==========

README: SPIKECOR script to correct motion spikes in fMRI data (2014/04/10)

==========

Author: Nathan Churchill, University of Toronto
email: nchurchill@research.baycrest.org

==========

This is a rapid automated algorithm, used to detect large outliers ("spikes") in fMRI timeseries, typically due to abrupt head movements. Spikes are estimated using Principal Component Analysis (PCA) on a sliding time-window. It is applied to fMRI data and/or motion parameter estimates (MPEs) that are produced under standard rigid-body motion correction.

This algorithm fits a maximum-likelihood Gamma distribution on motion variance and identifies significant outliers at p<.05. It is a purely statistical criterion, and does not require any manual thresholding on head motion. The algorithm and model details provided in:

> *Campbell K, Grigg O, Saverino C, Churchill N, Grady C. Age Differences in the Intrinsic Functional Connectivity of Default Network Subsystems. Frontiers in Human Neuroscience. 2013; 5:73*

Please cite this article if code is used in any publications.

README CONTENTS:

I. ALGORITHM AND SCRIPT OVERVIEW
II. GENERAL REQUIREMENTS
III. RUNNING SPIKECOR
IV. INTERPRETING DIAGNOSTIC PLOTS

I. ALGORITHM AND SCRIPT OVERVIEW

The algorithm is uses a single script, "spikecor.m". It also requires Jimmy Shen's tools for loading nifti files: (www.mathworks.com/matlabcentral/fileexchange/8797-tools-for-nifti-and-analyze-image). A "NIFTI_tools" folder is included, just make sure it is in your path if needed.

The outlier testing procedure (also available by typing "help spikecor"): we want to find spikes (brief, large signal changes) in multidimensional data. In order to confirm that they are motion-related artifact, we estimate outliers in both MPE and fMRI data.

1. For each datapoint $\mathbf{x}(t)$ ($1 <= t <= N_{time}$):
   (a) find median PC-space coordinate $\boldsymbol{x}_{med}(t)$, in a 15 TR timewindow centered at t
   (b) compute displacement by the Euclidean distance $D(t) = || \boldsymbol{x}(t) - \boldsymbol{x}_{med}(t) ||_2$
2. For displacement timeseries D, compute Gamma distribution on D^2
   using Maximum Likelihood fit to infer degrees of freedom.
3. Identify outliers, significant at p<0.05

- We measure displacement relative to a 15-TR time window:
  This has the advantage over (1) a simple derivative, e.g. we can discriminate
  displacement AWAY from the main datapoint cluster vs. displacement back TOWARDS
  this cluster. The windowing also (2) minimizes the impact of slow changes
  in amplitude over time, which might inflate our displacement estimates.
- If we do slice-specific outlier testing, this whole process is repeated independently
  for each axial brain slice

- For outlier removal:
  We identify outlier volumes/slices, based on the chosen statistical
  criterion (specified in OUT_PARAM). Outlier volumes/slices are discarded
  and replaced using cubic-spline interpolation from neighbouring volumes.
  Unlike simply discarding motion spikes, this avoids any issues of
  sharp discontinuities in the timeseries, and is less sensitive to data
  and power loss from motion spikes

- the script also produces some extra diagnostic outputs (e.g. PCA eigenspectra)

II. GENERAL REQUIREMENTS

* fMRI dataset
  - 4D fMRI timeseries; (in NIFTI or ANALYZE format)
* Brain Mask
  - 3D binary volume where 1=brain voxels (NIFTI or ANALYZE format)
* File containing Motion Parameter Estimates
  - these are the 6 rigid-body motion parameters, estimated when
    performing standard motion correction (e.g. AFNI's 3dvolreg; FSL's flirt)
  - formatted as text-file containing (time x 6) column matrix,
    e.g. line t contains [roll pitch yaw, IS RL AP] coordinates of timepoint t.

III. RUNNING SPIKECOR

- Inputs and Outputs of "spikecor.m" given below; this information is in the spikecor.m header as well.
- A “pseudo-code” example below demonstrates how to run the script

```
% ========================================================================
%  SPIKECOR
% ========================================================================
%
% SYNTAX:
%
%       spikecor( volname, maskname, mpename, outprefix, OUT_CORRECT, interpname  )
%
% INPUT:
%       volname   = string, giving path/name of input fMRI data
%                   must be 4D fMRI data in NIFTI or ANALYZE format
%       maskname  = string, giving path/name of binary brain mask (to exclude non-brain tissue)
%                   must be 3D fMRI volume in NIFTI to ANALYZE format
%       mpename   = string, giving path/name of MPE file
%                   must be a 6-column textfile, giving 6 rigid-body parameter timerseries
%
%      outprefix  = optional string giving output path for (1) QC output, and (2) diagnostic figures.
%                   If outprefix=[], uses 'volname' as the default
%
%      OUT_PARAM  = string determines which criteria are used to identify outliers
%                   and interpolate over them. OUT_PARAM options:
%
%                   'none'        : do not discard outliers  (for diagnostic purposes only)
%                   'motion'      : replace outlier volumes, based on MPE values
%                   'volume'      : replace outlier volumes, based on fMRI PCA distribution
%                   'volume+motion': replace outlier volumes, based on MPEs & fMRI PCA distribution
%                   'slice'       : replace outlier slices,  based on fMRI single-slice PCA distribution
%                   'slice+motion' : replace outlier slices,  based on MPEs & fMRI single-slice PCA distribution
%
%                   * we recommend conservative choice 'volume+motion' as a starting point.
%
%      interpname = string specifying the full path+name of the de-spiked fMRI data output.
%                   (e.g. interpname = 'mypath/subject_data_1_interp.nii')
%
% OUTPUT:
%
%  (1) Interpolated fMRI data, labeled [ interpname, '.nii'],
%     provided OUT_PARAM != 'none'
%
%
%  (2) an "output" structure, saved to [outprefix,'_QC_output.mat'] with fields:
%
%     [Censor vectors] Binary vectors of length (time x 1), where (1=non-outliers) and (0=outliers).
%       These are used to remove and interpolate outliers if OUT_PARAM != 'none':
%
%       output.censor_mot   : significant outlier in MPEs
%       output.censor_vol   : significant outlier in fMRI data
%       output.censor_volmot: significant outlier in BOTH fMRI data and MPEs**
```

```
%
%     [Censor matrices] A new feature! Binary matrices of size (time x brain slices)
%       where (1=non-outliers) and (0=outliers). This gives outliers for individual
%       axial brain slices, and can be used as input to remove outlier slices:
%
%       output.censor_slc   : significant outlier in fMRI slice data
%       output.censor_slcmot: significant outlier in BOTH fMRI slice data and MPEs**
%
%     ** for (censor_volmot) and (censor_slcmot), we discard fMRI outliers
%       if they occur at same time OR +1 TR after a motion spike. This
%       allows for delayed fMRI-related signal changes (e.g. spin-history effects)
%
%     [Other possibly useful outputs]
%
%      output.eigimages_fmri: matrix (voxel x K) PCA eigenimages for fMRI data (K=PCs explain. 95% of variance)
%      output.eigvect_fmri  : matrix (time x K) PCA eigen-timeseries for fMRI data
%      output.eigfract_fmri : vector (K x 1) of fraction of variance explained by each PC of fMRI data
%
%      output.eigweights_mot: matrix (6    x 6) PCA weights on motion parameters for MPEs
%      output.eigvect_mot   : matrix (time x 6) PCA eigen-timeseries for MPEs
%      output.eigfract_mot  : vector (6 x 1) of fraction of variance explained by each PC of MPEs
%
%  (3) Quality Control output figures. The figures include:
%
%      "<outprefix>_diagnostic_plot0.png" : summary results for temporal variance in data
%      "<outprefix>_diagnostic_plot1.png" : summary results from PCA decomposition of data
%      "<outprefix>_diagnostic_plot2.png" : results for estimated motion spikes
```

PSEUDO-CODE EXAMPLE

If you have the following:

> motion-corrected 4D fMRI data "subjectX_data.nii", located in "myfiles/"
> textfile containing MPEs "subjectX_mpe.txt", located in "myfiles/"
> 3D mask of brain tissue "subjectX_mask.nii", located in "myfiles/"

The spikecor function is called as:

```
spikecor( 'myfiles/subjectX_data.nii', 'myfiles/subjecX_mask.nii', 'myfiles/subjectX_mpe.txt',
'myfiles/subjectX_summary', 'volume+motion', 'myfiles/subjectX_despike.nii' );
```

Which produces the outputs:

> "subjectX_despike.nii": de-spiked version of "subjectX_data.nii",
  located in directory "myfiles".
  NB the 'volume+motion' parameter requires that removed spikes are outliers

in both fMRI data and MPEs

> "subjectX_summary_QC_output.mat": a matfile containing an "output" structure, with summary information on outliers in the data

> "subjectX_summary_diagnostic_plot(0,1,2).png": three diagnostic figures showing some information about the data -- see Section IV. INTERPRETING DIAGNOSTIC PLOTS below for more information

---

IV. INTERPRETING DIAGNOSTIC PLOTS

The spikecor algorithm also produces a set of general diagnostic output figures that might be useful for interpreting the signal/noise content of your fMRI data. They are as follows:

"<outprefix>_diagnostic_plot0.png" : summary results for temporal variance in data

(top) 3 images are colormaps of temporal Standard Deviation (SD) measured at each voxel in the fMRI timeseries data. The saggital, coronal and axial slices are plotted, that intersect at the center of the brain volume. Colormap ranges from [2.5th to 97.5th] percentile of SD in the brain (e.g. the 95% confidence interval)
>> Used to identify locations of greatest variance in the brain (e.g. is it in brain edges (motion) or vasculature (physiology)?

(bottom) Z-scored timeseries identified in the data:
. fMRI-PC: 1st principal component timeseries of fMRI data
. fMRI-GS: "Global Signal" (mean BOLD signal of each brain volume)
. MPE-PC: 1st principal component timeseries of rigid-body MPEs
We also plot R^2 (coefficient of determination) between fMRI-PC and fMRI-GS / MPE-PC. This measures shared variance between timeseries.
>> Tests whether fMRI signal (fMRI-PC) is highly correlated with MPEs.

"<outprefix>_diagnostic_plot1.png" : summary results from PCA decomposition of data

(top, left): scatterplot of PC#1 vs. PC#2 coordinates (e.g. loadings) for each datapoint in the fMRI timeseries (blue dots). The red curve plots the PC-space trajectory over time from T=1 to T=(#timepoints); this trajectory is computed as the mean PC-space coordinates in a 15-TR sliding window.
>> Demonstrates multivariate variance structure in fMRI data

(top, right): eigenspectrum for both fMRI and MPE principal component decompositions, measured as fraction of total variance.
>> Tests the complexity of fMRI signal and MPEs; a "flatter" spectrum indicates more complex signal, that requires many PCs to describe

(bottom): brain PCA eigenimages, associated with PC#1 and PC#2, plotted in (top,left). Colormap scaled to +- (95th percentile) of voxel weights.
>> shows the spatial patterns with greatest modulation in fMRI data

"<outprefix>_diagnostic_plot2.png" : results for estimated motion spikes

(top, left): displacement of each time-point in MPE data, relative to median PC-space coordinates (rescaled so that maximum=1). Dashed red line gives the p=.05 significance threshold for outliers

(middle, left): displacement of each time-point in fMRI data, relative to median PC-space coordinates (rescaled so that maximum=1). Dashed red line gives the p=.05 significance threshold for outliers

(bottom, left): colormap of outlier timepoints. Blue = not an outlier. Yellow=outlier in fMRI or MPEs. Red = outlier in BOTH fMRI and MPE data.

(top,right): heatmap showing displacement of each time-point for axial fMRI slices, relative to median PC-space coordinates
(slice x time matrix; values rescaled so that maximum=1).
In colormap, blue=minimum displacement / red=maximum displacement

(bottom, right): colormap of outlier timepoints per axial slice.
Blue = not an outlier. Yellow=outlier in fMRI or MPEs.
Red = outlier in BOTH fMRI and MPE data.

* right-side plots (slice x time outliers) can be used to test whether there is within-volume (non-rigid) head motion.
For example, if you see "columns" of outliers, you have full-volume displacement (e.g. every slice is an outlier at the same time)
If you see horizontal "bands" then you have potential motion within interleaved volume acquisition, and might want to consider OUT_PARAM='slice' or 'slice+motion' settings.