Hands-On Session: Regression Analysis

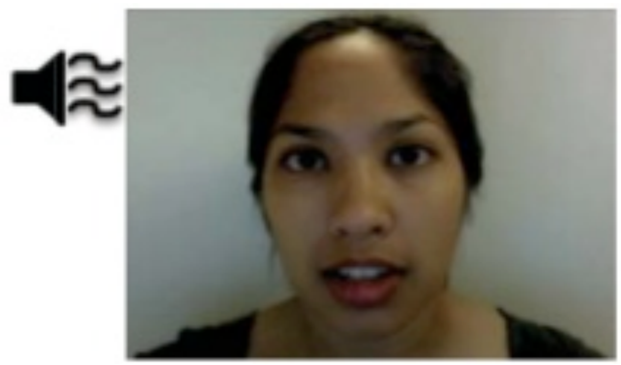
- What we have learned so far
  - Use data viewer 'afni' interactively
  - Model HRF with a fixed-shape basis function
    - Assume the brain responds with the **same shape**
      - in any active regions
      - assumed for each stimulus type
    - Difference in magnitude: β (and its significance) is what we focus on
- What we will do in this session
  - Data pre-processing overview for time series regression analysis
  - Basic concepts of regressors, design matrix, and confounding effects
  - Statistical significance testing in regression analysis
  - Navigation with GUI '**afni**'
    - Spot check for the original data
    - Statistics thresholding with data viewer '**afni**' (two-sided vs. one-tailed with t)
    - Model performance (visual check of curve fitting and test via full F or R^2)

A Case 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

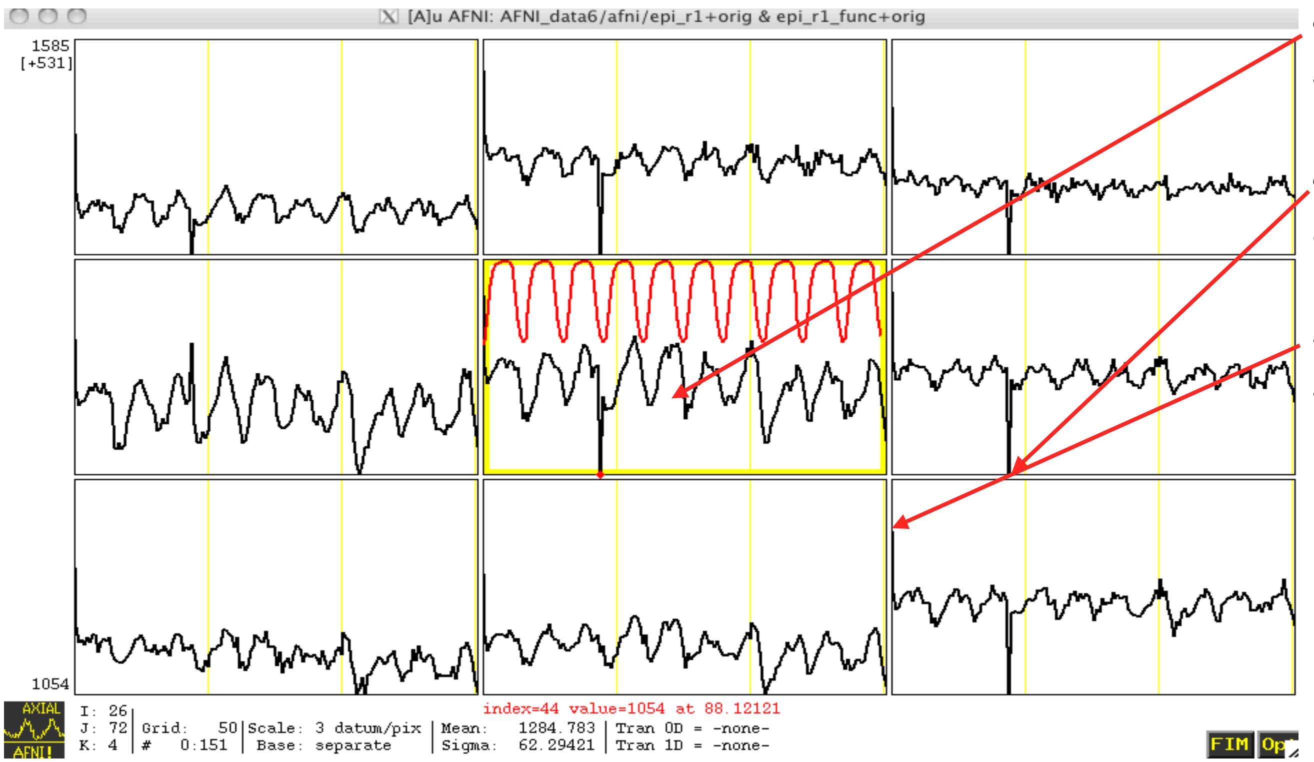
- 3 runs in a scanning session
- Each run consisted of 10 blocked trials:
  - 5 blocks contained Auditory-Reliable (*Arel*) stimuli, and
  - 5 blocks contained Visual-Reliable (*Vrel*) stimuli
- Each block contained 10 trials of *Arel* OR *Vrel* stimuli
  - Each block lasted for 20s (1s for stimulus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

Data Collected

- 2 anatomical datasets for each subject from a 3T
  - 175 sagittal slices
  - voxel dimensions = $1.0 \times 0.938 \times 0.938$ mm
- 3 time series (EPI) datasets for each subject
  - 33 axial slices x 152 volumes (TRs) per run
  - TR = 2s; voxel dimensions = $2.75 \times 2.75 \times 3.0$ mm
- Sample size, $\underline{n}$ = 10 (all right-handed subjects)

Data Quality Check

- To look at the data: type `cd AFNI_data6/afni`, then `afni`
- `Switch Underlay` to dataset `epi_r1`
  - Then `Axial Image` and `Graph`
  - `FIM→Pick Ideal` ; then click `afni/epi_r1_ideal.1D` ; then `Set`
  - Right-click in image, `Jump to (ijk)`, then `26 72 4`, then `Set`



- Data clearly has activity in sync with reference
  - 20s blocks
- Data also has a big spike at 89s
  - Head motion
- Spike at $t = 0$
- Some keyboard shortcuts
  - a: automatic scaling
  - v: video mode
  - m/M: decrease or increase number of voxels in graph matrix

Preparing Data for Analysis

- Common pre-processing steps (and relevant programs):
  - Outliers: **3dToutcount** (or **3dTqual**), **3dDespike**
  - Temporal alignment (slice timing correction) of sequential/interleaved slices: **3dTshift**
  - Image/volume registration (head motion correction): **3dvolreg**
  - Anatomy ←→ EPI registration: **align_epi_anat.py**
  - Spatial normalization (standard space conversion): **@auto_tlrc, adwarp**
  - Blurring/smoothing: **3dmerge, 3dBlurToFWHM, 3dBlurInMask**
  - Masking: **3dAutomask, 3dSkullStrip**
  - Temporal mean scaling: **3dTstat** and **3dcalc**
  - Global mean scaling*: **3dROIstats** (or **3dmaskave**) and **3dcalc**

- Not all steps are necessary or desirable in any given case

Regression Analysis

- Run script by typing **tcsh rall_regress** (takes a few minutes)

```
3dDeconvolve -input rall_vr+orig                                          \
    -concat '1D: 0 150 300'                                               \
    -polort 1                                                             \
    -num_stimts 8                                                         \
    -stim_times 1 stim_AV1_vis.txt 'BLOCK(20,1)' -stim_label 1 Vrel      \
    -stim_times 2 stim_AV2_aud.txt 'BLOCK(20,1)' -stim_label 2 Arel      \
    -stim_file 3 motion.1D'[0]' -stim_base 3 -stim_label 3 roll          \
    -stim_file 4 motion.1D'[1]' -stim_base 4 -stim_label 4 pitch         \
    -stim_file 5 motion.1D'[2]' -stim_base 5 -stim_label 5 yaw           \
    -stim_file 6 motion.1D'[3]' -stim_base 6 -stim_label 6 dS            \
    -stim_file 7 motion.1D'[4]' -stim_base 7 -stim_label 7 dL            \
    -stim_file 8 motion.1D'[5]' -stim_base 8 -stim_label 8 dP            \
    -gltsym 'SYM: Vrel -Arel' -glt_label 1 V-A                           \
    -tout -x1D rall_X.xmat.1D -xjpeg rall_X.jpg                          \
    -fitts rall_fitts -bucket rall_func                                   \
    -jobs 2
```

- 2 audiovisual stimulus classes were given using `-stim_times`
    - basis functions were specified as **`BLOCK(20,1)`**
- **Important to include motion parameters as regressors**?
    - May remove the confounding effects due to motion artifacts
    - 6 motion parameters as covariates via `-stim_file` + `-stim_base`
    - `motion.1D` generated from `3dvolreg` with the `-1Dfile` option
    - Test the significance of head motion parameters
        - Use **`-gltsym 'SYM: roll \ pitch \ yaw \ dS \ dL \ dP'`**
        - Use **`-fout`** to get a composite F-stat from the 6 tests

- Regression: fit model to data as well as possible (in least-squares sense)
    - model defines what is expected to be seen in the data
    - basis functions should accurately reflect BOLD responses
    - confounding effects are modeled, as well as effects of interest

- Inputs for regression: data and model
    - EPI data, with run break indexes (since dataset contains all 3 runs)
    - degree of polynomial baseline, to be applied for each run
    - regressor timing, with desired basis functions
    - other regressors, to be applied as they are (no basis functions)
        - here, 6 motion parameter curves as regressors of no interest

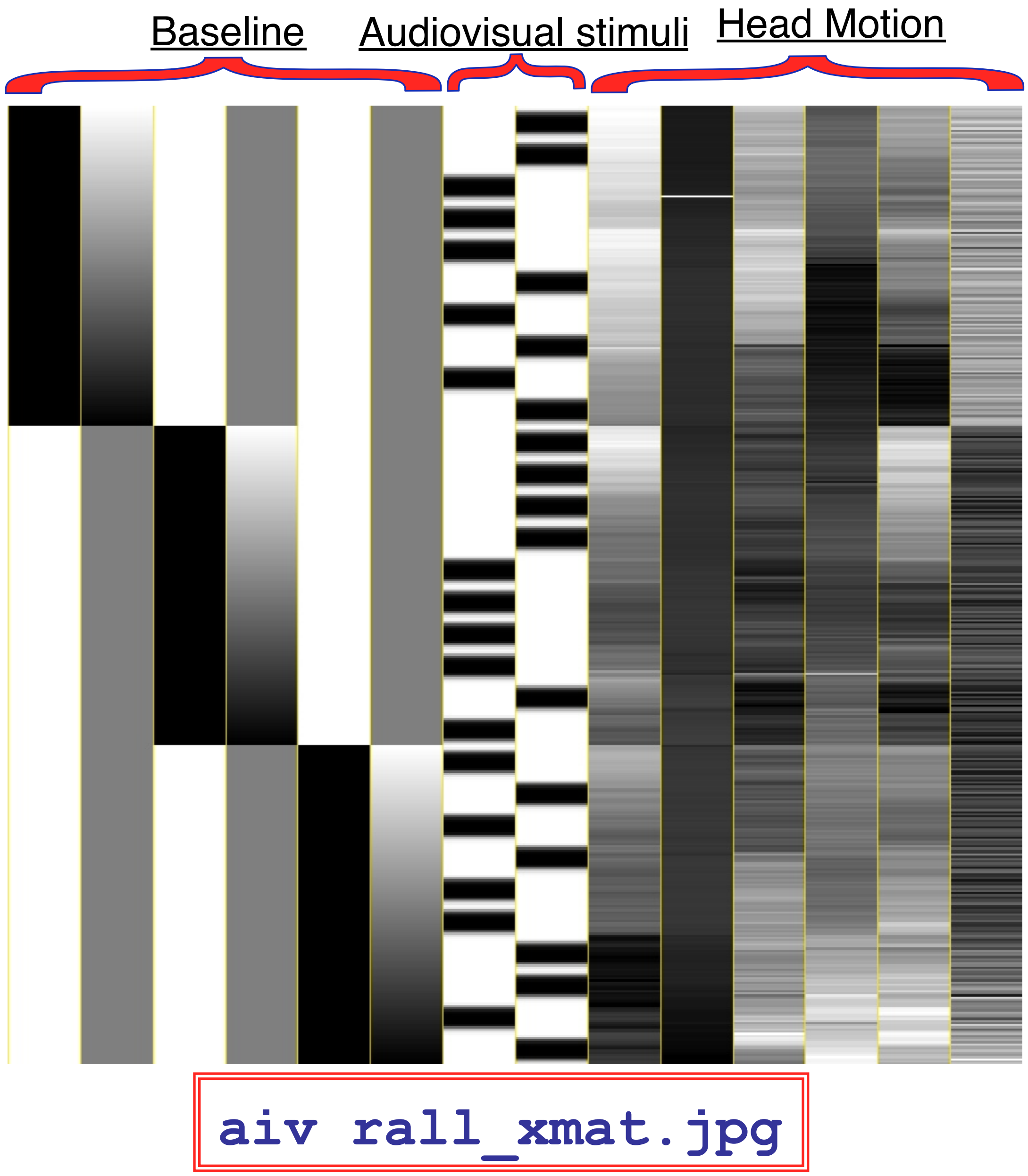
- Output from regression (what was requested):
    - applied model written to **`rall_X.xmat.1D`** (image to **`rall_X.jpg`**)
    - model fit to the data written to **`rall_fitts+orig`**
    - bucket dataset of betas and statistics written to **`rall_func+orig`**
        - includes requested contrasts and resulting statistics

Modeling Serial Correlation in the Residuals

- Temporal correlation exists in the residuals of the time series regression model
- Within-subject variability (or statistical value) would get deflated (or inflated) if temporal correlation is not accounted for in the model
- Better correct for the temporal correlation if bringing both effect size and within-subject variability to group analysis
- ARMA(1, 1) assumed in **3dREMLfit**
- Script automatically generated by 3dDeconvolve (may use –x1D_stop)
  - ⋆ File **rall_func.REML_cmd** under **AFNI_data6/afni**
  - ⋆ Run it by typing **tcsh –x rall_func.REML_cmd**

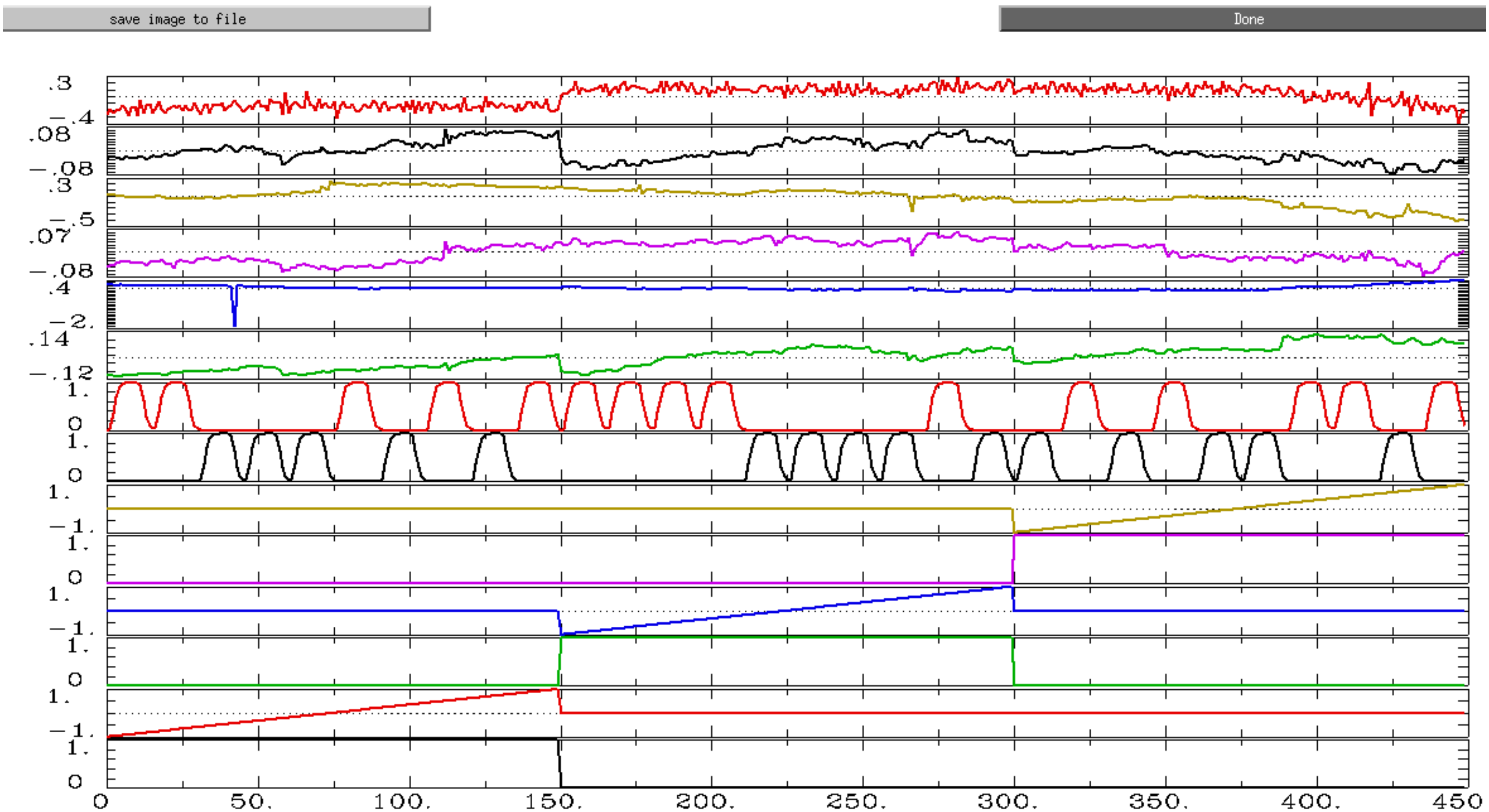
```
3dREMLfit -matrix rall_X.xmat.1D -input rall_vr+orig \
 -tout -Rbuck rall_func_REML -Rvar rall_func_REMLvar \
 -Rfitts rall_fitts_REML -verb
```

Regressor Matrix for This Script (via -xjpeg)



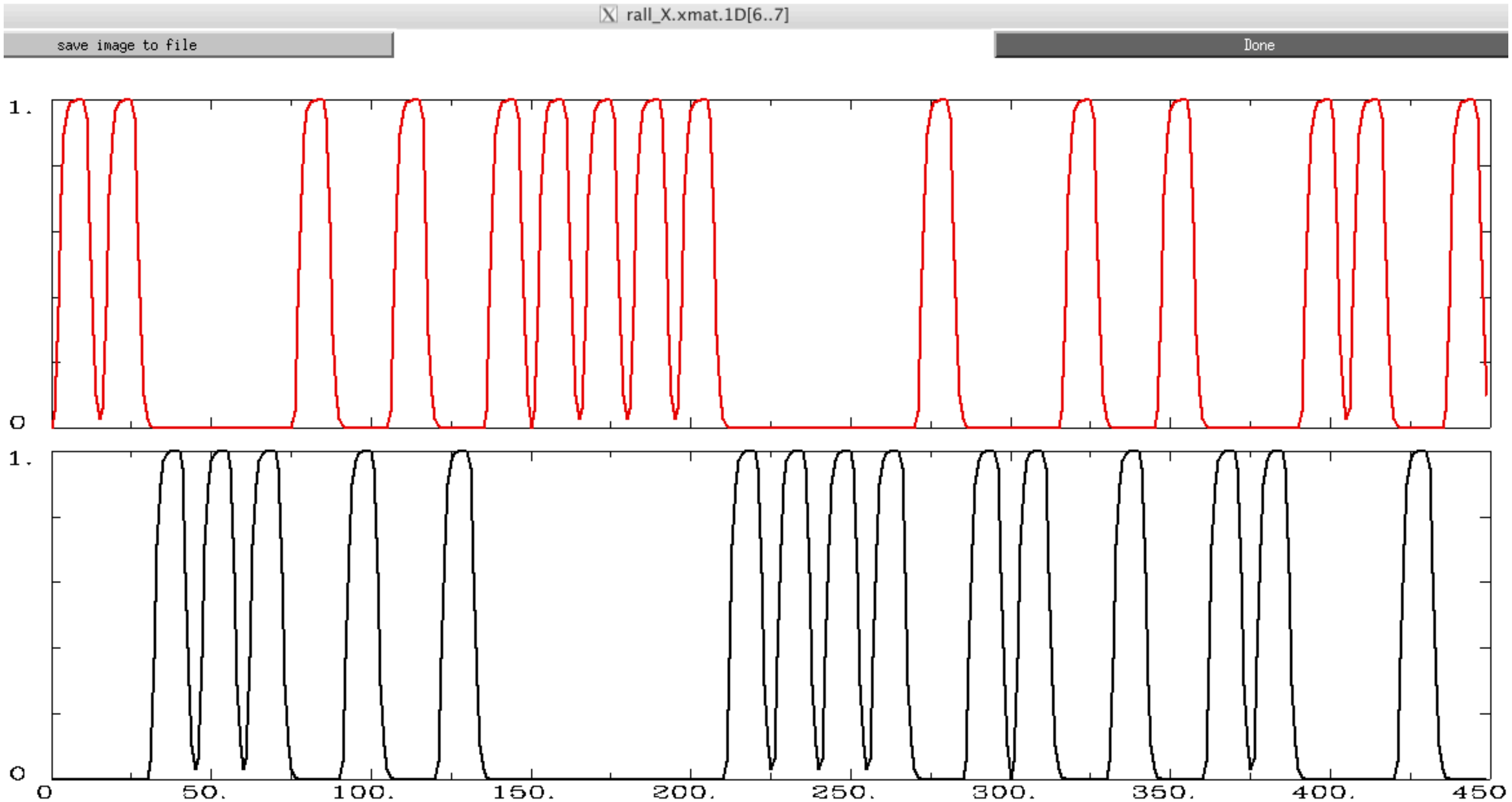
- 6 drift effect regressors
  - linear baseline
  - 3 runs times 2 params/run
- 2 regressors of interest
- 6 head motion regressors
  - 3 rotations and 3 shifts

Showing All Regressors (via -x1D)



All regressors: **1dplot -sepscl rall_X.mat.1D**

Showing Regressors of Interest



Regressors of Interest: **1dplot rall_X.mat.1D'[6..7]'**

Options in 3dDeconvolve - 1

`-concat '1D: 0 150 300'`

- ★ "File" that indicates where distinct imaging runs start inside the input file
  - ➢ Numbers are the time (TR) indexes inside the dataset for starts of runs
  - ➢ These time points are considered as discontinuities in the model
  - ➢ In this case, a text format .1D file put directly on the command line
    - ➥ Could also be a filename, if you want to store that data externally

`-polort 1`

- ★ Use a linear polynomial to model the baseline and drift in each run

`-num_stimts 8`

- ★ 2 audiovisual stimuli (+6 motion), thus 2 **`-stim_times`** below
- ★ Times given in the `-stim_times` files are *local* to the start of each run

`-stim_times 1 stim_AV1_vis.txt 'BLOCK(20,1)' -stim_label 1 Vrel`

- ★ Content of `stim_AV1_vis.txt`

```
60 90 120 180 240
120 150 180 210 270
0 60 120 150 240
```

  - o Each of 3 lines specifies start time in seconds for stimuli within the run

Options in 3dDeconvolve - 2

```
-gltsym 'SYM: Vrel -Arel' -glt_label 1 V-A
```

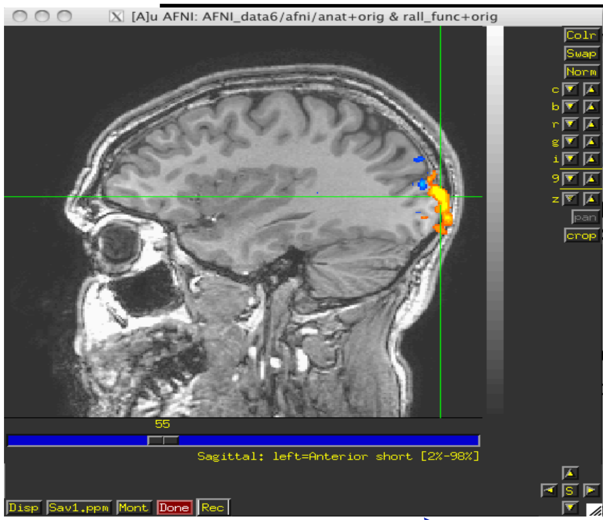
- **GLT**s: **G**eneral **L**inear **T**ests
- **3dDeconvolve** provides test statistics for each regressor separately, but to test combinations of the β weights in each voxel, use `-gltsym`
- Example above tests the difference between the β weights for the `Visual-reliable` and the `Audio-reliable` responses
  - `SYM:` means symbolic GLT form is on command line
    - Otherwise inputs will be read from a file
  - Symbolic names for each regressor taken from `-stim_label` options
  - Stimulus label can be preceded by `+` or `-` to indicate sign to use in combination of β weights, and may also
  - Leave space after each label!
- Goal is to test a linear combination of the β weights
  - Null hypothesis $\beta_{\text{Vrel}} = \beta_{\text{Arel}}$
  - e.g., does `Vrel` get different response from `Arel`?
- What do `'SYM: 0.5*Vrel +0.5*Arel'` and `'SYM: Vrel \ Arel'` test?

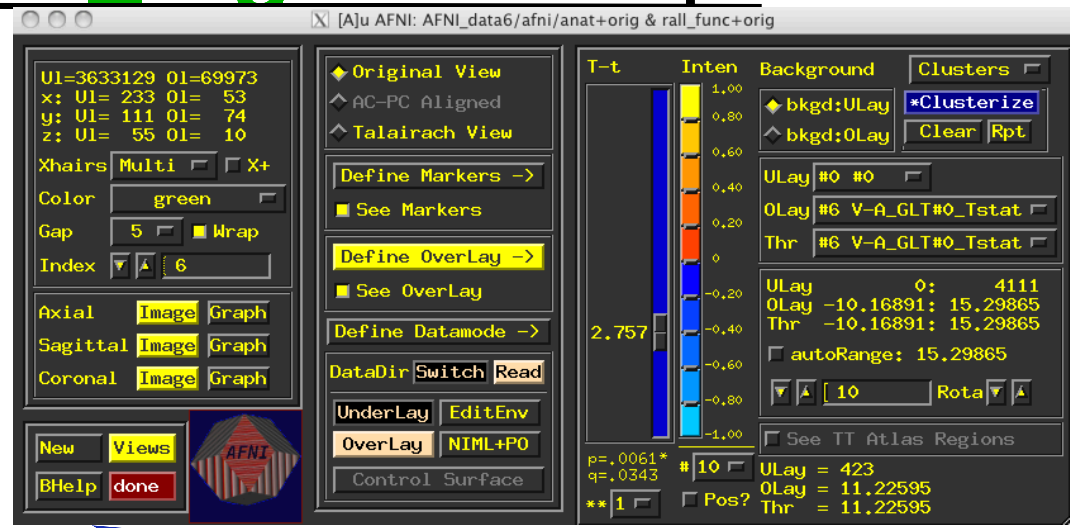
Options in 3dDeconvolve - 4

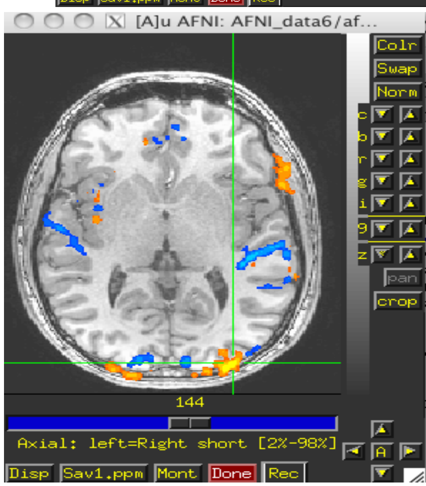
`-fout -tout` = output both *F*- and *t*-statistics for each stimulus class (`-fout`) and stimulus coefficient (`-tout`) — but not for the baseline coefficients (use `-bout` for baseline)

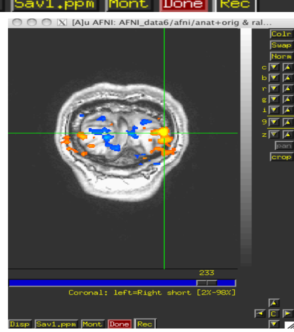
- The full model statistic is an *F*-statistic that shows how well all the regressors of interest explain the variability in the voxel time series data
  - Compared to how well *just* the baseline model time series fit the data times (in this example, we have 12 baseline regressor columns in the matrix — 6 for the linear drift, plus 6 for motion regressors)
  - $F = [SSE(r) - SSE(f)]/df(\mathrm{n}) \div [SSE(f)/df(\mathrm{d})]$
- The individual stimulus classes also will get individual *F*- (if `-fout` added) and/or *t*-statistics indicating the significance of their individual *incremental* contributions to the data time series fit
  - If DF=1 (e.g., F for a single regressor), ***t*** is equivalent to *F*: $t(n) = F^2(1, n)$

Results of **rall_regress** Script



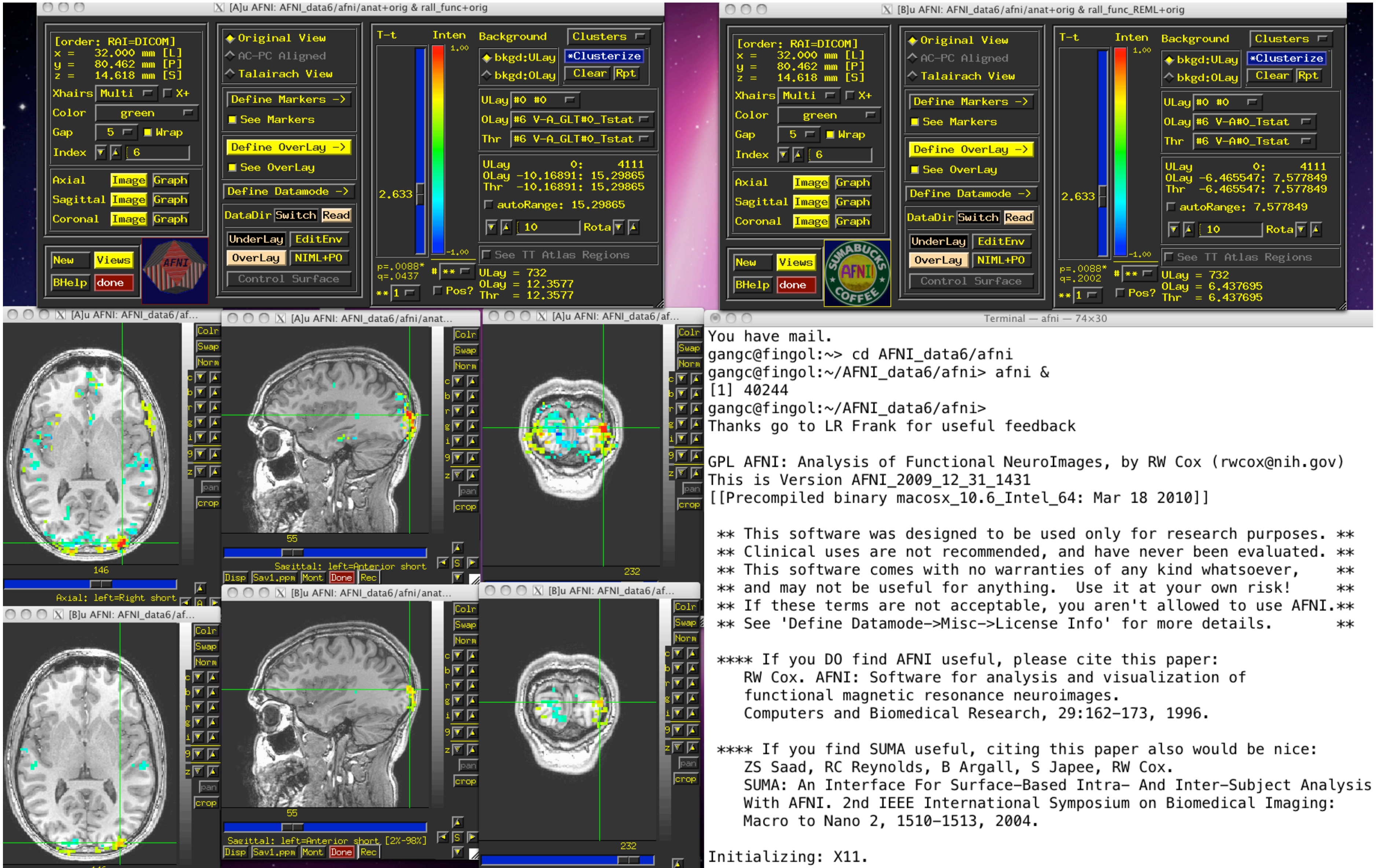






- Images showing results from third GLT contrast: **VrelvsArel**
- Menu showing labels from **3dDeconvolve**
- Play with these results yourself*!*

Compare 3dDeconvolve and 3dREMLfit



Group Analysis: will be carried out on β or `GLT` coef (+*t*-value) from single-subject analysis