selxavg

Comments or questions: analysis-bugs@nmr.mgh.harvard.edu\\
$Id: selxavg.tex,v 1.1 2005/05/04 17:00:49 greve Exp $

1 Introduction

selxavg2 is a program for performing selective averaging and deconvolution of event-related fMRI images for a single subject over multiple runs. It can also average data from blocked designs. Requires matlab 5.2 or higher.

2 Usage

Typing selxavg2 at the command-line without any options will give the following message:

```
USAGE: selxavg2 [-options] -i instem1 -p parfile 1 [-i instem2 -p parfile2 ...] -o outstem
   instem1   - prefix of input file for 1st run
   parfile1  - path of parfile for 1st run
   [-tpexclfile file1] - file with timepoints of 1st run to exclude
   [instem2] - prefix of input file for 2nd run
   [parfile2]- path of parfile for 2st run
   [-tpexclfile file2] - file with timepoints of 2nd run to exclude
   outstem   - prefix of .bfloat output files
      NOTE: you must specify all runs to process now
Options:
   -TR <float>           : temporal scanning resolution in seconds
   -TER <float>          : temporal estimation resolution <TR>
   -timewindow <float>   : length of hemodynamic response in seconds <20>
   -prestim <float>      : part of timewindow before stimulus onset (sec)
   -nobaseline           : do not remove baseline offset
   -detrend              : remove baseline offset and temporal trend
   -rescale <float>      : rescale target (must run inorm first)
   -nskip <int>          : skip the first n data points in all runs
   -hanrad <float>       : radius of Hanning spatial filter (must be >= 1)
   -fwhm  <float>        : full-width/half-max of in-plane smoother (mm)
   -ipr   <float>        : in-plane resolution (mm)
   -gammafit delta tau   : fit HDR amplitude to a gamma function
   -timeoffset <float>   : offset added to time in par-file in seconds <0>
   -nullcondid <int>     : Number given to the null condition <0>
   -firstslice <int>     : first slice to process <auto>
   -nslices <int>        : number of slices to process <auto>
   -percent pscstem      : compute and save percent signal change in pscstem
   -taumax  float        : don't assume white noise (spec max delay)
   -ecovmtx  stem        : compute and save residual error covariance
   -ecovmsn  stem        : residual error covariance, with spat norm
   -whtmtx   bfile       : whitening matrix
   -basegment            : force segmentation brain and air (with nonwhite)
   -mail                 : send mail to user when finished
   -eresdir dir          : directory in which to save eres vols
   -signaldir  dir       : directory in which to save signal vols
```

```
-monly mfile          : don't run, just generate a matlab file
-synth seed           : synthesize data with white noise (seed = -1 for auto)
-parname name         : use parname in each input directory for paradigm file
-cfg file             : specify a configuration file
-umask umask          : set unix file permission mask
-version              : print version and exit
```

3 Command-line Arguments

Note that command-line arguments can also be specified inside of a configuration file (see the -cfg argument).
-i instem: stem of the input functional volume for a single run. It is assumed that the data are stored in *bfile format*. If multiple runs are to be analyzed, each input stem must be preceded by a *-i* flag.

-p parfile: name of file in which the stimulus sequence is stored (in *paradigm file format*). The ID for fixation (or null-stimulus) defaults to 0 but can be specified using the *-nullconid*. The ID's must be consecutive integers (no skipping). If multiple runs are to be analyzed, a paradigm file must be specified for each run. A paradigm file will be paired with an input stem based on order of appearance on the command-line.

-tpexclfile: (optional) full path of file that specifies the time-points to ignore during an analysis. The format is a single columns of times within the run to exclude. One may want to exclude data points where the scanner has spiked.

-o outstem: this is the stem of the output volume (in bfile format) to be generated. The output consists of two volumes, one with stem *oustem* in which the selective averages and standard deviations are stored. The other has a stem *oustem-offset* in which the offset (or baseline) of each voxel is stored. In addition, there is a file called *outstem.dat* in which various parameters of the selective averaging are stored.
-TR float: temporal sampling resolution (ie, time between scans in seconds).

-TER float: temporal resolution (in seconds) at which to estimate the hemodynamic response. Defaults to the TR. Note that using anything besides the TR for the TER requires that the stimulus presentation scheduling to have taken a finer TER into account. This cannot be done post-hoc. See **optseq**.

-timewindow float: the amount of time needed to capture the hemodynamic response as if a stimulus were presented in isolation. This must include the prestimulus baseline period (see prestim).

-prestim tPreStim: specify a portion of *timewindow* to estimate the hemodynamic response *before* the stimulus is presented. This can be used as a "sanity check" to verify that the subject is not anticipating the stimulus. It can also be used to remove baseline shifts when computing statistics.

-nobaseline: instruct selxavg NOT to remove the baseline offset. By default, it will remove the offset on a run-by-run basis from the raw fMRI data. Note that the offset will be stored in *outstem-offset*.

-detrend: instruct selxavg to fit and remove any linear temporal trend from the raw fMRI data. This is done on a run-by-run basis. Specifying this option forces the **-baseline** option to be specified.

-rescale target: this option used to be obsolete but is now viable again when used in conjunction with **inorm**. **inorm**, the global intensity normalization program, will produce a file called *instem.meanval* in which the global mean value (after segmentation) is stored. When the -rescale option is specified, selxavg will read the meanval file and the target value and rescale the entire input volume so that it's new global mean is target.

-nskip nSkip: specify the number of *functional* slices to skip at the beginning of each run. This applies to all runs and all anatomical slices. This functionality can also be implemented with a tpexclude file.

-fwhm FWHM: specify the radius of a 2-D Hanning kernel with which to spatially filter each functional image. The radius is in mm. This is identical to -hanrad except that in -hanrad the radius is measured in in-plane voxels.

-hanrad HanRadius: specify the radius of a 2-D Hanning kernel with which to spatially filter each functional image. The larger the Hanning radius, the more smoothing. Radii less than 1.0 have no effect.

-gammafit Δ τ: fit the amplitude of a gamma function to each hemodynamic response. The output file will then hold the best-fit amplitude instead of the hemodynamic response. One can still run *stxgrinder* on the output, but the *-ircorr* and *delrange* options cannot be used nor can the *tm* or *Fm* tests be specified. Typical values: $\Delta = 2.25s$, $\tau = 1.25s$. This option is needed for processing data blocked designs. The gamma function has the form:

$$h(t) = \begin{cases} 0 & t < \Delta \\ (\frac{\tau e^2}{4})(\frac{t-\Delta}{\tau})^2 e^{-(\frac{t-\Delta}{\tau})} & t > \Delta \end{cases} \quad (1)$$

-timeoffset float: time (in seconds) to add to times in parfile for cases when the acquisition and presentation sequences did not start at the same time. Positive offsets indicate that the stimulus presentations began *after* the first scan.

-firstslice int: first *anatomical* slice to process (usually 0). If unspecified, **selxavg** will autodetect the first slice. It is highly recommended that the first slice number be 0.

-nslices int: total number of *anatomical* slices to process. If unspecified, **selxavg** will autodetect the number of slices.

-percent pscstem: instruct selxavg to compute and save the hemodynamic response averages as a percentage of the baseline. The results are stored in a separate volume with stem *pscstem*. The variances are scaled accordingly. Note that this is somewhat redundant because the offsets are automatically stored in *outstem-offset*, and the percent signal change can be computed from the offsets and the hemodynamic averages when viewing (eg, in **yakview**).

-taumax maxdelay: *note: this feature is still a little buggy, so use for evaluation only.* Specifies that selxavg should compensate for temporal correlations in the noise by estimating and fitting the temporal noise autocorrelation. The argument maxdelay is the maximum delay (in

seconds) that should be used when fitting the noise autocorrelation function. It typically ranges between 20-30 sec.

-basegment: this flag forces selxavg to segment the brain from the air in each functional run and estimate the noise autocorrelation function only from those voxels labeled as brain. It is highly recommended that this flag be set.

-eresdir: specify a directory in which to save the residual error for each slice and run. Not very useful for most users.

-signaldir: specify a directory in which to save the estimated signal for each slice and run. Not very useful for most users.

-synth seed: synthesize input data as white gaussian noise. If the seed is set to -1, then sexavg will generate a seed based on the system clock. This option is used for debugging.

-monly: only generate the matlab file which would accomplish the analysis but do not actually execute it. This is mainly good for debugging purposes.

-mail: mail user upon completion of analysis.

-cfg file: specify a selxavg configuration file. This is a file with a different command-line option (and its arguments) on each line. For example,

```
-TR 2
-hanrad 1.5
-detrend
-gammafit 2.7 1.9
```

This is often more convenient typing all the options on the command-line. Any command-line options specified with -cfg will override any conflicting values in the configuration file.

4 Output

selxavg will create several output files per anatomical slice plus one file for all slices. The names of these files are based on the output stem provided on the command-line. The hemodynamic averages are stored in *outstem_sss.bfloat* which has the corresponding header file *outstem_sss.hdr* (where *sss* indicates the anatomical slice number). A file called *outstem-offset_sss.bfloat* is created in which the offset (or baseline) is stored. A file called *outstem_.dat* is created into which various parameters are stored (mostly command-line parameters).

The waveforms within the hemodynamic average file are stored in a certain format (same as *selavg* format). Each HDR estimate for each condition will have N_h components, where N_h is the integer ratio of the timewindow and the TR. In a *bfloat* file, the data are stored as image planes with each plane having a number of rows and columns equal to the those of the original functional scan. Each condition will occupy $2N_h$ planes for a total of $2N_hN_c$ planes where N_c is the total number of conditions including fixation. For a given condition, the first N_h planes are the averages for each point in the timewindow. The next N_h planes are the standard deviations for each corresponding point in the timewindow. With *selxavg*, the averages for the fixation condition are *always* set to zero; they are included in the file strictly for backwards compatibility with *selavg* format. The standard deviation planes for the fixation condition are all equal to the

standard deviation of the residual error and are used in further processes steps (eg, *stxgrinder*). Again, this redundant information is included for backwards compatibility.

5 Examples

Consider the case in which the current directory is */space/raid/4/users/inverse/fspace/950818SA*, the base directory of a functional session. Below this directory is, among others, a subdirectory called *image*. Below the *image* subdirectory are five subdirectories: *r01, r02, r03, r04, r05*, one for each of five runs. The functional data are stored in the run subdirectories. For example, the functional data for anatomical slice 10 of run 2 would be in a file called *s_010.bshort* under the *r02* directory. There would be a header file companion called *s_010.hdr*. Assume that there are 16 anatomical slices, 0 through 15, and that the TR is 2 seconds. Assume that the paradigm files for the five runs are stored in a subdirectory of *image* called *parfiles* and that the five paradigm files are named *par01.dat, par02.dat, par03.dat, par04.dat, par05.dat*. Assume also that there are three types of stimuli presented in the experiment, one of which is fixation.

To selectively average all the slices for the first run only, one would execute the following command line:

```
selxavg -TR 2 -timewindow 24 -prestim 4 -detrend \
        -i image/r01/s -p parfiles/par01.dat \
        -o sxa1/havg
```

The *-TR 2* option tells selxavg to use a TR of 2 seconds. The *-timewindow 24* and *-prestim 4* indicate that the hemodynamic response will be estimated 24 seconds, 4 seconds of which will be before stimulus onset. The *-detrend* option forces the baseline and temporal trends to be removed. The input stem is *image/r01/s* and the paradigm file is *parfiles/par01.dat*. The output stem is *sxa1/havg*. After successfully running selxavg with these options, a new subdirectory will have appeared under the *image* directory called *sxa1*. Within *sxa1*, there should be 16 files of the form *havg_sss.bfloat* (along with 16 header (.hdr), as well as 16 files of the form *havg-offset_sss.bfloat* (along with 16 header (.hdr), and one file called havg.dat.

To average all five runs, execute:

```
selxavg -TR 2 -timewindow 20 -detrend \
        -i image/r01/s -p parfiles/par01.dat \
        -i image/r02/s -p parfiles/par02.dat \
        -i image/r03/s -p parfiles/par03.dat \
        -i image/r04/s -p parfiles/par04.dat \
        -i image/r05/s -p parfiles/par05.dat \
        -o sxa1/havg
```

Note that it is *not* possible to execute *selxavg* five different times to average all the runs. It must be executed once with all five runs specified at the time of execution.

6 Theory of Operation

6.1 Forward Model

$$y = X_1 * h_1 + X_2 * h_2 + b + at \ldots, \tag{2}$$

where y is the $N_t \times N_v$ matrix of raw data, X_i is the design matrix (ie, presentation shedule) for stimulus type i, h_i is the (FIR) impulse response to stimulus type i, b is the baseline offset, a is

the slope of the linear drift, and t is time. This can all be collapsed into one matrix equation:

$$y = X * h \tag{3}$$

where y is still the raw data, X contains the presentation schedules for all stimuli plus a column of ones for the baseline offset plus a column of linearly increasing numbers for the temporal trend, h contains the (FIR) impulse response to all stimulus types as well as the baseline offset and linear slope.

6.2 Inverse Model

In the equation above, we have X and y and want to find h. This is accomplished by solving the inverse model:

$$\hat{h} = (X^T X)^{-1} X^T y \tag{4}$$

6.3 Construction of X

For a given stimulus type, the number of rows of X_i equal the number of time points N_t, and the number of cols of X_i equal the number estimates in the hemodynamic response N_h. N_h, in turn, is equal to the time window (as specified with the -timewindow option) divided by the TER (as specified with the -TER option). The timewindow should be an intetger multiple of the TER. If the TER is unspecified, it defaults to the TR. If the prestimulus window is zero, X_i is constructed in the following way. When a stimulus of type i is presented, a 1 is placed in the first column of X_i at the row corresponding to the time point at which it was presnted. Another 1 is then placed in the second column at the next row (ie, diagonally down and to the right). This process is repeated to fill all the N_h columns.