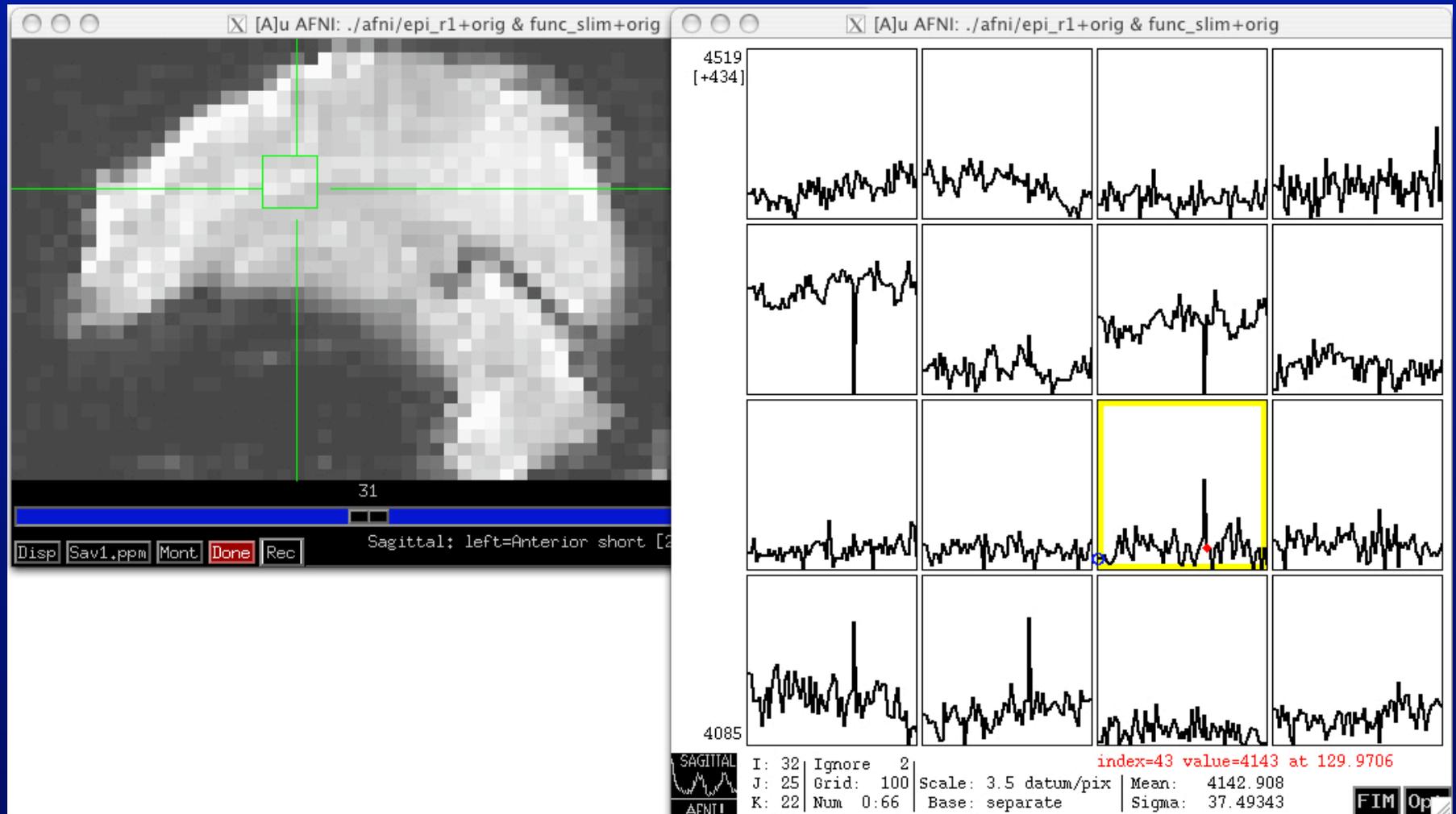
- Choose HRF model $h(t)$ [AKA *fixed-model regression*]
- Build model responses $r_n(t)$ to each stimulus class
 - Using $h(t)$ and the stimulus timing
- Choose baseline model time series
 - Constant + linear + quadratic (+ movement?)
- Assemble model and baseline time series into the columns of the \mathbf{R} matrix
- For each voxel time series \mathbf{z} , solve $\mathbf{z} \approx \mathbf{R}\boldsymbol{\beta}$ for $\hat{\boldsymbol{\beta}}$
- **Individual subject maps:** Test the coefficients in $\hat{\boldsymbol{\beta}}$ that you care about for statistical significance
- **Group maps:** Transform the coefficients in $\hat{\boldsymbol{\beta}}$ that you care about to Talairach/MNI space, and perform statistics on the collection of $\hat{\boldsymbol{\beta}}$ values across subjects

Motion, The Second Nuisance in FMRI

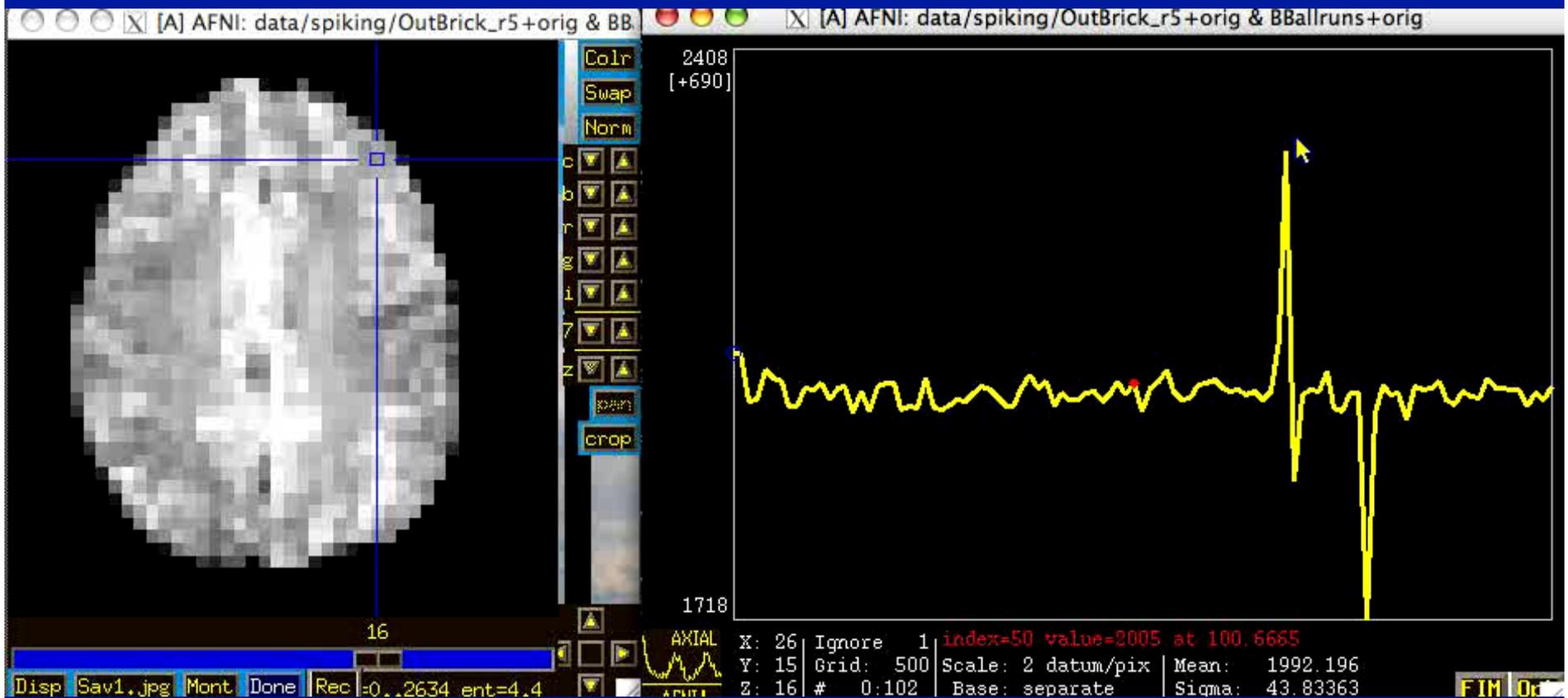
Movement Spikes



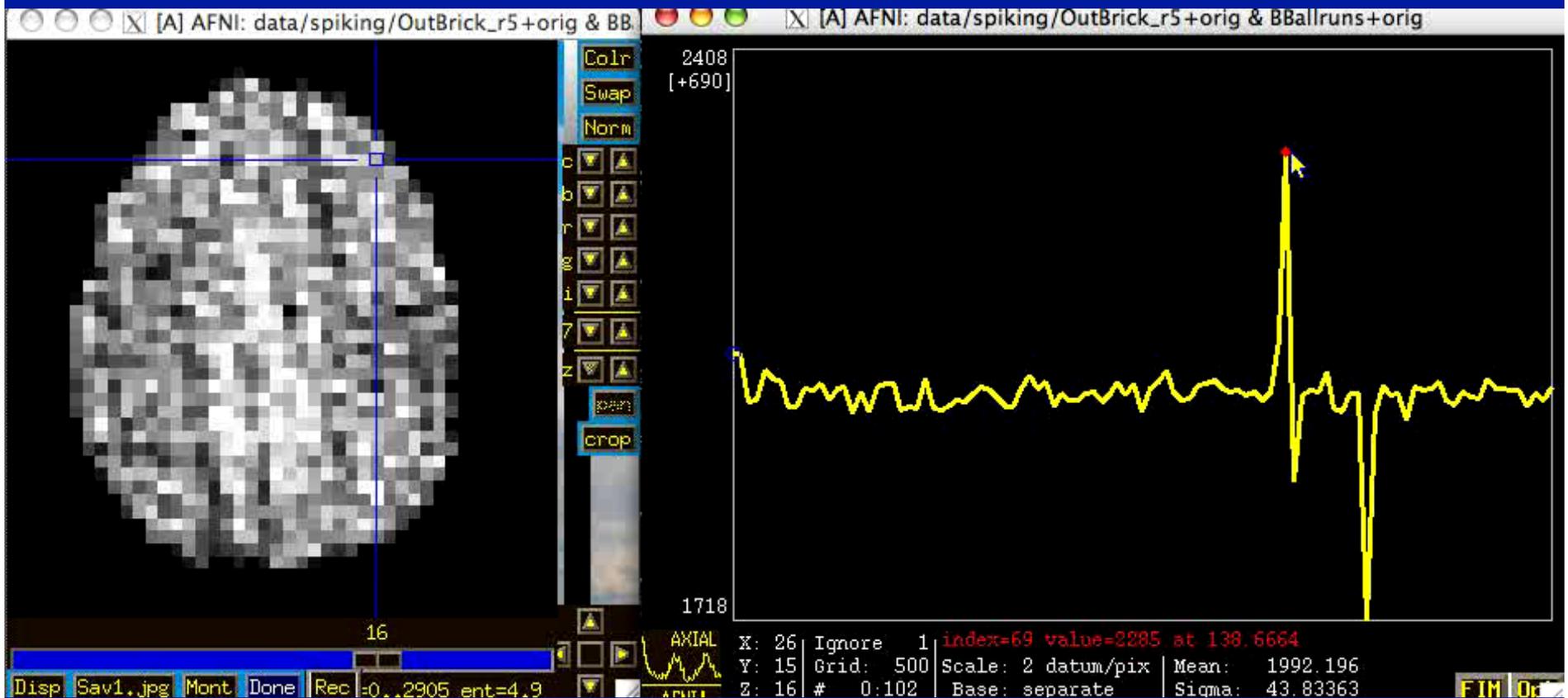
Movement Spikes



Hardware Spike

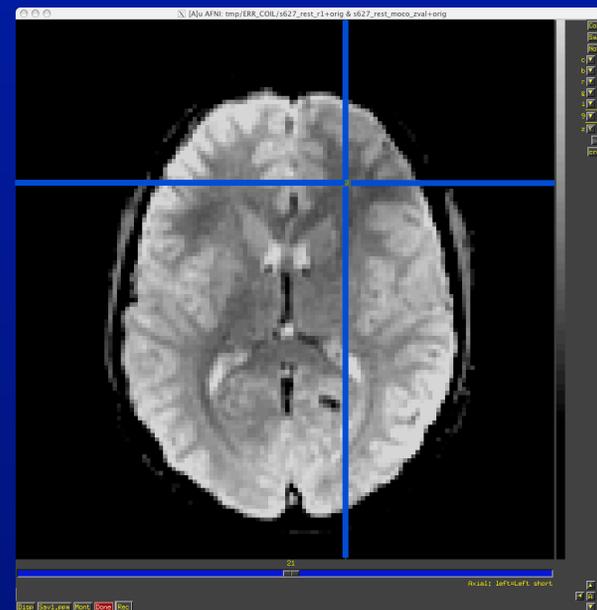
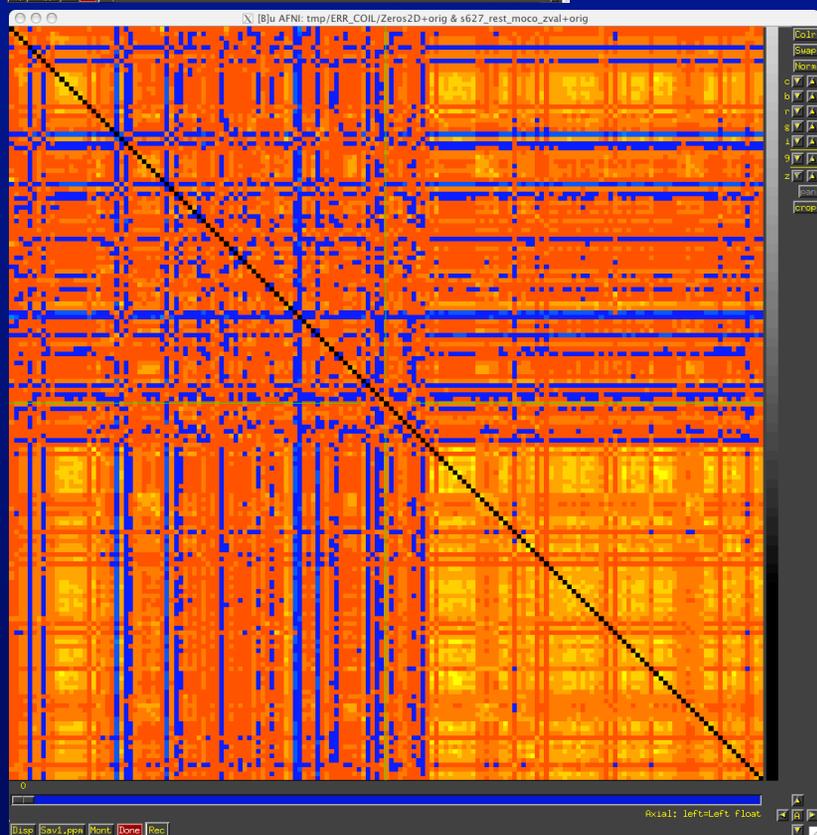
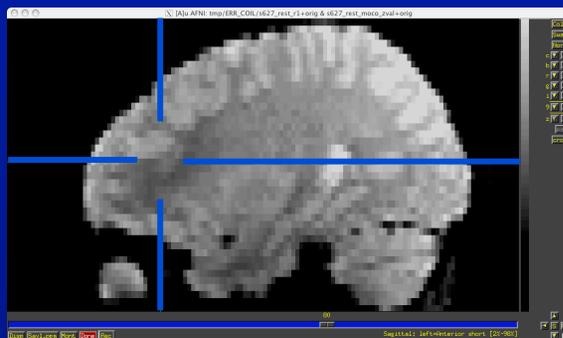


Hardware Spike

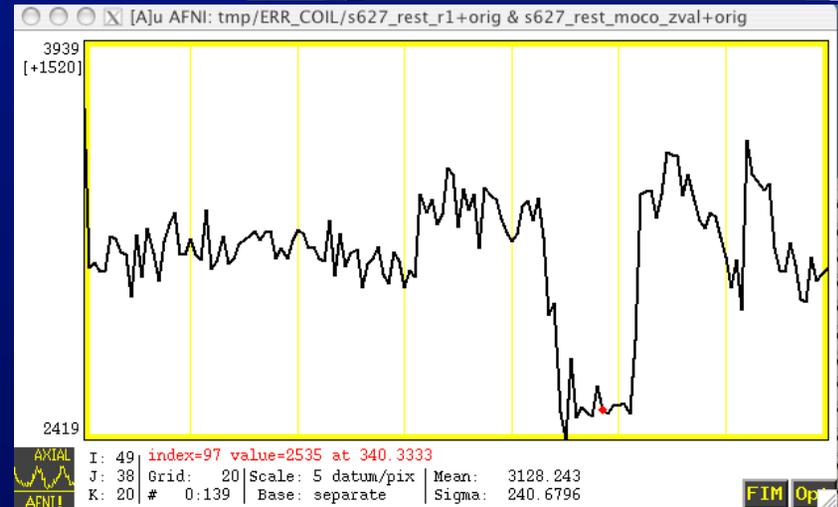
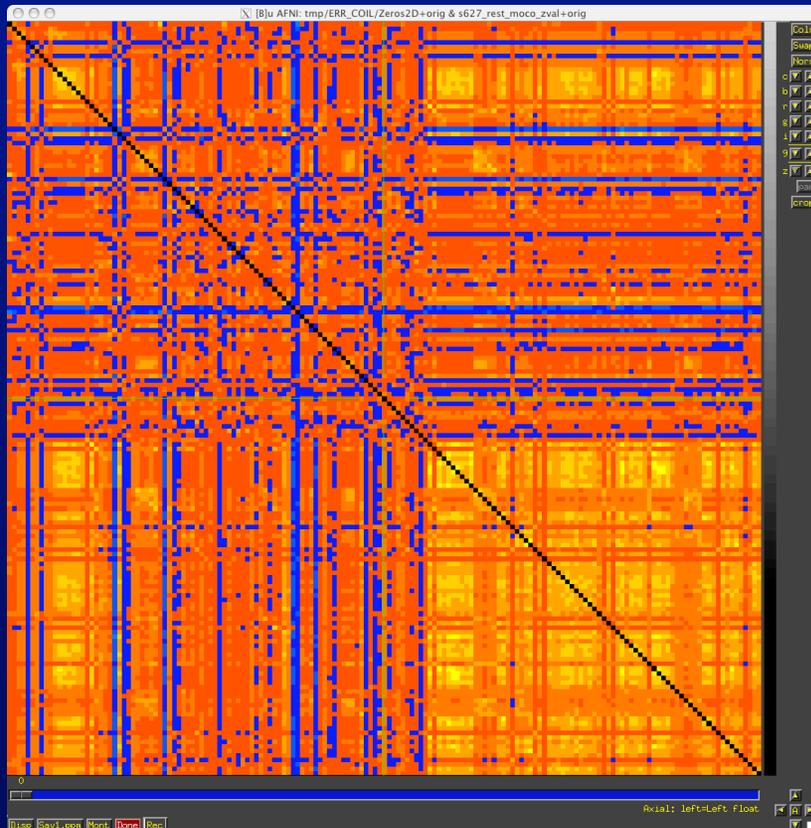
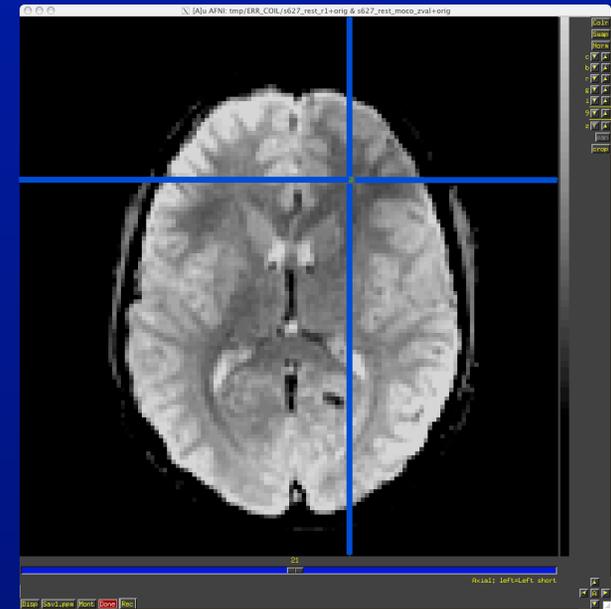
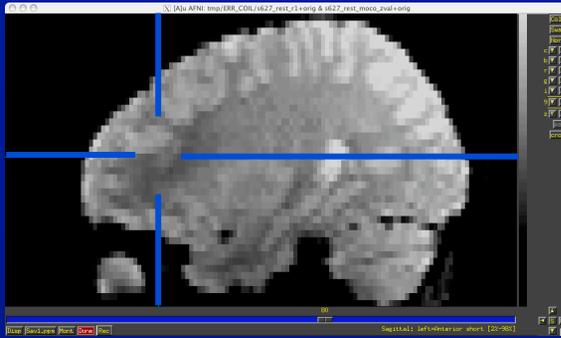


- Spikes caused by loose gradient coil connection

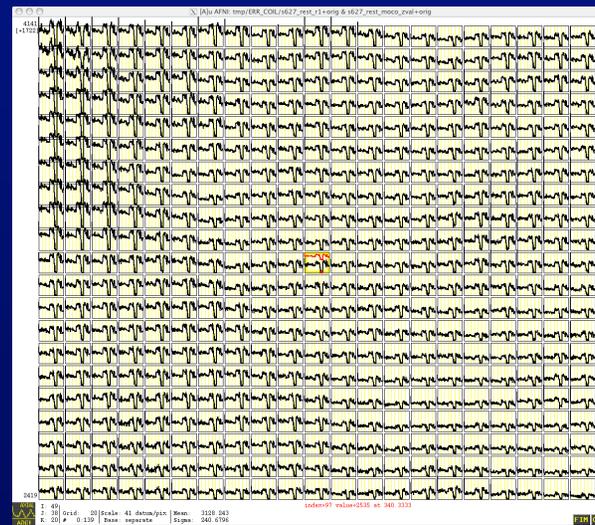
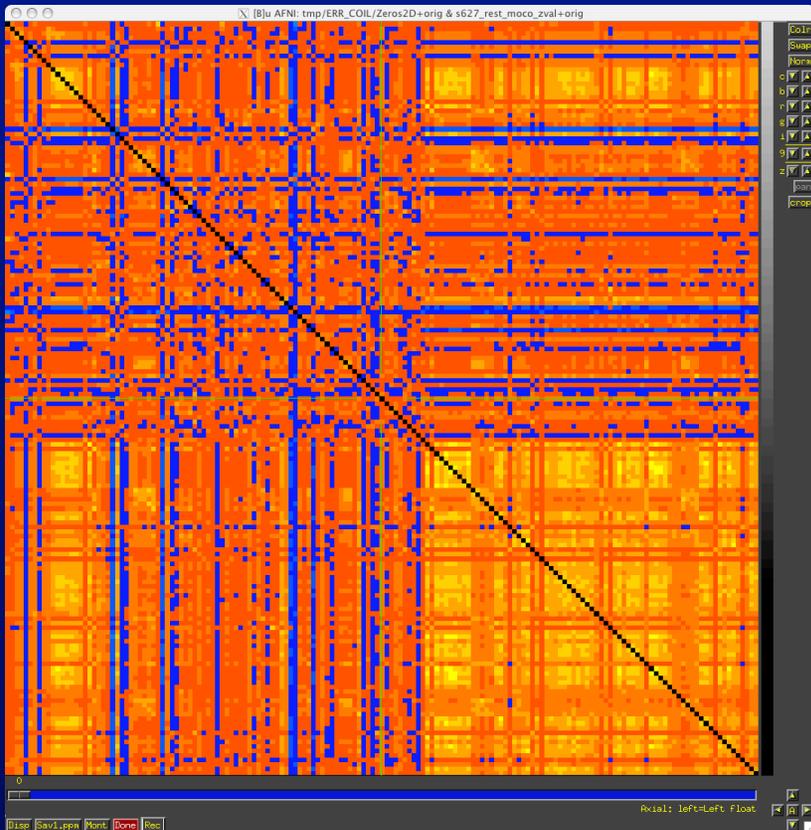
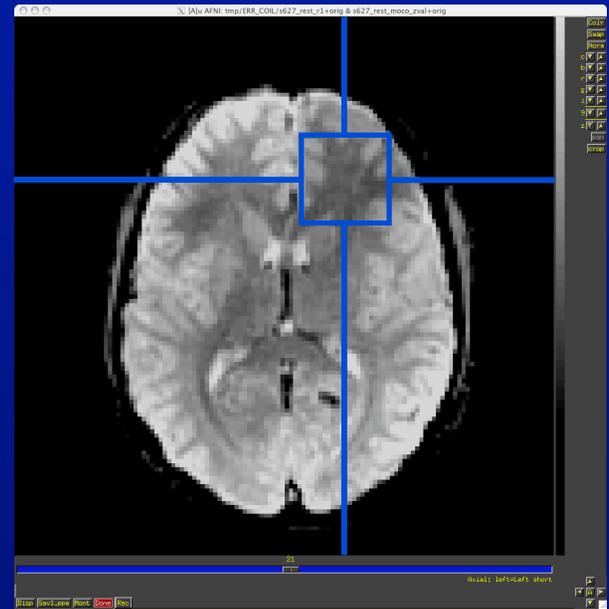
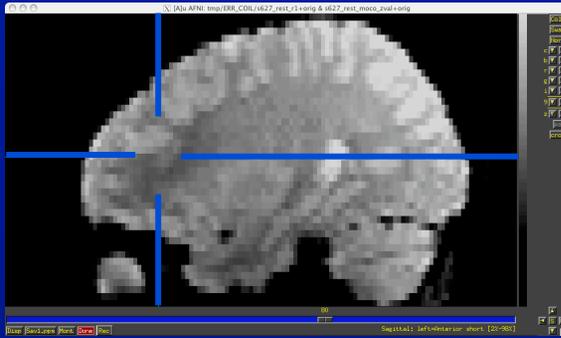
Weirder Spikes



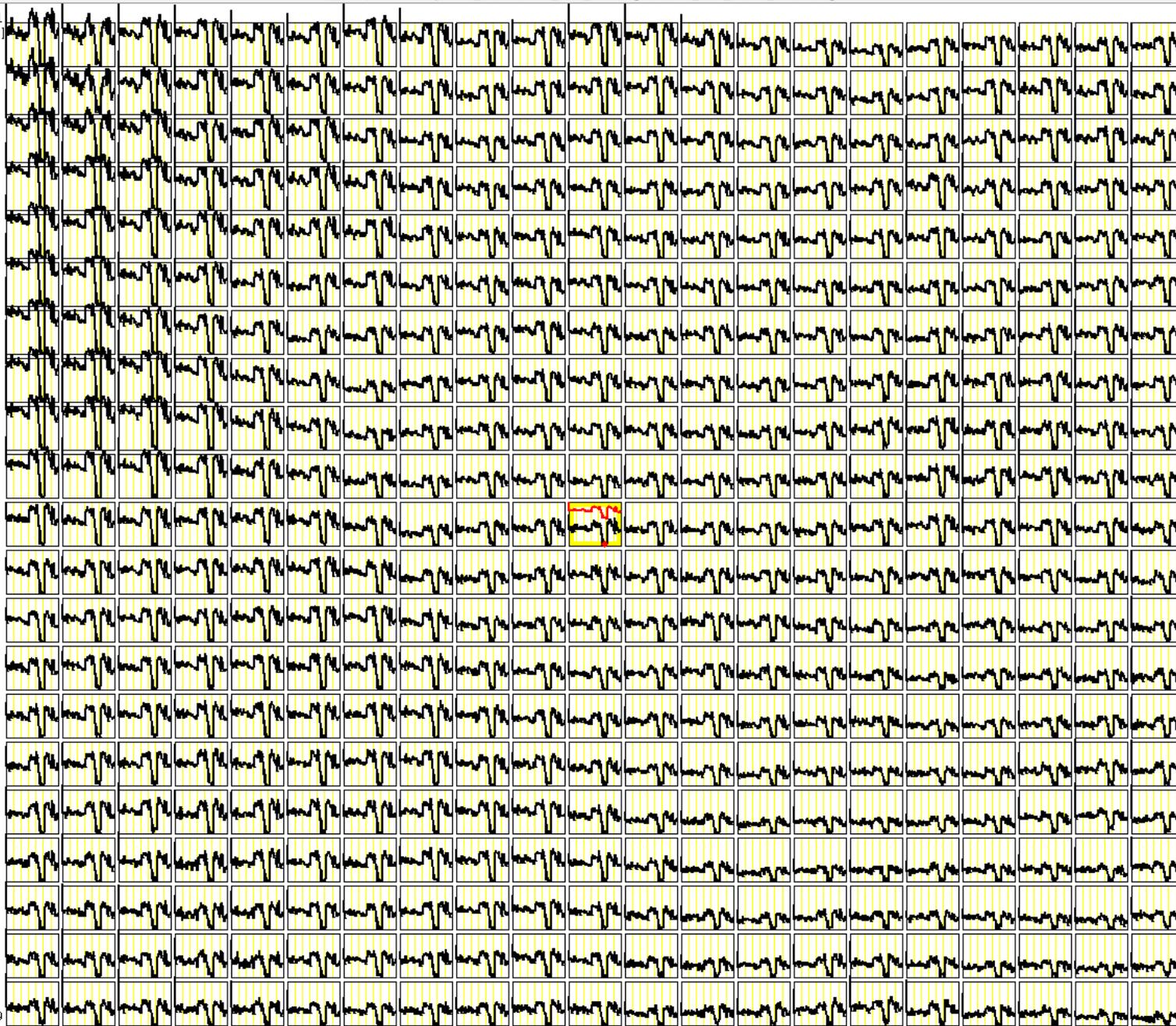
Weirder Spikes



Weirder Spikes



4141
[+1722]



index=97 value=2535 at 340.3333

2419

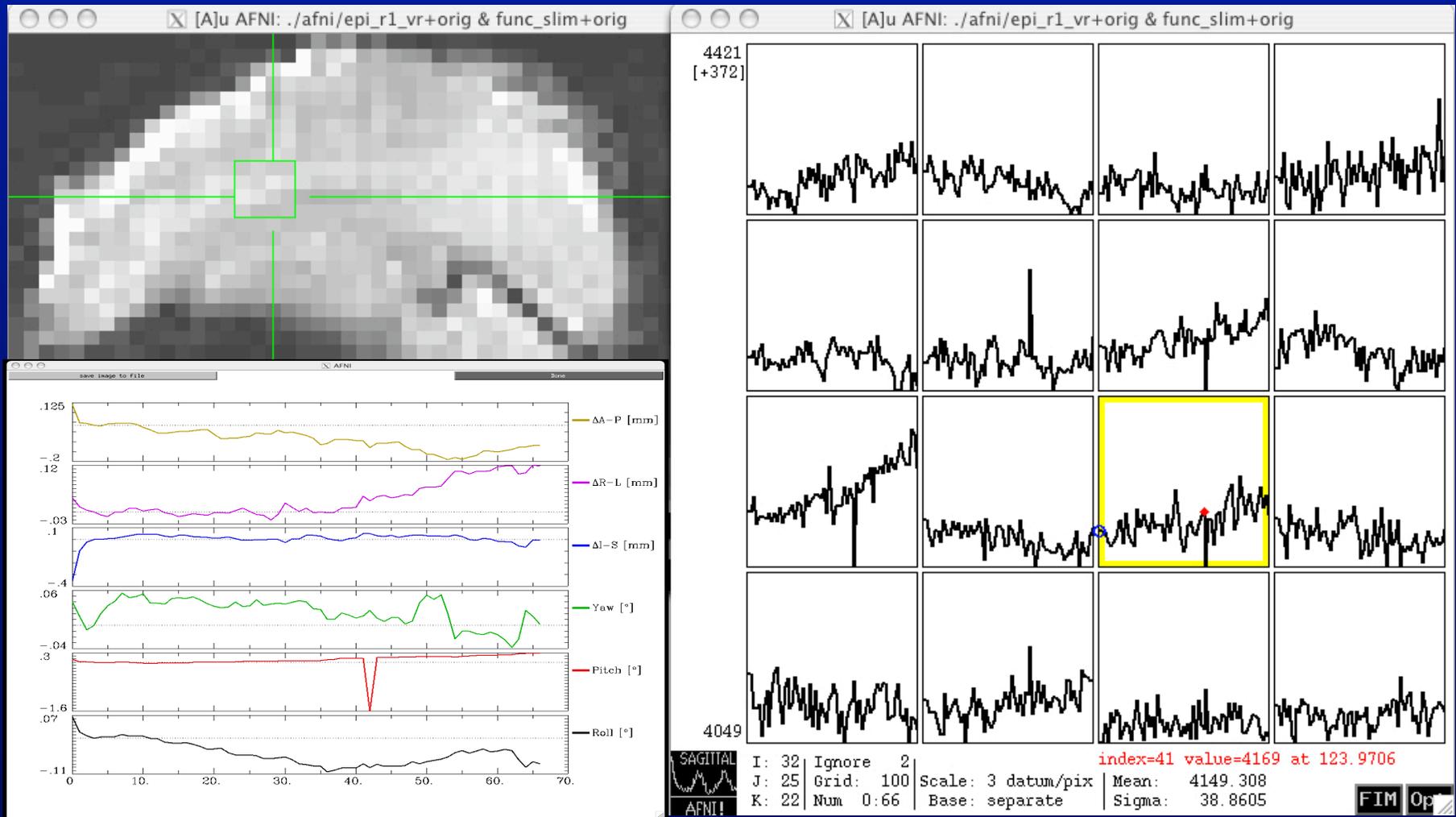


I: 49
J: 38 Grid: 20 | Scale: 41 datum/pix Mean: 3128.243
K: 20 # 0.139 Base: separate Sigma: 240.6796

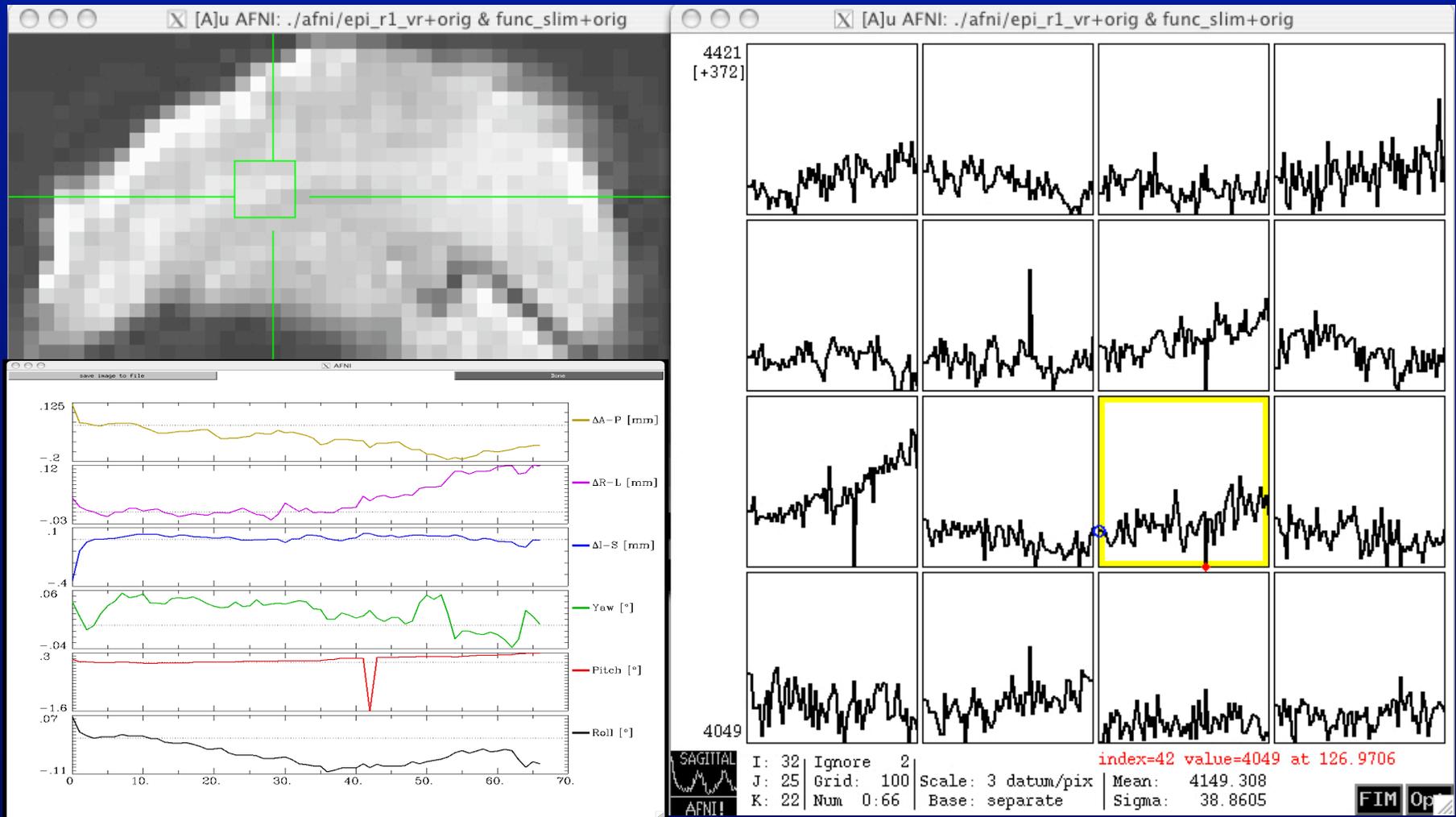
Motion Correction

- Within-modality: $T2^*$ to $T2^*$ or $T1$ to $T1$
 - Least squares cost functional is simple and robust
 - For EPI time series, rigid body (6 parameters) is typically used.
- Cross modality registration $T1$ to $T2^*$ for example
 - A variety of joint histogram based cost functionals
 - Elegant and general purpose.
 - But they can reach lowest cost at bad alignment
 - We propose the use of Local Pearson Correlation for an EPI to $T1$ cost functional

Movement Corrected spikes remain



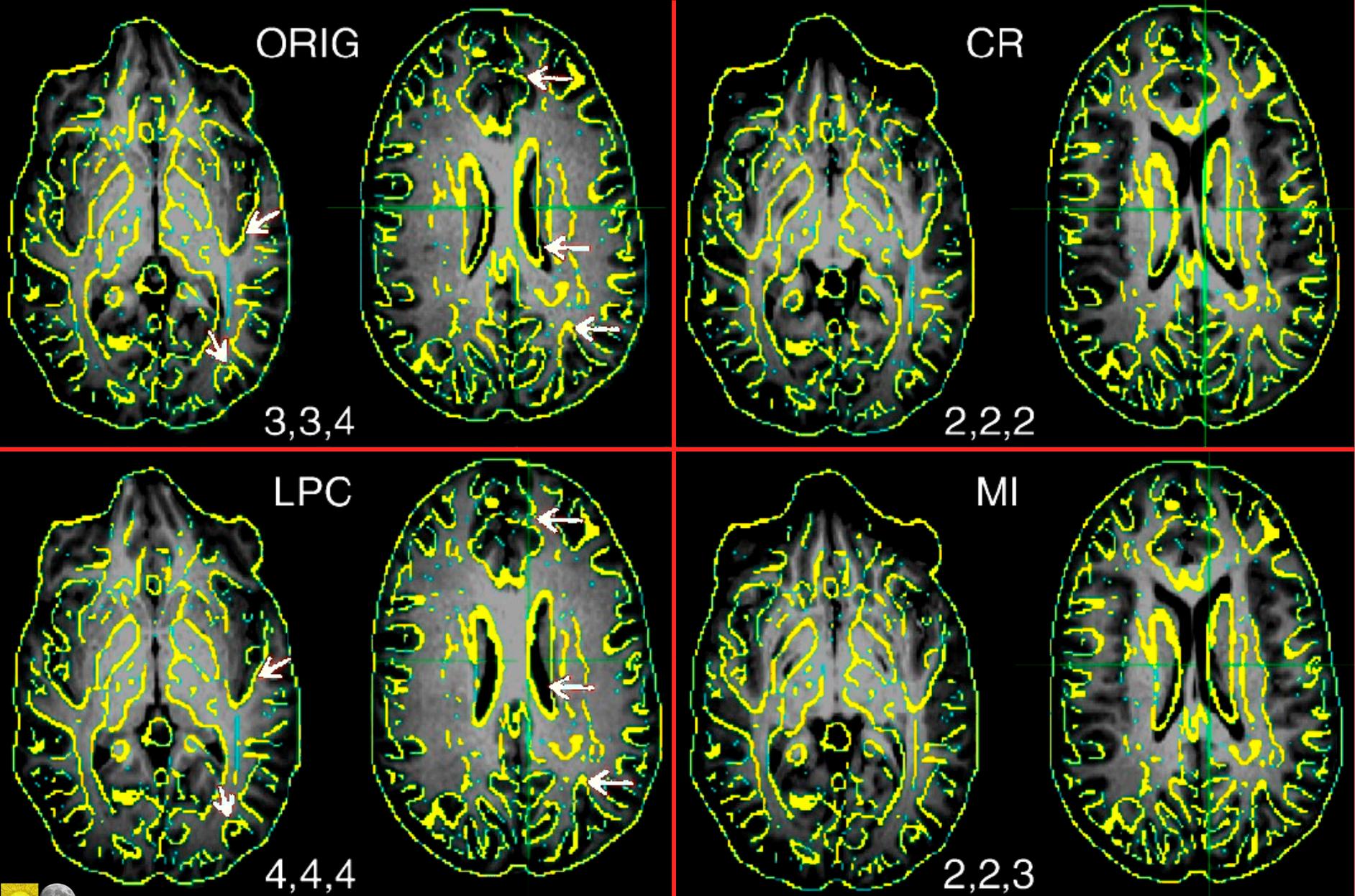
Movement Corrected spikes remain



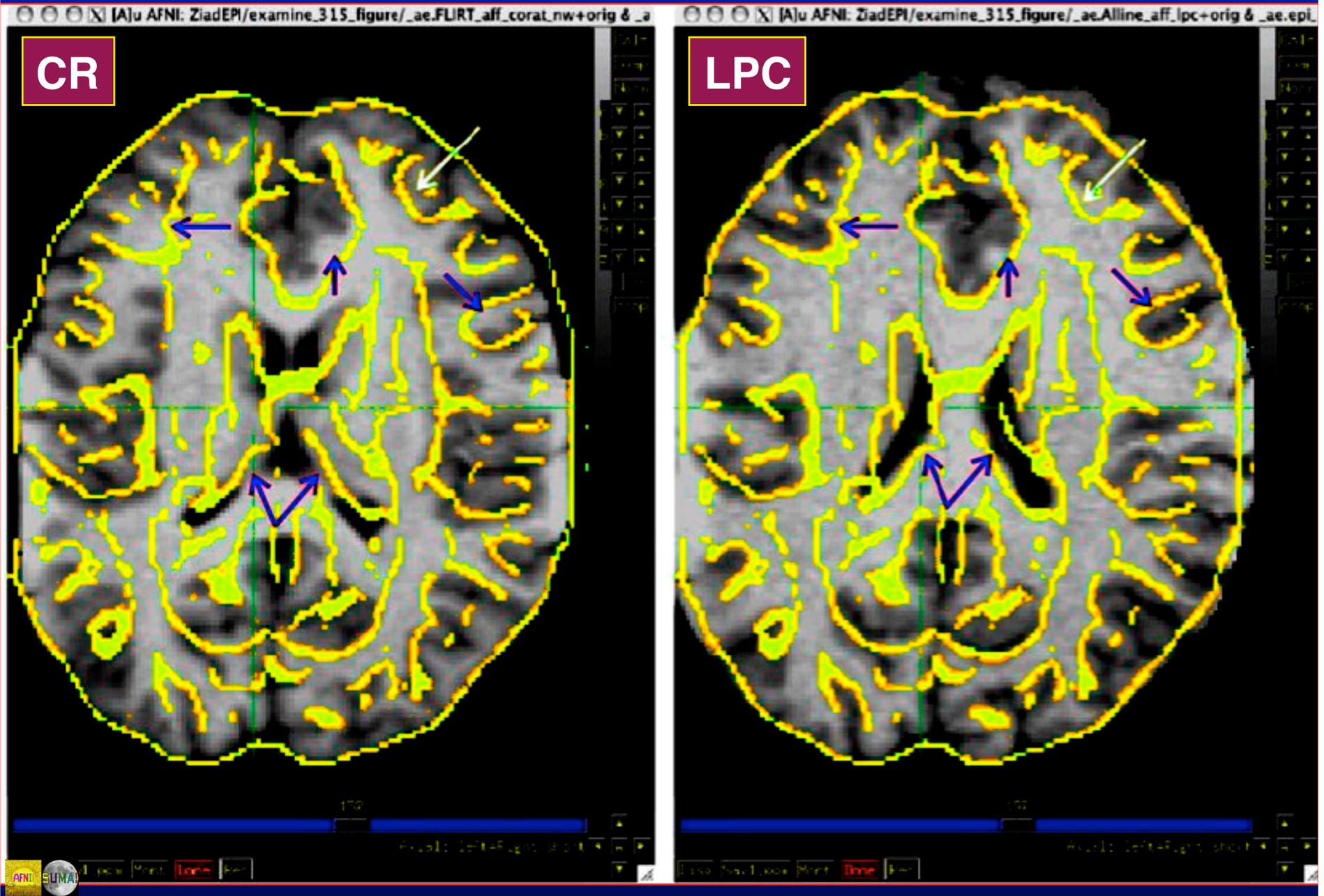
Motion Correction

- Within-modality: T2* to T2* or T1 to T1
 - Least squares cost functional is simple and robust
 - For EPI time series, rigid body (6 parameters) is typically used.
- Cross modality registration T1 to T2* for example
 - A variety of joint histogram based cost functionals
 - Elegant and general purpose.
 - But they can reach lowest cost at bad alignment
 - We propose the use of Local Pearson Correlation for an EPI to T1 cost functional

Results: EPI Edges Atop Anatomical Slices



Results: EPI Edges Atop Anatomical Slices

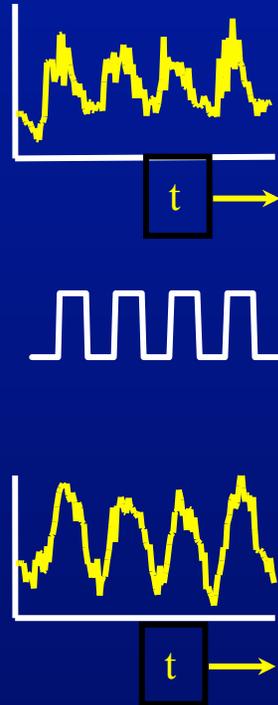
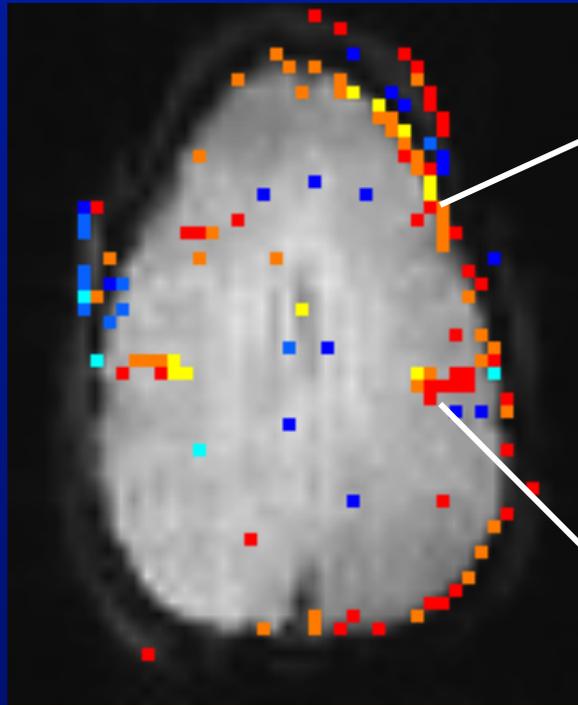


Stimulus Correlated Movement

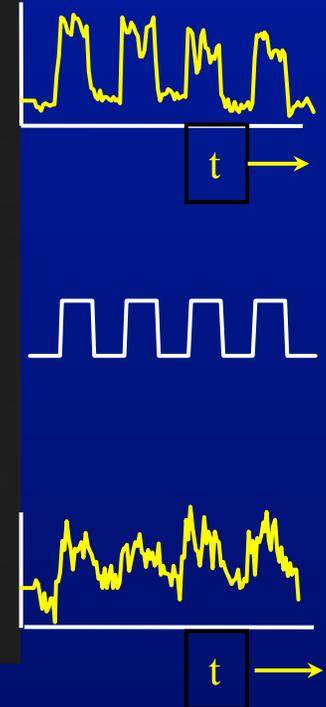
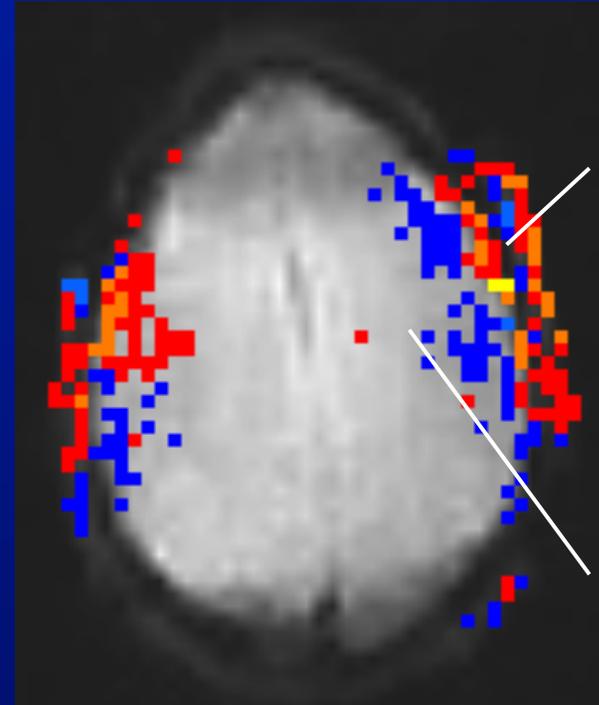
- By accident
 - Stimulus induced
 - Could confound results
 - Can happen in subtle ways as tensing up shoulders or changing breathing depth
 - Warning sign is stimulus-correlated signals on edge of brain
 - Careful consideration of stimulus timing can reduce this problem
 - Uncorrelated with Stimulus
 - Adds variance to data, resulting in less power
- By design
 - Speech production, swallowing, etc.

"Activation" Artifacts

R.M. Birn, et al. Human Brain Mapping 7(2), 106-114, 1999



overt speaking



jaw clenching

- Non-BOLD signal changes correlated with task timing

Slide courtesy of R. Birn

Z.S.S 6-13

Look at single subject results

- Consider response to task
- Multiple comparison corrections

Multi-Voxel Statistics

Spatial Clustering
&

False Discovery Rate:

“Correcting” the Significance

Basic Problem

- Usually have 50-200K FMRI voxels in the brain
- Have to make at least one decision about each one:
 - Is it “active”?
 - That is, does its time series match the temporal pattern of activity we expect?
 - Is it differentially active?
 - That is, is the BOLD signal change in task #1 different from task #2?
- Statistical analysis is designed to control the error rate of these decisions
 - Making *lots* of decisions: hard to get perfection in statistical testing

Multiple Testing Corrections

• Two types of errors

- **What is H_0 in FMRI studies?** H_0 = no effect (activation, difference, ...) at a voxel
- Type I error = Prob(reject H_0 when H_0 is true) = false positive = p value
- Type II error = Prob(accept H_0 when H_1 is true) = false negative = β
- **power** = $1-\beta$ = probability of detecting true activation
- Strategy: controlling type I error while increasing power (decreasing type II errors)
- Significance level α (magic number 0.05) : $p < \alpha$

Justice System: Trial

Hidden Truth

	Defendant Innocent	Defendant Guilty
Reject Presumption of Innocence (Guilty Verdict)	Type I Error (defendant very unhappy)	Correct
Fail to Reject Presumption of Innocence (Not Guilty Verdict)	Correct	Type II Error (defendant very happy)

Statistics: Hypothesis Test

Hidden Truth

	H_0 True Not Activated	H_0 False Activated
Reject H_0 (decide voxel is activated)	Type I Error (false positive)	Correct
Don't Reject H_0 (decide voxel isn't activated)	Correct	Type II Error (false negative)

- **Family-Wise Error (FWE)**

- Multiple testing problem: voxel-wise statistical analysis
 - With N voxels, what is the chance to make a false positive error (Type I) in one or more voxels?

Family-Wise Error: $\alpha_{FW} = 1 - (1 - p)^N \rightarrow 1$ as N increases

- For Np small (compared to 1), $\alpha_{FW} \approx Np$
- $N \approx 50,000+$ voxels in the brain
- To keep probability of even one false positive $\alpha_{FW} < 0.05$ (the “corrected” p -value), need to have $p < 0.05 / 5 \times 10^4 = 10^{-6}$
- This constraint on the per-voxel (“uncorrected”) p -value is so stringent that we would end up rejecting a lot of true positives (Type II errors) also, just to be safe on the Type I error rate

- **Multiple testing problem in FMRI**

- 3 occurrences of multiple tests: Individual, Group, and Conjunction
- Group analysis is the most severe situation (have the least data, considered as number of independent samples = subjects)

- **Two Approaches to the “Curse of Multiple Comparisons”**
 - Control **FWE** to keep expected total number of false positives below 1
 - Overall significance: $\alpha_{FW} = \text{Prob}(\geq \text{one false positive voxel in the whole brain})$
 - **Bonferroni correction**: $\alpha_{FW} = 1 - (1-p)^N \approx Np$, if $p \ll N^{-1}$
 - Use $p = \alpha/N$ as individual voxel significance level to achieve $\alpha_{FW} = \alpha$
 - Too stringent and overly conservative: $p = 10^{-8} \dots 10^{-6}$
 - What can rescue us from this hell of statistical super-conservatism?
 - Correlation: Voxels in the brain are not independent
 - Especially after we smooth them together!
 - Means that Bonferroni correction is *way way* too stringent
 - Contiguity: Structures in the brain activation map
 - We are looking for activated “blobs”: the chance that pure noise (H_0) will give a set of seemingly-activated voxels next to each other is lower than getting false positives that are scattered around far apart
 - Control FWE based on spatial correlation (smoothness of image noise) **and** minimum cluster size we are willing to accept

 - Control false discovery rate (**FDR**) — Much more on this a little later!
 - FDR = expected proportion of false positive voxels among all **detected** voxels
 - Give up on the idea of having (almost) no false positives at all

Group Analysis

[File: GroupAna.pdf](#)

Gang Chen

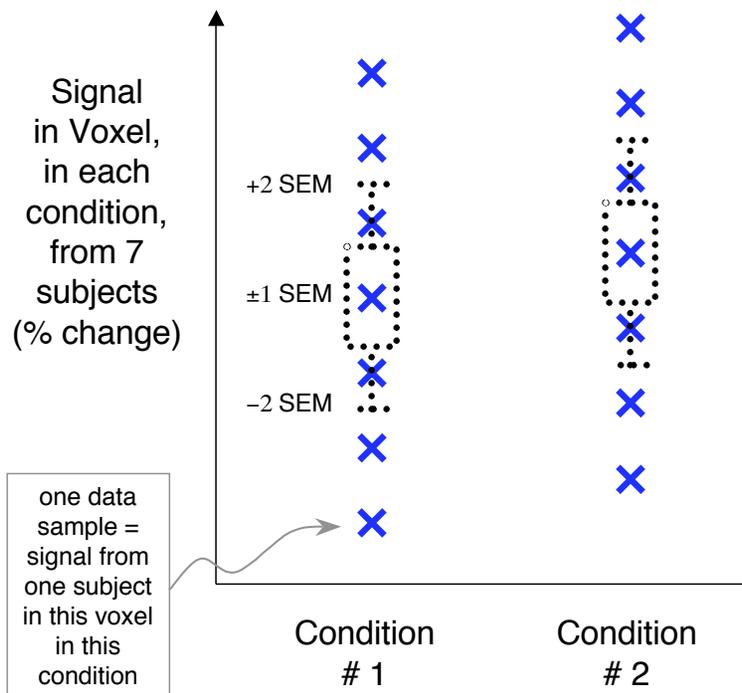
SSCC/NIMH/NIH/HHS



Why Group Analysis?

- Summarizing individual subject results
- Why not one analysis with a mega model for all subjects?
 - Computationally unmanageable
 - Heterogeneity in data or experiment design across subjects
- What is a valid summarizing method?
 - Effect of subject $i =$ group effect + deviation of subject i
 - A simple (one-sample t -test) model $\beta_i = b + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$
 - If individual effects are consistent across most or all subjects, the deviations would be relatively small
 - Significance measure = group effect relative to variability
 - Student t -test as a simple illustration

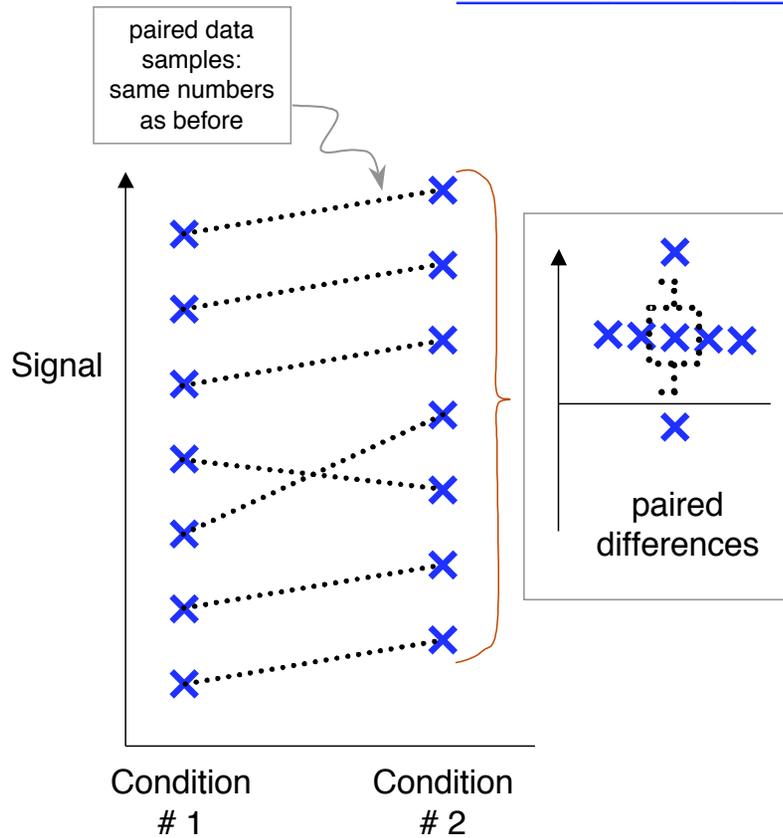
Unpaired 2 Sample *t*-Test: Cartoon Data



• Not significantly different!

- Condition = some way to categorize data (*e.g.*, stimulus type, drug treatment, day of scanning, subject type, ...)
- SEM = Standard Error of the Mean
= standard deviation of sample divided by square root of number of samples
= estimate of uncertainty in sample mean
- Unpaired *t*-test determines if sample means are “far apart” compared to size of SEM
 - *t* statistic is difference of means divided by SEM

Paired *t*-Test: Cartoon Data



- Paired means that samples in different conditions should be linked together (*e.g.*, from same subjects)
- Test determines if differences between conditions in each pair are “large” compared to SEM of the differences
- Paired test can detect systematic *intra*-subject differences that can be hidden in *inter*-subject variations
- Lesson: properly separating *inter*-subject and *intra*-subject signal variations can be very important!

- Significantly different!
- Condition #2 > #1, per subject

Terminology: Fixed factor/effect - discrete variable

- Treated as a **fixed** variable (constant) in the model
 - Categorization of conditions/tasks (modality: visual/auditory)
 - Within-subject (repeated-measures) factor
 - Subject-grouping: Group of subjects (gender, normal/patients)
 - Between-subject factor
- All **levels** of a factor are of interest (house vs. face)
 - main effect, contrasts among levels
- Fixed in the sense of statistical inferences
 - apply only to the specific levels of the factor
 - don't extend to other potential levels that might have been included
- Fixed effects may also include **continuous** variables (covariates)
 - Of direct interest
 - Improving statistical power by controlling for data variability

- Terminology: Random factor/effect

- ☞ Random variable in the model: exclusively subject in FMRI

- average + effects uniquely attributable to each subject: e.g. $N(\mu, \tau^2)$

- Requires enough number of subjects

- ☞ Each individual subject effect is of NO interest

- ☞ Random in the sense

- subjects serve as a random sample (representation) from a population

- inferences can be generalized to a hypothetical population

- Fixed vs. random effects

- ☞ Conventional model $\beta_i = b + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$

- ☞ Linear mixed-effects model $\beta_i = b + \delta_i + \varepsilon_i$, $\delta_i \sim N(0, \tau^2)$, $\varepsilon_i \sim N(0, \sigma^2)$

- ↙ b : universal constant

- ↙ δ_i : each subject's unique and consistent personality

- ↙ ε_i : random fluctuations in life

Covariates

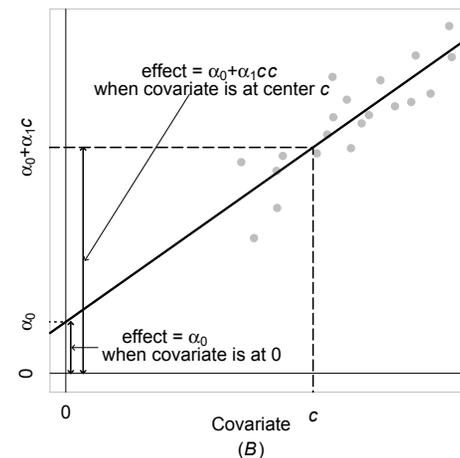
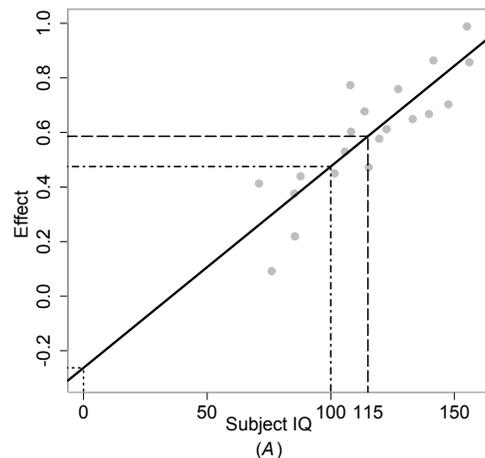
- ❑ Confusing usage in literature
 - ❑ May or may not be of direct interest
 - Direct interest: relation between response and the covariate
 - ❑ Is response proportional to response time?
 - Of no interest: confounding, nuisance, or interacting variables
 - ❑ Controlling for or covarying or partialling out: what does it mean?
 - ❑ Subtle issue in this case: centering
 - ❑ Continuous or discrete
 - Continuous: historically originated from ANCOVA
 - I solely use it as a continuous variable to avoid confusion
 - Very careful when treating a discrete (categorical) variable as covariate
 - ❑ Dummy coding
 - ❑ Interaction
-

Covariate: Modeling framework

- Most people learned covariate modeling with ANCOVA
 - ❑ Historical extension to ANOVA
 - ❑ Quite limited and not flexible
 - ❑ Not a good approach in general
 - GLM or LME: broader context
 - ❑ All explanatory variables are treated equally in the model
 - ❑ Doesn't matter: variable of interest or not, discrete or continuous
 - ❑ Discrimination or categorization occurs only at human (not model) level
-

Handling covariates: one group

- Model $y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i$, for i th subject: **no other variables**
 - α_1 - slope (change rate, marginal effect): effect per unit of x
 - Simple and straightforward: no manipulation needed
 - α_0 – intercept ($x=0$): group effect while controlling x
 - Controlling is NOT good enough
 - Interpretability - α_0 at what x value: mean or any other value?
 - **Centering** is crucial for interpretability
 - Center does not have to be mean



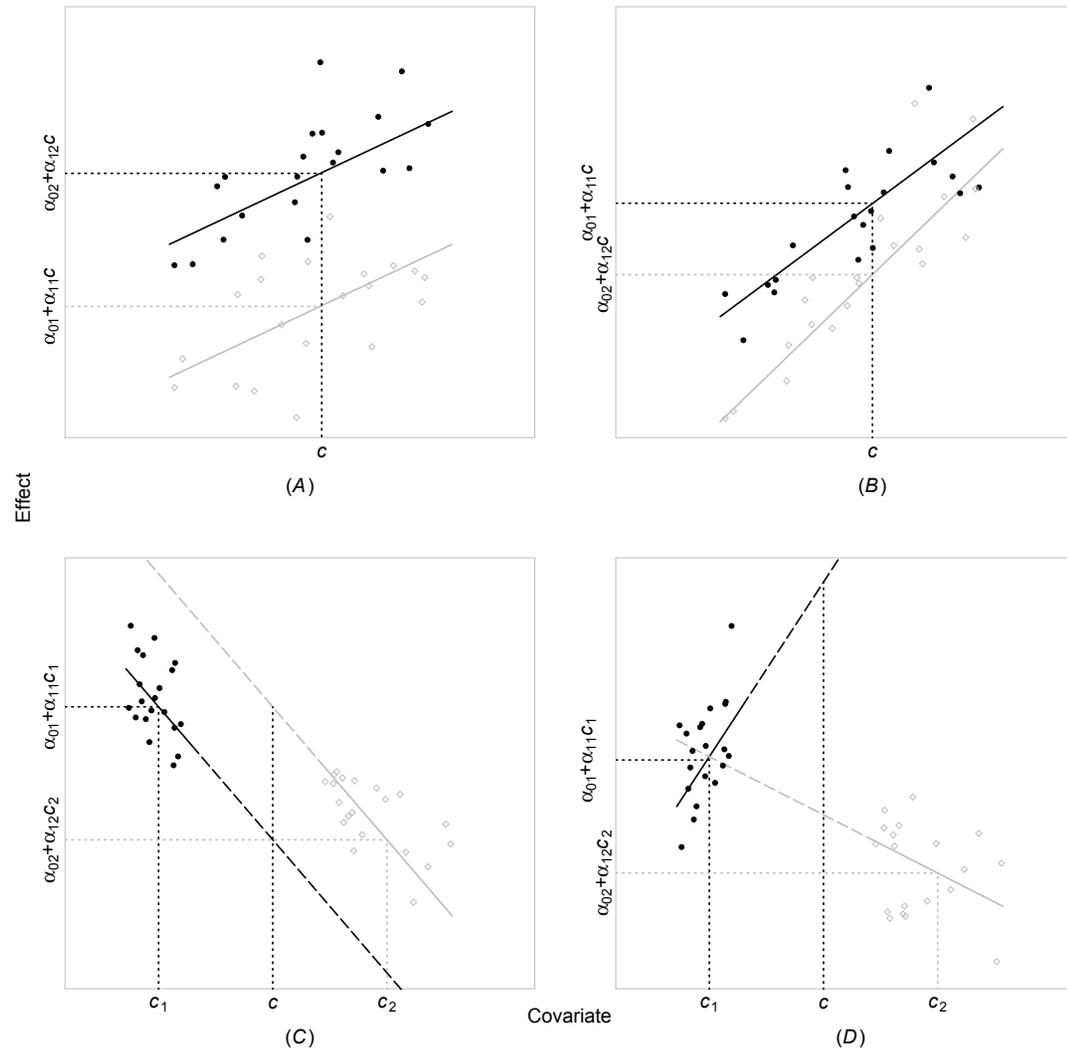
Covariates: two or more groups

- ❑ Slope
 - ❑ Same or different across groups?
 - ❑ Usually we don't know in advance
 - Start with different slopes – interaction between group and covariate
 - If same, then model tuning
 - ❑ Intercept: **centering** again
 - ❑ Same or different center across groups?
 - ❑ How to decide? Plot out covariate distribution
 - ❑ If about the same, nice and easy!
 - ❑ If dramatically different, now what?
 - ❑ If possible, this issue should have been thought of when designing the experiment
 - You may balance covariate values (e.g. age) across groups
 - How about if it is not under your control (e.g., response time)?
-

Covariates: different center across groups

- Most statisticians (including in FMRI) consider it **horrible**
 - ❑ For example, Miller GM and Chapman JP. '*Misunderstanding analysis of covariance*', J Abnormal Psych 110: 40-48 (2001)
 - ❑ SPM and FSL communities
 - ❑ It may well be the case
 - Groups were not balanced in experiment design: design failure!
 - E.g., males and females have different age distribution, and we can't resolve: in the end the group difference is due to sex or age difference?
 - ❑ But I beg to differ under other scenarios
 - Now stop and think!
 - What is the point of considering the covariate? Using RT as example, we can **account for within-group variability of RT, not variability across all subjects in both groups**
 - Do not center by default without careful forethought
-

Slope and intercept with two groups



Acknowledgments

Robert Cox
Gang Chen
Rick Reynolds
Daniel Glen
Michael Beauchamp
Rasmus Birn

Most of these slides were taken from the AFNI bootcamp class material.

For complete set of documentation and test data:

http://afni.nimh.nih.gov/pub/dist/HOWTO/howto/ht00_inst/html/index.shtml

and

http://afni.nimh.nih.gov/pub/dist/HOWTO/howto/ht00_inst/html/data_handouts.shtml