**Cleaning stripe artifacts**

**Richard Dinga**

**r.dinga@ggzingeest.nl**

Cleaning of artifacts is based on identifying noise components in melodic decomposition and and then cleaning them using fsl\_regfilt and automatizing this process using FIX. Best results are obtained by cleaning artifacts twice; first before any preprocessing and then after motion correction (no mask, no smoothing).

**All preprocessing and cleaning steps are:**

Cleaning part

1. Convert data to nii

2. MELODIC (mo mask)

3. filter classified noise components

4. Motion correction

5. MELODIC (mo mask)

6. filter classified noise components

**Rest of the processing**

7. Grand mean scaling

8. Spatial smoothing

9. ICA AROMA

10. high pass filter

11. first level glm FEAT

12. registration of statistical maps to MNI space

Noise components need to be classified manually for approximately 10 subjects per each scanning site and task. Rest of the subjects can be cleaned automatically using FIX. It's recommended to train FIX on data from each site and task separately.

**Step 0**

Obtain and install all required software.

FSL

FIX - [http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FIX/UserGuide#Downloading\_and\_Installing\_FIX](http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FIX/UserGuide" \l "Downloading_and_Installing_FIX)(require matlab and R with additional toolboxes)

Melview – Recommended for manual component classification <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Melview>

ICA\_AROMA - <https://github.com/rhr-pruim/ICA-AROMA>

Create training training subjects file – text file with one subject number per line of subjects that will be used to manually classify components and train FIX. Fix will be trained on all subjects from this list, therefore subjects from different scan sites should be in different files.

Create text files with rest of the subjects that will be cleaned automatically by FIX

Create one text file with list of subjects TR's: TRs.txt with subject number + subjects TR per line. For example:

110123 2.3

110356 2.303

123456 2.29

**Step 1**

Run script 1.sh per each subject from the training set. This script takes only subject number as argument, however it's necessary to change paths inside to FIX installation directory, FSL directory, functional data and T1 data. Also user should change beginning of the script if the data are not already in the nifty format.

This script will run first MELODIC, motion correction and registration and it will create first set of features for FIX training. Motion correction and registration is run at the moment only to create necessary files for the FIX to function correctly, however raw data are being usedto classify and clean components.

Script finished successfully if there is a features.csv file created in the ${SUBJECTNUMBER}/1st\_cleaning/FIX

If that is not the case and all the other steps finished correctly, then it's necessary to look at the errorLog.txt and logMatlab.txt files and find out what is going on. It's possible that the fix is not installed properly or some of the paths are not set up correctly or some of the steps failed for various reasons.

**Step 2**

Classify noise components.

Look at each MELODIC component in all 3 axes and decide if it's a stripe or other artifact to clean or not. Save numbers of noise components in file called hand\_labels\_noise.txt in ${SUBJECTNUMBER}/1st\_cleaning directory. Last line of the file should be comma separated numbers of noise components inside squared brackets. For example: [1,2,3,77,88,]

It doesn't matter what is before the last line. This is also the format in which noise components will be saved using melview, which is also recommended for this task. It's necessary to look at the components in all 3 axes, because stripes are basically invisible in the default axial view and vice versa for other artifacts.

**Step 3**

Train FIX

Run script 2\_training.sh . This script takes two arguments 1.path to file with numbers of training subjects, 2. number 1 or 2 indicating if it's a first or second cleaning step. This step will run correctly only if the previous script run correctly on all the training subjects. So if all the training subjects has features.csv file

This script should create directory of the name of the training subjects file + 1st\_cleaning\_LOO (so if my file with training subjects numbers is called subjects.txt there should be file called subjects\_1st\_cleaning\_LOO), and also file with the same name and ending with .Rdata. Also, in each of the subjects 1st\_cleaning directory, there should be called fix4melview\_somename\_thr1.txt (with different numbers). Those are components classified by FIX for different thresholds. Lower threshold means more conservative guess. Those files can be loaded and examined in melview. Optimal threshold will be data dependent, in case of Tower of London data it was 50, however it might vary. Threshold should be high enough so most of the stripes components are classified as noise without any of the true signal (which might be hard to recognize, since we are running melodic on data that are not preprocessed).

**Step 4**

Clean the data

Run script 3\_cleaning.sh as a previous script, it takes two arguments. Text file with training subjects numbers and number indicating if it's a first or second cleaning. It's required to change the threshold number inside the script ( in the “clfs=$(tail ${cleaning\_folder}/fix4melview\_\*\_LOO\_thr50.txt -n 1 | tr -d '[]')” line). If the desired threshold is bigger than a 50, then uncomment the optional part of the script and follow the instruction written there.

In case of not cleaning any components, script will print out warning. This is expected in the second cleaning step for some subjects, however it should not happened during the first cleaning step. It means nothing is getting cleaned and threshold should get adjusted or classification files are missing or have wrong names. In the end there should be a file called cleaned\_data.nii.gz in the 1st\_cleaning directory

**Step 5**

Run motion correction, second melodic and create second set of fix features

Run script 4.sh . As in the first script, there should be features.csv file created in the ${SUBJECTNUMBER}/2nd\_cleaning/FIX folder

**Step 6**

Train Fix

Run script 2\_training.sh with 2 as an second argument and repeat step 3 (classification of components, looking for optimal threshold). If there are almost no stripe components present, second cleaning can be skipped.

**Step 7**

Clean the data

Repeat step 4 with 2 as an second argument argument

After this step, training data should be reasonably clean.

**Step 8**

Run rest of the preprocessing

Run script called 5.sh that will run rest of the preprocessing pipeline and first level feat. Feat should be of course change for each task.

**Step 9**

Cleaning rest of the data

In order to clean rest of the data, run script clean\_rest.sh . This will run script 1.sh, classify components automatically using .RData files created running training script.

