

IMAGEN

Processings FAQ

Revision 01/13/11

Content

1 Transfer and conversion

- 1.1 Which converter is used to obtain nifti files ?
- 1.2 Why does the images filename includes the sequence number ?

2 Structural Image processings

- 2.1 Which algorithm is used for the structural image spatial normalisation ?
- 2.2 What does the file-names prefix in the database means ?
- 2.3 Why not using a custom template for anatomical spatial-normalisation ?

3 EPI sequences processings

- 3.1 Which preprocessings are applied to EPI sequences ?
- 3.2 Which reference-slice is used as a reference during slice-timing-correction ?
- 3.3 Which warping algorithm is used to spatially normalise fMRI sequences ?
- 3.4 How was the custom EPI template exactly computed ?
- 3.5 Why use an fully EPI-based spatial normalisation instead of an anatomically-guided one ?

4 Subject-level Modeling

- 4.1 Which movement confounds are included in the design matrix ?
- 4.2 I want to test the opposite effect of some contrast map, but I don't know how to threshold negatively. How to quickly reverse the sign of wimages ?

5 Artefacts detection, Recovering

- 5.1 Were spike artefacts corrected ?
- 5.2 How and when is Outliers detection for fMRI contrast maps used ?
- 5.3 How were the contrast-specific Region-Of-Interest (ROI) created ?
- 5.4 How does Artefactual timepoints due to incorrect realignment, work ?
- 5.5 How do you make use of the incorrectly realigned volumes list ?
- 5.6 Can I have a short paragraph summarizing all QC-related auto processings ?

1 Transfer and conversion

1.1 Which converter is used to obtain nifti files ?

5

The **dcm2nii** converter included in Chris Rorden's MRICron package is applied on all the DICOM available directories. This is the result of trying various converter, none of which had a 100% success rate. The dcm2nii had a reasonably low failure rate. Manual recovering of failed cases is yet to be done.

1.2 Why does the images filename includes the sequence number ?

This numbering scheme is the one from the dcm2nii converter and i tend not to modify outputs more than necessary as it avoids potential incompatibilities later. Also, it sometimes avoided filename clashes. This does apply to original, converted nifti images and some of the preprocessed EPI files. 10

Note that if you use commandline (bash), you can rename all your files, dropping the s* suffix, quite trivially with a command similar to:

```
for f in processed/spmpreproc/*/SessionA/EPI_faces/wea*.nii ; do mv $f ${f%*s*}.nii ; done 15
```

2 Structural Image processings

2.1 Which algorithm is used for the structural image spatial normalisation ?

The SPM8 “New Segmentation” algorithm is applied to all MPRAGE anatomical image. This algorithm fits & wraps six tissue classes using a bayesian framework to yield probabilistic segmented tissues maps and MNI-space

warping parameters.

20

The output of the “New Segmentation” algorithm provides both Segmentation and Spatial-normalisation informations, in the form of tissues probability maps in MNI-space and in native space. Additionally, deformed tissues may be “modulated”, ie. scaled by the inverse-Jacobian of the local transformations, to maintain matter volume.

The default parameters of SPM8 “New Segmentation” are unmodified. When applying the transformation, the new images are created using a 4th Degree B-Spline interpolation.

Note that the tissues maps had not been smoothed, leaving users free to decide any optional smoothing parameters.

2.2 What does the file-names prefix in the database means ?

A few of the SPM-preprocessed structural images are available for every subjects, using a SPM-like naming scheme:

mmprage* (note the 'm' prefix) : is the **intensity-bias corrected** image

wc_1*, wc_2, wc_3* : those images are segmented tissues probability maps, **warped in the MNI** coordinate space. (1, 2, 3 code respectively for Gray Matter, White Matter, Corticospinal Fluid). Since it's probability maps, the range of values is 0 to 1. The resolution of it is 1mm x 1mm x 1mm (also, the dimensions voxel-wise match exactly the MNI152_T1_1mm Template file). Note that those have **not been smoothed**.

mwc_1*, mwc_2 : those images are “**modulated**” tissues maps, ie. segmented tissues maps warped to the MNI coordinate space, appropriately scaled per-voxel in order to compensate for deformation regarding matter volume. The resolution in those is 1.5mm x 1.5mm x 1.5mm. (1 and 2 code respectively for Gray Matter and White Matter). Those images have **not been smoothed**.

c_1*, c_2*, c_3* : those images are segmented tissues in **native** space.

2.3 Why not using a custom template for anatomical spatial-normalisation ?

For the T1-processings, specifically the spatial warping, we eventually choose to use the standard MNI space. This choice had been made after we got convinced of the superior robustness of the normalization from SPM8's bayesian warping algorithm, compared to the classical intensity-based template registration .

The warped anatomical data available from Imagen are therefore normalized in the same coordinate systems of the MNI.

Using a population-specific template is an attractive idea, but it is still not clear that it could have been efficiently combined with the new segmentation which requires a priori tissue maps.

3 EPI sequences processings

3.1 Which preprocessings are applied to EPI sequences ?

Time series (potentially spike-corrected) are **slice-timing corrected**, temporally **realigned**, warped on the **MNI space**, **gaussian-smoothed at 5mm-FWHM**. To do that, the SPM8 software is used.

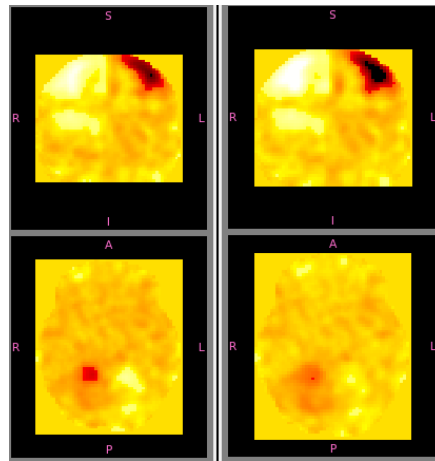
Default parameters of SPM8 are applied unless otherwise specified.

The output time-series has a 3mm x 3mm x 3mm voxel size; interpolation used 4th Degree B-Spline.

3.2 Which reference-slice is used as a reference during slice-timing-correction ?

The slice-timing correction of top-down scanning uses the upper (first) slice as reference for the interpolation.

On some contrasts, e.g. motor, using an inverted slice order (and opposite reference slice accordingly) may seems to give better results on statistical maps. However, this may be due to the inaccuracy at modeling the Hemodynamic response in this particular region, as this counter-intuitive behaviour is not replicated in other area.



The figure below shows the behaviour : Left fig assumes top-down slice order, Right fig assumes bottom-up slice order, with first slices for reference, on 8 subjects, same color scale ; The motor area signal (upper fig) is stronger on the bottom-up slice ordering, but the cerebellum activations (lower fig) is stronger on the correct slice ordering.

3.3 Which warping algorithm is used to spatially normalise fMRI sequences ?

The generic non-linear warping algorithm included in SPM8 is used to register a mean image of the sequence to a custom specific EPI template. 65

3.4 How was the custom EPI template exactly computed ?

A set of 240 randomly-selected subjects (30 for each acquisition center) EPI sequences for two protocols (SST and MID, ie. one from each session) were spatially-realigned, and their temporal-mean image were rigidly coregistered to their respective anatomical image. Those anatomical images were warped to the MNI space using the SPM8 non-linear normalisation algorithm, and the corresponding transformations parameters were applied on the mean-EPI images. From those 480 MNI-warped mean EPI images, a simple xor-based metric is computed to quantify the overlap quality between individual EPI masks and the MNI mask. Using this metric we discarded 17% outliers, and the remaining 400 mean-images are averaged, then smoothed with a 5mm FWHM gaussian kernel. 75

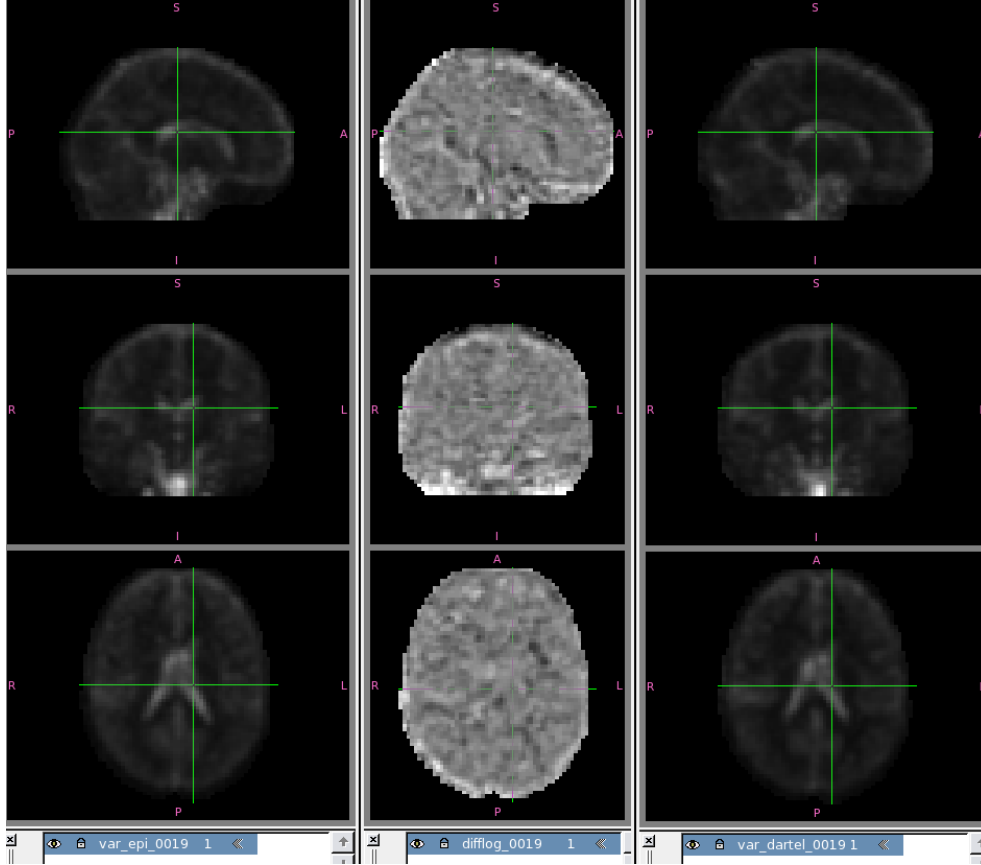
The final template images is 53x63x46 voxels, with 3mm x 3mm x 3mm voxel resolution. It is available at [XXXfixme:urlXXX](http://www.spm8.com/fixme:urlXXX)

3.5 Why use an fully EPI-based spatial normalisation instead of an anatomically-guided one ?

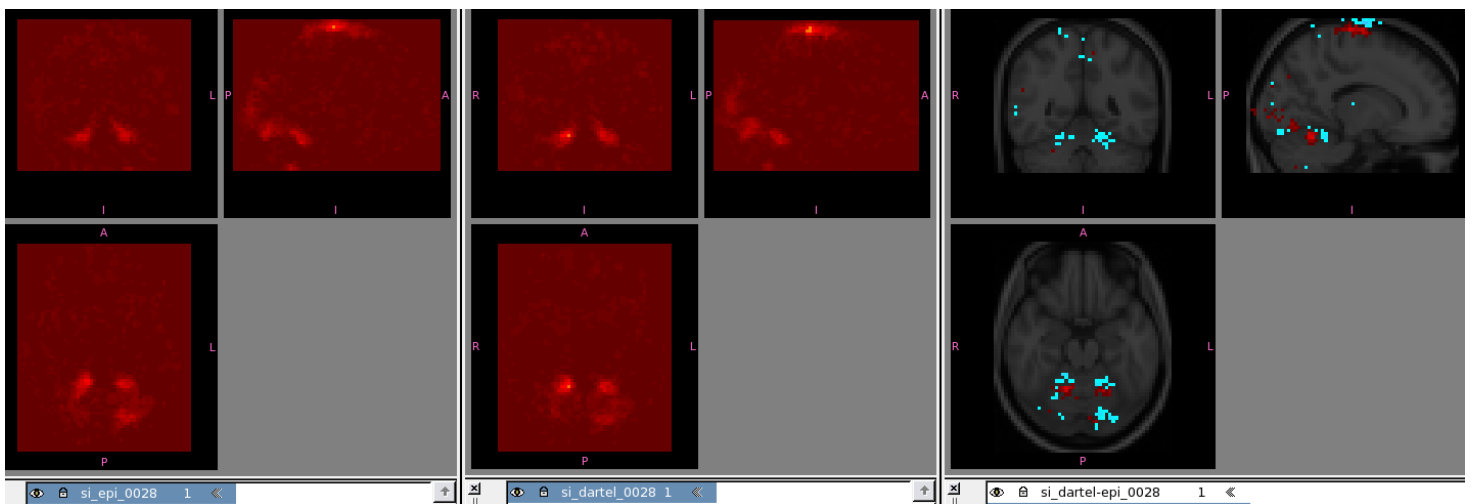
On the tests we conducted on ~80 subjects, for the EPI_short_MID task, we couldn't find significant influence of the warping algorithm on the shape or values of the output parametric maps. 80

The influence of spatial normalisation on the statistical maps has been experimented on a subset of subjects:

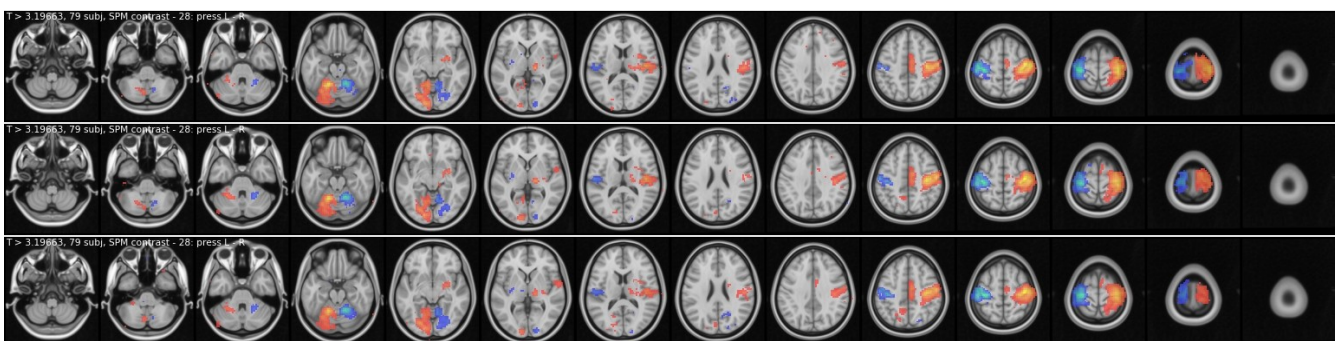
- EPI_short_MID, contrast con_0019 ("feedback_hit-missed"), variance image of 71 subjects contrast maps. Left : EPI-template based, Right : Dartel-warping based, Middle: log difference ($\log(\text{EPI_based}) - \log(\text{Dartel_based})$).



- Contrast con_0028 (press Left – Right) “Consensus”-image (sum of individual thresholded T-maps). 85
 Left : EPI-based warping, Middle : Dartel-based warping. 71 subjects, Individual threshold at $|T| > 3$, maxima value are 46 (EPI, left) and 52 (Dartel, middle). Right : main difference of the two. (Red : $dartel > epi$; Cyan : $epl > dartel$; thresholded on difference > 2). Those images are downloadable at XXXfixme:urlXXX



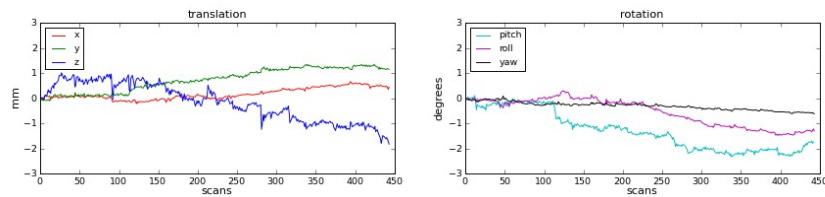
- EPI_short_MID, 79 Subjects, con_0028 (press Left – Right). Top to bottom : Directly Warped on the 90
 EPI custom template, warped applying Anatomical-based warping parameters, warped applying
 Dartel-computed tissues-based warping parameters.



4 Subject-level Modeling

4.1 Which movement confounds are included in the design matrix ?

Movement parameters are estimated during the preprocessing step, which outputs 6 time-dependent parameters (3 for translations and 3 for rotations), similar to the example below:



In all models version 1.0, those estimated values are inserted in the design-matrix as confound regressors. This include the 3 translation parameters, the 3 rotation parameters, and also the 3 squared translation parameters, and the 3 cubic-ed translation parameters (Those are intended to fit e.g. potential spikes). Additionally, two temporally shifted copies of the 3 translations parameters are included, corresponding to one TR forward and one TR backward.

4.2 I want to test the opposite effect of some contrast map, but I don't know how to threshold negatively. How to quickly reverse the sign of wimages ?

Separately, if you have the FSL software, i found that the commandlinetool "fslmaths" allows multiplying by -1.

Example usage is :

```
fslmaths /tmp/con_0028.nii -mul -1 /tmp/con_0028_reversed_fsl.nii
```

And, supposing you're using unix scripting, the command can beembedded in a loop in such kind of way:

```
cd /data/mid_contrasts/con_0010
for a in 000*/con_0010.nii.gz; do fslmaths $a -mul -1 ${a/con_0010/con_0010_reversed};
done
```

...but nevermind it if you're not familiar enough with those:-)

5 Artefacts detection, Recovering

5.1 Were spike artefacts corrected ?

Each individual fMRI timeseries undergo an automatic spike detection algorithm. It uses a mean-squared based metric to identify unexpected values temporally and spatially limited to a single slice.

At the preprocessing time, input timeseries are corrected for these spike using a temporal linear interpolation at the corresponding slices, before undergoing the next preprocessings. An exemple of such a detected slices is shown below (at two successive temporal points, left is axial, middle is sagittal, right is difference of the two timepoints) :



The algorithm detected and recovered bad slices in about 150 sequences, out of a total of about 4 * 800 sequences tested, across protocols.

5.2 How and when is Outliers detection for fMRI contrast maps used ?

Outliers detection is based on a region-of-interest-oriented distance metric. Subjects which departs heavily from a group average are expected to be either large outliers or boggus data, and are therefore suggested for a visual review.

The distance metric compute a multivariate, mahalanobis **distance** of the signal from **inside a ROI set to the rest of the group**. That group average (and covariance-matrix) is now fixed per protocol, per contrast, and has been

computed using the trimmed average from about 1000 subjects.

For a **given protocol and contrast number**, to the available distance score of each individual subjects is associated with **two suggested threshold**. One is based on the null Chi2 distribution probability at $p = 0.001$, and the other is based on a non-parametric estimation of the distribution. Note that the second threshold is consistently higher (less severe) than the first one. 130

This procedure is able to detect the most blatant **outliers regarding multivariate effects**, most of which includes obvious artefactual cause of problem.

For convenience, the ROI set used to compute the distance is based either on the HarvardOxford ROI **cortical** atlas, ROI **subcortical** atlas, or contrast-specific **hierarchical functional ROI** set (see later question), and the corresponding distance values are available for each of those. 135

Those distance measures and associated thresholds, or, alternatively, already-thresholded list of subjects are available using Data->Spm->Score outliers from the Imagen database website. The six columns are named respectively score_cortical, score_subcortical, score_multiroi (for scores values), and supra_cortical, supra_subcortical, supra_multiroi (for flags indicating whether it crossed one or two of my own precomputed thresholds). Keep in mind that this is an automatic procedure and should be considered mostly as hints for QC. 140

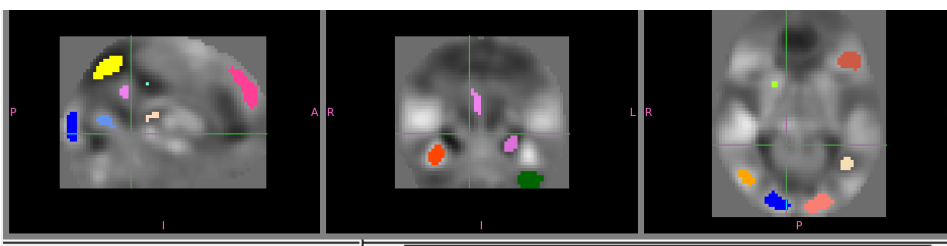
5.3 How were the contrast-specific Region-Of-Interest (ROI) created ?

In order to overcome the difficulty of manually digging individual activations maps (first level) from thousands of subjects coming from four EPI sequences, we sometimes make use of an automatic Functional ROI extraction algorithm, intended to locate the peak activities of effect maps. 145

For reference, the algorithm is outlined below:

- For one specific contrast, all individual significance maps of every subjects are averaged.
- A tree-based algorithm aims at summarizing this average activation landscape through the creation of hierarchical isolevels-like regions. 150
- Others averaged-significance maps are created from bootstrapped samples of the subject population. The same tree algorithm is run on each of those bootstrapped average maps.
- Computing an overlap of the regions, less reproducibles ones are discarded, while others are averaged, leading to a final set of hierarchical ROIs which are relevant sample-wide for the signal analysis of the contrast. Other sets of hierarchical ROIs can be created the same way for every contrast. 155

Example of such functional-ROIs label map. Because of its hierarchical nature, not all ROIs is displayed on the picture below.



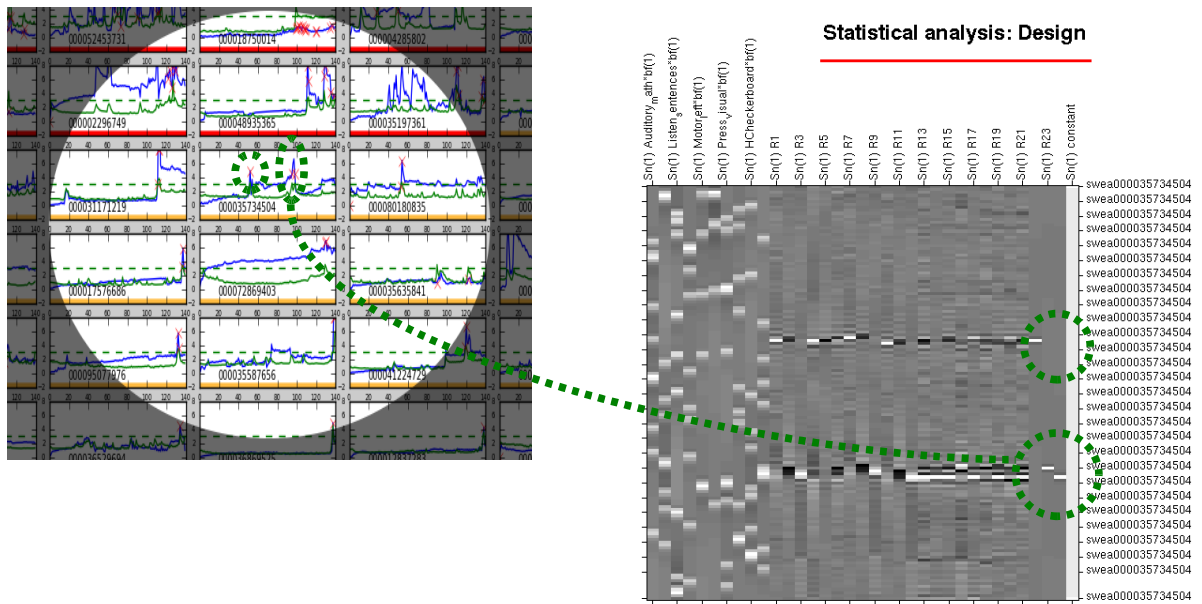
5.4 How does Artefactual timepoints due to incorrect realignment, work ?

The preprocessing part involves a realignment step which aims at spatially registering all volumes from the timeseries to a reference volume (the temporal-mean volume). This is achieved with the SPM8 software, which outputs both the realigned, resampled sequence, and estimates of movement parameters. Yet this procedure cannot recovers from movements which are either too wide or too artefactual ; in that case, it is probably best to simply ignore those volumes. The automatic detection of those volumes is based on the simple observation that a badly realigned volume depart significantly from the average volume of the sequence. Thus, a mean-squared based metric is flagging the timepoints where the volume **mean-square difference to the (robust-) average** one raise above a fixed threshold (This upper threshold is set somewhat empirically from the data distribution, so that only 0.5% of all the time-points of all of the (180-subjects available sub-)population have a larger value for the difference measure) 160

5.5 How do you make use of the incorrectly realigned volumes list ?

170

The bad volumes may be used, for e.g., as input in the design-matrix, in order to regress out those bad timepoints. This is done in all models version 2.



Additionally, the number of volumes flagged as bad across the sequence is also saved, and an informative text is set: “BAD” if more than one bad volumes per minute, “OK” if no such volume is found at all, “DUBIOUS” in between.

175

These informations are available in the IMAGEN database website, in Data->Spm->BadRP category. Bad rp (Bad Realignment Parameters) columns of interest are “badvolumes”, ie. A list of volume timing from the sequence where incorrect realignment occurred, and “outcome”, the simpler textual output (“OK”, “DUBIOUS”, “BAD”)

5.6 Can I have a short paragraph summarizing all QC-related auto processings ?

180

The following paragraphs can be copied or adapted:

"In order to detect and remove blatant outliers, an automatic procedure has been used. In this context, outliers refers mainly to data with obvious problem, such as missing values or heavy misregistration, but also subject maps which exhibit largely disparate activations profile for expected contrasts, which usually reflects task behavioural misdoing. Those had been applied blindly database-wide, and is therefore not biased to traits of interest of the current study.

185

First, at the timeserie level, a mean-square based metric identifies spikes as specific slice- and time-limited burst, and those are corrected for using temporal linear interpolation. Second, following the SPM realignment procedure, another metric flags remaining time-points where the volume mean-square difference to the (robust) average raises above a threshold pre-calibrated on 180 random subjects, and the subject data is discarded if more than one such bad volume occurs per minute.

190

At the activations maps level, an automatic procedure ranks the maps which departs unreasonably from the group. The procedure rely on a distance metric from each individual activation profile to the average one, computed as follows. In MNI space, a set of Region of Interest is defined based either on anatomical or functional knowledge for the relevant contrast, and every subject's average effect over each of those ROI is gathered. The $n * p$ resulting matrix, where n is the number of subjects and p is the number of ROIs, is used to estimate to $p * p$ covariance matrix of the ROIs signal across the subjects. To improve the robustness of the covariance measurement, the 0.5% upper and lowest values had been trimmed out each component before the estimation ; Additionally, the covariance matrix is inverted using an eigen-decomposition for which the 25% least informative component are nullified. This allows to further compute the Mahalanobis distance of every subjects to the group mean. As it had been checked using subgroups that the results doesn't depend much on the group size anymore after a certain size, the number of subjects involved to estimate to metric has been fixed to 1000. The ability to extract this multivariate, Mahalanobis distance from every subjects to the fixed group average allows to quickly compute their score and compare it to a relevant non-parametric threshold, to flag them as outliers. A visual checking did actually confirm that the more extreme scores corresponds to obvious artefacts."

195

200