Remarks on FMRI processing, with example scripts in afni_proc.py

Paul A. Taylor, Daniel R. Glen, Richard C. Reynolds
Scientific and Statistical Computing Core (NIMH, NIH, USA)
Contact: paul.taylor@nih.gov Web: https://afni.nimh.nih.gov/pub/dist/doc/htmldoc/

OHBM-2025
Poster #1566

Introduction

FMRI processing is complicated. It relies on various computational procedures, including alignment, "data cleaning" (such as despiking and censoring), time series analysis and statistical modeling. It is important to make pipeline steps consistent with the study aims.

Here, we discuss a few specific processing considerations, matching relevant choices with various study designs and data properties. Several of these are general tips and considerations, which may apply across all software tools.

We also present a set of demo scripts for these using the open source AFNI toolbox[1] and its FMRI pipeline generation tool, afni_proc.py[2] (AP). These include task-based and resting state examples, with volumetric and surface processing, for both ROI-based and voxelwise analyses.

Q&A on Important FMRI Processing Choices

Should I blur/smooth the FMRI data, and by how much?

Blurring helpfully *increases* local TSNR (but *reduces* spatial specificity, so we must balance); in AP, it is added via the “blur” block. Note that some blurring always occurs from motion correction and alignment.

- For voxelwise single echo FMRI, one might blur with a FWHM (full-width at half-maximum) of 1.5-2 times the minimum voxel size.
- For voxelwise multi-echo FMRI, there will already be higher TSNR, so one might blur just slightly above voxel dimension.
- For ROI-based studies, blurring should *not* be applied (in AP, do *not* add the "blur" block), since blurring leaks outside signal into each ROI. Note that TSNR should increase by averaging within ROIs, anyways.
- Blurring on a surface helps ensure that GM is blurred with local GM (which is harder to constrain in volumetric processing).

Should I use single- or multi-echo (ME) FMRI?

ME-FMRI can increase TSNR, just from optimal combination (OC) of echos (Posse et al., 1999). MEICA methods may remove non-physiological features, further boosting TSNR. In AP, use the "combine" block to apply any of these (including the tedana[3] version of MEICA). Note ME-FMRI scans often use multiband and/or slice-acceleration. Keeping these factors low helps reduce artifacts, but check for cross-slice correlations. For example, use InstaCorr in the APQC HTML[4,5]. But if there are no artifacts, ME-FMRI generally helps increase signal strength. See Gilmore et al. (2022)[6] for a practical comparison of single- and multi-echo FMRI results in a naturalistic study.

How much do I need to know about areas of interest?

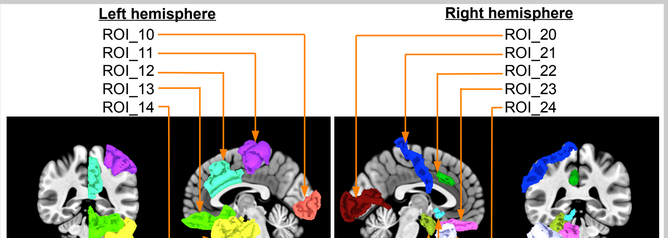
It is challenging to get strong, undistorted FMRI signal everywhere in the brain. Having a list of locations of interest (even for voxelwise studies) helps determine acquisition settings and voxel sizes. It also guides quality control, so can be more sure that your data are stable and reliable in your areas of interest. Consider:

- Inferior frontal, subcortical and temporal lobe regions often have low TSNR, signal dropout and distortion.
- Studying ROIs that are small and/or contain narrow features might require high-resolution EPI.

In AP, tables of TSNR and ROI shape properties can be displayed in the APQC HTML report. This helps assess that both TSNR is stable and the EPI spatial resolution is fine enough to capture the region well:

- A default set of ROIs around the brain will be checked, if a known template space is used.[5]
- You can also load your own set of ROIs (via "-anat_follower_ROI", "-ROI_import" and "-mask_segment_anat" options) for these checks, which will also be shown in the APQC HTML report.

ROIs in default MNI TSNR and shape table



TSNR and shape table (warnings highlighted)

ROI	Nvox	Nzer	Dvox	Tmin	T25%	Tmed	T75%	Tmax
10	448	8	2.2	44	143	206	245	328
11	1071	4	2.8	29	182	237	274	452
12	794	0	2.4	65	121	159	200	305
13	1337	5	3.0	5	18	40	82	325
14	1971	1345	3.0	8	50	73	105	185
20	839	15	2.2	9	130	179	212	330
21	282	15	1.4	67	171	220	265	405
22	92	0	1.7	108	174	203	240	314
23	342	0	2.2	8	15	21	28	93
24	660	433	2.2	15	42	62	110	271
30	180	11	2.2	10	27	47	76	140
31	34	0	1.0	62	166	195	217	237

How can I reduce EPI distortion?

One can acquire phase images (field maps) or an opposite phase-encoded EPI[7]. Neither can fully remove distortion, but each helps and adds negligible scan time. AP can directly integrate either.

- Using opposite phase-encoded EPI seems better in most software[8]. Acquire 5-10 reverse encoded volumes (the number reduces odds of subject motion ruining the data), which takes only 10-20 s total. In AP, use "-blip_forward_dset" and "-blip_reverse_dset".
- View both the raw EPI and results of EPI-anatomical alignment to judge distortions (Fig 1), as shown in AP's quality control HTML [5].

Are there downsides to bandpassing rest FMRI within 0.01-0.1 Hz?

<u>*Yes.*</u> Bandpassing is statistically expensive and removes a large fraction of degrees of freedom (DFs): 60% when TR=2s, 80% when TR=1s.

Example DF usage with 0.01-0.1 Hz bandpass

initial DF	216	100.0%
DF used for regs of interest	0	0.0%
DF used for censoring	2	0.9%
DF used for polort	5	2.3%
DF used for motion	12	5.6%
DF used for bandpass	129	59.7%
total DF used	148	68.5%
final DF	68	31.5%

Example DF usage with**out** 0.01-0.1 Hz bandpass

initial DF	216	100.0%
DF used for regs of interest	0	0.0%
DF used for censoring	2	0.9%
DF used for polort	5	2.3%
DF used for motion	12	5.6%
total DF used	19	8.8%
final DF	197	91.2%

This may create mathematical issues when motion censoring is applied[9], so AP checks DFs carefully to warn users. Note that meaningful signal patterns exist above 0.1 Hz, so useful features may be removed from the data[10,11]. If bandpassing, consider using a less expensive range (e.g., 0.01-0.2 Hz).
Also, bandpassing is not always performed correctly. It should typically be applied as part of the regression step (not separate Fourier), to avoid spectral leakage and other artifacts[2,12]. AP does bandpassing correctly, in the “regress” block.

Should I use tissue-based regressors in the processing?

Non-GM, tissue-based regressors may help remove of non-neuronal BOLD features from FMRI (esp. non-task data). This includes PCs of regions (CompCor; in AP, "-regress_ROI_PC"), local WM (in AP, "-regress_fast_anaticor"), etc.
A key assumption for these methods is that non-GM signals contain only non-neuronal effects, like motion. Caveats:

- One must thus make sure that non-GM tissue maps do *not* intersect with actual GM (in AP, they are eroded before or during processing). Check carefully for EPI distortion or other artifacts that spread signals around, to avoid GM overlap.
- Recent work looking in detail at BOLD signals in non-GM tissue suggests that assumptions of non-GM-BOLD-like signal may not be obvious: Gore et al. (2019) provide a review of early work for WM, and see also Wang et al. (2022); Gonzalez-Castillo et al. (2022) have shown that signals in ventricles can correlate strongly with physiological measures and even GM. Chen et al. (2023) used local HRF modeling to show that WM signals are typically not null and can carry useful information.

Therefore, while using local tissue regressors can help reduce some artifacts (e.g., Jo et al. 2020[13]), care should be taken with assumptions of non-GM signals. Likely more work will be required for this topic.

Conclusions

FMRI researchers must make a large number of choices, aiming to match data acquisition and processing with study design and goals. We have provided a Q&A of several considerations, based on experience with both processing and developing methods.

afni_proc.py allows researchers to control a large number of details about the processing, which can be easily commented and shared. For more details on processing and quality control, see these papers:

Reynolds et al. 2024

Taylor et al. 2024

Acknowledgments and References
The research and writing of the manuscript were supported by the NIMH Intramural Research Programs (ZICMH002888) of the NIH (HHS, USA). This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov).
[1] Cox RW (1996). Comput Biomed Res 29(3):162-173.
[2] Reynolds RC, Glen DR, Chen G, Saad ZS, Cox RW, Taylor PA (2024). Imaging Neuroscience 2:1-52.
[3] DuPre E, Salo T, Ahmed Z, Bandettini PA, …, Handwerker DA (2021). The Journal of Open Source Software 6(66), 3669.
[4] Song S, Bokkers RPH, Edwardson MA, ..., Latour LL (2017). PLoS ONE 12(10):e0185552.
[5] Taylor PA, Glen DR, Chen G, Cox RW, Hanayik T, Rorden C, Nielson DM, Rajendra JK, Reynolds RC (2024). Imag Neurosci 2: 1-39.
[6] Gilmore AW, Agron AM, González-Araya EI, Gotts SJ, Martin A (2022). Front Neurosci 16:854387.
[7] Chang H, Fitzpatrick JM (1992). IEEE Trans. Med. Imaging, 11:319-329
[8] Roopchansingh V, French JJ, Nielson D, Reynolds R, Glen D, D’Souza P, Taylor P, Cox R, Thurm A (2020). bioRxiv 2020.09.28.306787
[9] Caballero-Gaudes C, Reynolds RC (2017). Neuroimage 154:128-149.
[10] Gohel SR, Biswal BB (2015). Brain Connectivity 5(1):23-34.
[11] Shirer WR, Jiang H, Price CM, Ng B, Greicius MD (2015). Neuroimage. 117:67-79.
[12] Hallquist MN, Hwang K, Luna B (2013). Neuroimage. 82:208–225.
[13] Jo HJ, Gotts SJ, Reynolds RC, Bandettini PA, Martin A, Cox RW, Saad ZS (2013). Journal of Applied Mathematics, vol. 2013, ID 935154.