Connectome Operations For FSL ExEcution (COFFEE): a turnkey pipeline for preprocessing of fMRI data

McAvoy MM, Liu L, Zhou R, Philip BA.

Author Details:

Mark M. McAvoy, PhD. mcavoy@wustl.edu. Affilitation: Program in Occupational Therapy, Washington University School of Medicine, St. Louis, MO, United States.

Lei Liu, PhD. lei.liu@wustl.edu. Affilitation: Department of Biostatistics, Washington University School of Medicine, St. Louis, MO, United States.

Ruiwen Zhou, PhD. zhour@wustl.edu. Affilitation: Department of Biostatistics, Washington University School of Medicine, St. Louis, MO, United States

Benjamin A. Philip, PhD.* bphilip@wustl.edu. Affilitation: Program in Occupational Therapy, Washington University School of Medicine, St. Louis, MO, United States.

*: Corresponding author.

Data availability statement: All software is openly available at https://github.com/PhilipLab/COFFEE-pipeline. The data that support the findings of this study are available on request from the corresponding author, BAP, upon reasonable request.

Funding statement: This work was funded by NIH/NINDS R01 NS114046 to BAP.

Conflict of interest disclosure: Author BAP and Washington University in St. Louis have a licensing agreement with PlatformSTL to commercialize the drawing program used in this study.

Patient consent & ethics approval statement: All participants gave informed consent, and all procedures were approved by the Institutional Review Board at Washington University in St. Louis School of Medicine.

Acknowledgements: The authors would like to thank Taewon Kim, Samah Gassass, and Setsu Uzume for their assistance collecting the data in this manuscript.

ABSTRACT

Numerous methods exist to analyze functional MRI (fMRI) data, but no software currently exists to integrate the commonly-used FSL statistical analysis software with alternative preprocessing methods from the Human Connectome Project. Here we developed the Connectome Operations For FSL ExEcution (COFFEE) pipeline to integrate Human Connectome preprocessing, including Freesurfer brain extraction and registration, for FSL general linear model analysis of volumetric fMRI data. We tested COFFEE on a real fMRI dataset (N=26) wherein participants performed a precision drawing task during fMRI scanning. COFFEE preprocessing, compared to traditional FSL preprocessing, led to lower inter-individual variability across the brain, more precise brain extraction, and greater detected activation in sensorimotor areas contralateral to movement. The COFFEE pipeline provides a turnkey alternative to FSL preprocessing with substantial advantages for analysis of task fMRI data. COFFEE software is available online for research use and further testing.

KEYWORDS: image analysis pipeline, fMRI, brain, software, magnetic resonance imaging, humans, brain mapping

INTRODUCTION

Numerous methods and software packages exist for the analysis of fMRI data, and many of these approaches offer different advantages and tradeoffs. One commonly used analysis package is the fMRIB Software Library (FSL; Jenkinson et al., 2012), particularly its FEAT software for general linear model (GLM) analysis of volumetric fMRI data (Woolrich et al., 2001; Woolrich et al., 2004). However, despite FEAT's popularity, FSL preprocessing has some shortcomings, one of which is its reliance on the FSL Brain Extraction Tool (BET; Smith, 2002) for identifying brain voxels in the scan volume. BET is prone to under- and over-extraction of brain voxels (Boesen et al., 2004; Zhuang, Valentino and Toga, 2006; Klein et al., 2010), and has limited options for adjusting its algorithmic criteria (Jenkinson, 2018). Together these shortcomings mean that BET either produces inaccurate brain extractions or requires time-consuming and error-prone manual adjustment (Quilis-Sancho, Fernandez-Blazquez and Gomez-Ramirez, 2020; Mohapatra, Gosai and Schlaug, 2023). Therefore, alternative FSL-compatible preprocessing approaches (including brain extraction) could improve the reliability and accuracy of fMRI data analysis.

One alternative preprocessing approach was developed by Glasser et al. (2013) for the Human Connectome Project (HCP). HCP preprocessing includes the integration of Freesurfer software for brain extraction and parcellation (e.g. Fischl and Dale, 2000; Fischl et al., 2002; Fischl et al., 2004). HCP preprocessing provides many additional advantages, including: removal of spatial artifacts and distortions, creation of cortical surface and myelin maps, precise within-individual cross-modal registrations, and surface- and volume-based registrations to standard spaces (Glasser et al., 2013). However, HCP preprocessing was designed for idiosyncratic HCP data, including a focus on resting state over task-based fMRI data (Glasser et al., 2013). No turnkey solution yet exists to connect the advantages of HCP preprocessing (including Freesurfer brain extraction and registration) to task fMRI analysis packages such as FSL.

Here, we created the COFFEE (Connectome Operations For FSL ExEcution) pipeline as a turnkey solution to integrate HCP preprocessing with FEAT volume-based GLM analysis. We evaluated the effects of COFFEE on motor task fMRI results by performing COFFEE and FSL preprocessing on the same dataset, and with the same subsequent FEAT GLM analysis, and compared the two methods for changes in task-based activation, brain extraction, and inter-individual variability.

MATERIALS & METHODS

Connectome Operations For FSL ExEcution (COFFEE) Pipeline

The goal was to create a Macintosh compatible turnkey pipeline for processing fMRI data that integrated Freesurfer (e.g. Fischl and Dale, 2000; Fischl et al., 2002; Fischl et al., 2004) with the GLM implemented in FSL FEAT (Woolrich et al., 2001; Woolrich et al., 2004; Jenkinson et al., 2012) from FSL 6.0.6. To this end, version 3.27 of the HCP pipeline was downloaded (https://github.com/Washington-University/HCPpipelines/tree/v3.27.0) and the structural and fMRI volume pipelines were customized. The steps of the HCP pipeline include: Freesurfer brain extraction and parcellation,

motion correction, field map distortion correction, and warping to the 2mm MNI atlas via the "one step resampling" procedure (Glasser et al., 2013).

COFFEE updates to the HCP pipeline included:
(1) Rudimentary Bash code changes for Macintosh compatibility
(2) Aesthetic changes including the elimination of the hard coded paths and file names
(3) New option to perform structural analysis without a T2 weighted image
(4) New option to specify the native resolution as 0.7, 0.8, or 1mm
(5) Extended compatibility from Freesurfer version 5.3.0-HCP to versions 7.2.0, 7.3.2 and 7.4.0 (https://surfer.nmr.mgh.harvard.edu/)
(6) New option to edit the Freesurfer output and then rerun subsequent portions of the structural pipeline
(7) New option to leave fMRI time series in native HCP space, vs. default behavior of registering to the 2mm MNI atlas
(8) Reading the phase encoding direction from a json file rather than assuming it from the file name
(9) Additional optional fMRI processing steps to match FEAT:
    9a. Nonlinear spatial (SUSAN) filtering (Smith and Brady, 1997)
    9b. Normalization of intensity values to a median value 10000. This replaces the mean intensity normalization in the HCP pipeline.
    9c. Nonlinear temporal high pass filtering (Woolrich et al., 2001).
    9d. B0 distortion correction via spin echo field maps (Jezzard and Balaban, 1995).

This completes the COFFEE pipeline, and the fMRI time series are now ready for analysis in FEAT using its "Statistics" option (rather than "Full Analysis" or "Preprocessing").

COFFEE requires (and provides) one additional step after first-level FEAT analysis, if the user wants to run higher-level FEAT analyses. The first-level FEAT results will lack FEAT-formatted registration files, which are required for higher-level analyses. (These registration files cannot be created before first-level FEAT analysis, because FEAT requires them to be placed in the output directory of a FEAT analysis.) Therefore, after COFFEE and first-level FEAT analyses, there is a final "FEAT-adapter" step to use the processed data to compute FEAT-formatted registration images and matrices. COFFEE includes the option to run the full series "COFFEE pipeline, first-level FEAT analyses, FEAT-adapter, higher-level FEAT analyses" via a single command.

Pipeline Implementation

For ease of usability, a suite of setup scripts generate participant-specific pipeline scripts and batch files for group execution. The setup scripts accept a so-called *dat* file as input which matches our protocol but can be easily modified for any protocol. The *dat* file identifies the type of each nifti-format scan (e.g., T1 weighted). Example files and operational documentation are available at: https://github.com/PhilipLab/COFFEE-pipeline

Structural pipeline. The T1 weighted image (required) and the T2 weighted image (if available) are used to extract, segment and parcellate the brain along with registration to the MNI atlas (Fonov et al., 2009) via FSL and Freesurfer (Glasser et al., 2013). Basic usage to generate the structural pipeline scripts is with *dat* as the sole argument,

COFFEEstructpipeSETUP.sh *dat*

The scripts are written to the output folder specified by the *dat* with the addition of a label identifying the version of Freesurfer. For example, the results presented here utilize Freesurfer 7.4.0,

*output folder*7.4.0/*participant label*_hcp3.27struct.sh

where *output folder* and *participant label* are both specified in the *dat* file, and “hcp3.27” acknowledges the pipeline’s parentage to the Human Connectome Project version 3.27 pipeline. A second script is also created (*output folder*7.4.0/*participant label*_hcp3.27struct_fileout.sh), for convenience, that automatically redirects any runtime messages sent to the command line.

Functional pipeline. The outputs of the structural pipeline along with the reference images (if available) and field maps (if available) are used to process the fMRI time series. This includes motion and distortion correction along with cross-modal registration to the MNI atlas notably by a so-called one-step resampling approach (Glasser et al., 2013). Similarly to the structural pipeline, basic usage utilizes the *dat* to generate the functional scripts,

COFFEEfMRIpipeSETUP.sh *dat*

To utilize FSL FEAT-like processing, options are provided for SUSAN filtering and high pass filtering. For example,

COFFEEfMRIpipeSETUP.sh *dat* -f 6 -p 60 -o *fsf level 1.txt* -t *fsf level 2.txt*

will output fMRI time series that have been SUSAN filtered (-f) with a 6 mm FWHM kernel and high pass filtered (-p) with a cutoff of 60s. In addition, first (-o) and second (-t) FEAT analyses may also be performed by providing the appropriate files (e.g. *fsf level 1.txt* and *fsf level 2.txt*, respectively). Following the above example, and as described earlier in the structural pipeline, generated scripts include *output folder*7.4.0/*participant label*_hcp3.27fMRIvol.sh and *output folder*7.4.0*/participant label*_hcp3.27fMRIvol_fileout.sh along with an additional script,

*output folder*7.4.0/*participant label*_FEATADAPTER.sh

provided for convenience as it just repeats the FEAT portions of the fMRIvol script, thus facilitating easy hand editing for the user to perform additional first and second level analyses. If the user prefers, this may also be done automatically by the functional pipeline setup script,

COFFEEfMRIpipeSETUP.sh *dat* -o *fsf level 1.txt* -t *fsf level 2.txt* -F

where F(i.e., FEATADAPTER) limits to just the FEAT portions of the pipeline.

Pipeline testing

COFFEE was tested by performing two parallel analyses on the same dataset. Twenty-six right-handed adults (18 female; ages 52 ± 16, range 24-76) performed a precision drawing task (see *fMRI Task* below) in the fMRI scanner. 11/26 participants had peripheral nerve injuries to their right arm, but the differences between groups (categorical or performance-related) were not a focus of the current study. Two participants were unable to complete the drawing task with their right hand, so right hand (RH) drawing analyses were limited to 24 participants. All participants gave informed consent, and all procedures were approved by the Institutional Review Board at Washington University in St. Louis School of Medicine.

fMRI Acquisitions. Scans were performed on a Siemens PRISMA 3T MRI scanner. BOLD EPIs for fMRI were collected using a T2*-weighted gradient echo sequence, a standard 64-channel birdcage radio-frequency coil, and the following parameters: TR = 662 ms, TE = 30 ms, flip angle = 52°, 72 x 72 voxel matrix, FOV = 216 mm, 60 contiguous axial slices acquired in interleaved order, resolution: 3.0 x 3.0 x 3.0 mm, bandwidth = 2670 Hz/pixel, multi-band acceleration = 6x. Siemens auto-align was run at the start of each session.

High-resolution T1-weighted structural images were also acquired, using the 3D MP-RAGE pulse sequence: TR = 4500 ms, TE = 3.16 ms, TI = 1000 ms, flip angle = 8.0°, 256 x 256 voxel matrix, FOV = 256 mm, 176 contiguous axial slices, resolution: 1.0 x 1.0 x 1.0 mm. A T2-weighted image was also acquired at: TR = 3000 ms, TE = 409 ms, 256 x 256 voxel matrix, FOV = 256 mm, 176 contiguous axial slices. Spin echo field maps were also collected before the functional runs.

fMRI Task. Each participant completed 6 BOLD functional scans. During each scan, the participant used one hand (3 scans with right hand RH, 3 with left hand LH, interleaved order) to perform a precision drawing task based on the STEGA app (Philip and Frey, 2014; Philip et al., 2023). In brief, participants saw hollow shapes, and were instructed to draw a line within the bounds of the shape, as fast as possible while prioritizing staying in-bounds over speed.

Participants saw a video overlay of their (transparent) hand/pen over the drawn shape, using an MRI-compatible drawing tablet with bluescreen technology (Karimpoor et al., 2015), as shown in **Figure 1**. The task was presented in a block design with 15.2 seconds (23 images) of drawing, followed by 15.2 seconds (23 images) of rest (fixation cross). A scan comprised ten cycles of draw/rest, with an additional rest block at the start, and 3.3 seconds (5 images) of additional rest at the end, leading to a total duration of 5:23 (230 images of drawing, 258 rest).

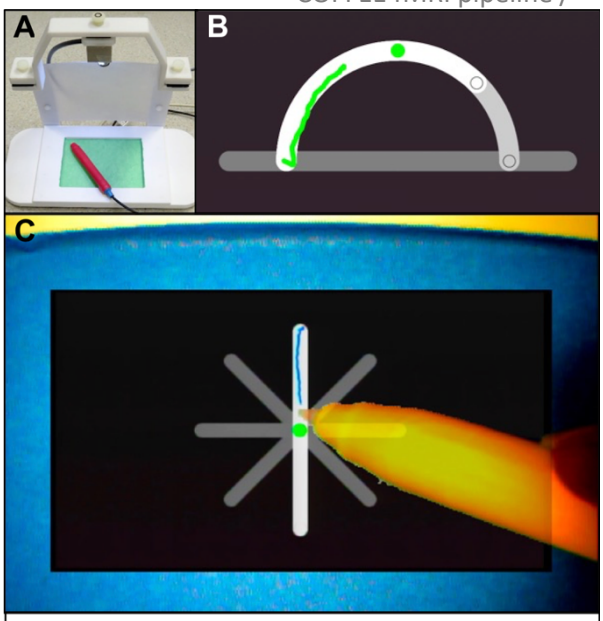

Figure 1. Motor task used for pipeline testing. A: MRI-compatible tablet. B: STEGA precision drawing task. C: Participant view of bluescreen integration.

Parallel analyses: COFFEE vs. FSL-only. In the "COFFEE analysis," the COFFEE pipeline described above was used to preprocess data with the following options: native resolution of 1 mm, distortion correction using spin echo field maps, spatial smoothing (SUSAN) was applied with a kernel of 6 mm FWHM, intensity was normalized via “grand median scaling” to a value of 10000, high pass temporal filtering was applied with 60 sec cutoff, and outputs registered to 2 mm MNI atlas. Subsequently, GLM statistical analysis was performed with FEAT version 6.0.7 (Smith et al., 2004). Explanatory variables (EVs) were modeled, along with their temporal derivatives, according to the block design described in the previous section (*fMRI Task*). Volumes with excess head motion were addressed (Siegel et al., 2014) via additional confound EVs calculated within each time series as framewise displacement of 75th percentile + 1.5* interquartile range, using the FSL script FSLMotionOutliers. The hemodynamic response was accounted for by convolving the model with a double-gamma function. First-level contrasts of parameter estimates (COPEs) were calculated for Task vs. Rest. The first-level COPE (Task > Rest) served as input to higher-level analyses performed using a fixed-effects model. Z-statistic (Gaussianized T) images were thresholded using clusters determined by $Z \geq 3.1$ and a corrected cluster significance threshold of $p < 0.05$. The first-level COPEs were averaged across runs for each participant (second level). Region of interest (ROI) analyses were performed on second-level (i.e. participant-level) data (see *ROI Analyses* below).

In the "FSL-only analysis," preprocessing included the following steps: Non-brain structures were removed using BET. Head movement was reduced using MCFLIRT motion correction. EPI unwarping was performed to correct for distortions due to magnetic field inhomogeneities using FSL PRELUDE and FUGUE, and distortion correction using spin echo field maps. Paralleling the COFFEE implementation, spatial smoothing, intensity normalization, and the high-pass temporal filter were applied with the same settings as above. Functional data were registered with the high-resolution structural image using boundary-based registration (Greve and Fischl, 2009), and resampled to 2x2x2 mm resolution using FLIRT; the participant images were then registered to standard images (Montreal Neurological Institute MNI-152) using FNIRT nonlinear registration (Andersson, Jenkinson and Smith, 2007). Subsequently, GLM statistical analysis was performed with FEAT as described for the COFFEE analysis above.

Individual example participants were selected for illustration on an *a priori* basis via the following criteria: participants were ranked by the metric "magnitude difference between analysis methods, in the whole-cortex ROI" (averaged between LH condition and RH condition), from which the 25th percentile participant (#1007), median participant (#2007), and 75th percentile participant (#2019) were selected. Single-run example images were taken from the Draw-RH condition, run 2 (of 3).

ROI Analyses. To quantitatively compare the two analyses, a region-of-interest (ROI) approach was used, involving 42 ROI variables (21 ROIs, calculated separately for LH-draw and RH-draw conditions). These 21 ROIs comprised: 6 motor sub-regions (3 per hemisphere), 14 networks (7 per hemisphere), and 1 whole-cortex ROI. Specifically, the six primary motor cortex (M1) ROIs were chosen as 5 mm radius spheres based on normative activation maps (Smith and Frey,

2011): Hand-M1 ("Hand;" left hemisphere x = -38, y = -24, z = 54; right hemisphere x = 38, y = -24, z = 54), Lip-M1 ("Lip;" left hemisphere x = -54, y = -8, z = 42; right hemisphere x = 44, y = -12, z = 42) and Foot-M1 ("Foot;" left hemisphere x = -4, y = -28, z = 72; right hemisphere x = 8, y = -24, z = 74). The 14 network ROIs were used, taking the full networks from the NITRC (2022) atlas of cortical networks in the Yeo et al. (2011) network parcellation: medial visual, sensorimotor, dorsal attention, ventral attention, frontoparietal, default mode (DMN), and subcortical; each in left and right hemispheres. Finally, the "whole-cortex" ROI was created as the union of all twelve cortical network-based ROIs. For each participant, the signal magnitude was calculated as the mean % BOLD signal change for the Task > Rest contrast in second-level (within participant) analyses.

COFFEE analysis was compared with FSL-only analysis via two approaches. The first approach focused on inter-individual variability. The primary variable was the standard deviation (across the signal magnitudes of the 26 participants, within each ROI). Direct statistical comparisons within each ROI used a pairwise Pitman-Morgan test (Morgan, 1939) to compare the inter-individual variability of COFFEE signal magnitudes vs. the inter-individual variability of FSL-only signal magnitudes. A big-picture overview was developed by comparing the 14 network ROI standard deviations for COFFEE analysis, vs. the 14 standard deviations for FSL-only analysis, via a paired-sample t-test. Finally, to explore the spatial patterns of where the two analyses differed, a 2 (Method: COFFEE vs. FSL-only) * 9 (Area) * 2 (Hemisphere: left, right) * 2 (Condition: LH-draw, RH-draw) ANOVA was performed on the standard deviations. The "area" factor had 9 levels instead of 10 because the Foot region was omitted as an outlier (high difference between COFFEE and FSL-only in Foot, Right Hemisphere).

The second approach for comparing COFFEE vs FSL-only analysis focused on mean signal magnitude. The same process was followed as above, except: the primary variable was the mean signal magnitude (across the 26 participants). Direct statistical comparisons used paired-sample t-tests. In the signal magnitude analysis, no outliers were omitted from the ANOVA, making it 2 * 10 * 2 * 2 instead of 2 * 9 * 2 *2.

Individual example ROIs were selected for illustration on an *a priori* basis via a similar method to the selection of example participants: ROIs were ranked by the metric "magnitude difference between analysis methods", from which the 25th percentile ROI (ventral attention, right hemisphere, RH-draw condition), median ROI (ventral attention, right hemisphere, LH-Draw condition), and 75th percentile ROI (default mode network, right hemisphere, left hand) were selected.

Brain extraction comparisons. To quantify the difference in brain extraction results between analysis methods, we compared the size of post-registration (i.e. standard space) brain masks from each fMRI run, following (Zhang et al., 2019).

RESULTS

COFFEE led to reduced inter-individual variability compared to FSL-only analysis.

In all forty-two ROIs, COFFEE analysis led to numerically lower inter-individual variability (standard deviation) than FSL-only analysis, as shown in **Table 1**. This difference was statistically significant in 41/42 ROIs (98%).

To provide a big-picture overview of how our analysis methods affected data variability across the brain, we compared standard deviations (across participants) across our network-based ROIs. We found significantly lower standard deviations for COFFEE than FSL ($T = -16.7$, $p = 9.94 \times 10^{-16}$); in these network-based ROIs, COFFEE standard deviations were 0.283 ± 0.0548, while FSL standard deviations were 0.504 ± 0.101, an effect size of 1.60. Moreover, COFFEE also had lower variability in the whole-cortex ROI under either task condition ($p \leq 0.005$; **Table 1** bottom two rows), demonstrating broadly lower variability for COFFEE analysis than FSL-only analysis.

To explore the spatial pattern of where the two analyses differed, we performed a 4-way ANOVA: 2 (Method: COFFEE vs. FSL-only) * 9 (Area: see **Table 1** Column 1) * 2 (Hemisphere: left, right) * 2 (Condition: LH-draw, RH-draw) on standard deviations. We found significant main effects of Method ($F(1,33) = 106.81$, $p < 0.001$), Area ($F(8,33) = 15.45$, $p < 0.001$),

and Condition ($F(1,33) = 16.22$, $p = 0.0003$), but not Hemisphere ($p = 0.288$). We found a significant interaction for Area * Hemisphere ($F(8,33) = 3.23$, $p = 0.008$), and a trend toward a Hemisphere * Condition interaction ($F(1,33) = 3.69$, $p = 0.0633$), but no other interactions ($p > 0.6$). Post-hoc tests revealed that the main effect of Method was driven by lower variability for COFFEE (0.338 ± 0.137) than FSL-only (0.569 ± 0.188), the main effect of Area was driven by significantly higher variability for motor ROIs than for network ROIs, likely due to the motor ROIs' smaller size; and the main effect of Condition was driven by higher variability for LH-draw (0.498 ± 0.206) than RH-draw (0.408 ± 0.187). The Area * Hemisphere interaction was driven by high variability for bilateral Hand and left hemisphere Lip; the Hemisphere * Condition trend was driven by higher variability for LH-draw (either hemisphere) than "right hemisphere, RH-draw." In sum, COFFEE showed improved reliability (lower inter-individual variability) throughout the brain; this effect was more pronounced in smaller ROIs, but otherwise not specific to task-related areas or networks (e.g. sensorimotor network).

ROI			**Standard deviation (across participants)**			
Area	**Hemisphere**	**Condition**	**COFFEE**	**FSL-only**	**Improvement (FSL-COFFEE)**	***p***
Hand	Left	LH	0.456	0.671	0.216	7.43×10^{-6}
Hand	Right	LH	0.789	1.113	0.324	1.97×10^{-5}
Hand	Left	RH	0.600	0.985	0.386	1.29×10^{-4}
Hand	Right	RH	0.402	0.519	0.118	0.011
Lip	Left	LH	0.712	1.013	0.301	3.96×10^{-4}
Lip	Right	LH	0.427	0.659	0.232	8.24×10^{-5}
Lip	Left	RH	0.549	0.950	0.401	5.36×10^{-7}
Lip	Right	RH	0.320	0.452	0.132	0.009
Foot	Left	LH	0.506	1.204	0.698	1.14×10^{-7}
Foot	Right	LH	0.721	1.996	1.275	6.97×10^{-13}
Foot	Left	RH	0.256	0.951	0.695	6.24×10^{-10}
Foot	Right	RH	0.344	1.289	0.946	1.13×10^{-9}
Med Visual	Left	LH	0.385	0.557	0.172	5.92×10^{-5}
Med Visual	Right	LH	0.370	0.571	0.201	2.69×10^{-6}
Med Visual	Left	RH	0.277	0.509	0.232	1.37×10^{-9}
Med Visual	Right	RH	0.275	0.526	0.251	5.49×10^{-7}
Sensorimotor	Left	LH	0.334	0.554	0.220	4.42×10^{-6}
Sensorimotor	Right	LH	0.364	0.673	0.309	1.96×10^{-7}
Sensorimotor	Left	RH	0.258	0.441	0.183	6.26×10^{-6}
Sensorimotor	Right	RH	0.260	0.422	0.162	1.02×10^{-6}
Drsl Attention	Left	LH	0.369	0.668	0.299	4.70×10^{-6}
Drsl Attention	Right	LH	0.372	0.727	0.355	3.57×10^{-7}
Drsl Attention	Left	RH	0.311	0.634	0.323	2.19×10^{-8}
Drsl Attention	Right	RH	0.298	0.612	0.314	3.29×10^{-10}
Vtrl Attention	Left	LH	0.306	0.482	0.176	1.30×10^{-4}
Vtrl Attention	Right	LH	0.315	0.569	0.255	4.52×10^{-8}
Vtrl Attention	Left	RH	0.230	0.379	0.150	2.08×10^{-6}
Vtrl Attention	Right	RH	0.285	0.402	0.117	0.002
Frontoparietal	Left	LH	0.269	0.568	0.299	5.26×10^{-9}
Frontoparietal	Right	LH	0.276	0.557	0.282	6.19×10^{-11}
Frontoparietal	Left	RH	0.218	0.448	0.229	9.49×10^{-10}
Frontoparietal	Right	RH	0.234	0.474	0.240	2.95×10^{-12}
DMN	Left	LH	0.261	0.479	0.218	1.20×10^{-7}

DMN	Right	LH	0.269	0.492	0.223	1.32×10^{-9}
DMN	Left	RH	0.187	0.442	0.256	7.45×10^{-12}
DMN	Right	RH	0.231	0.442	0.211	7.36×10^{-8}
Subcortical	Left	LH	0.332	0.385	0.052	0.350
Subcortical	Right	LH	0.240	0.356	0.116	0.005
Subcortical	Left	RH	0.181	0.355	0.174	6.67×10^{-5}
Subcortical	Right	RH	0.206	0.385	0.179	6.93×10^{-5}
Whole cortex	n/a	LH	0.276	0.508	0.233	5.51×10^{-8}
Whole cortex	n/a	RH	0.211	0.419	0.208	2.57×10^{-10}

Table 1. Inter-individual variability (standard deviation of BOLD % signal change during Task > Rest) was consistenyly lower for COFFEE analysis than FSL-only analysis. Condition = data collected during drawing with left hand (LH) or right hand (RH). P-values calculated via Pitman-Morgan test. Med = medial, Drsl = dorsal, Vtrl = ventral.

COFFEE sometimes led to increased BOLD signal magnitude compared to FSL analyses, but not consistently

In terms of signal magnitude (participant mean % BOLD signal change for Task > Rest), COFFEE analysis led to numerically greater activity in 28/42 ROIs (67%), as shown in **Table 2**. Of the 28 ROIs with numerically higher activity for COFFEE, 15 (54%) were statistically significant. Of the 14/42 ROIs with numerically lower activity for COFFEE, 4 (29%) were statistically significant.

To provide a big-picture overview of how our analysis methods affected data variability across the brain, we compared mean signal magnitudes (across participants) across our network-based ROIs. We found significantly higher magnitudes for COFFEE than FSL ($T = 2.25$, $p = 0.033$); in these network-based ROIs, COFFEE means were 0.059 ± 0.195 (mean ± SD across participants), while FSL means were -0.003 ± 0.336, an effect size of 0.23. However, we found no differences between analysis methods in the whole-cortex ROI for either condition ($p > 0.25$; **Table 2**, bottom two rows), demonstrating that COFFEE analysis increased BOLD signal magnitudes in some networks but not overall.

To explore the spatial pattern of where the two analyses differed, we performed a 4-way ANOVA: 2 (Method: COFFEE vs. FSL-only) & 10 (Area: see **Table 2** Column 1) * 2 (Hemisphere: left, right) * 2 (Condition: LH-draw, RH-draw) on mean signal magnitudes. We found a significant main effect of Area ($F(9,37)=11.28$, $p < 0.001$), and a significant Hemisphere * Condition interaction ($F(1,37) = 6.36$, $p = 0.016$), but no other main or interaction effects ($p > 0.3$ except the following example), including no significant main effect of Method ($F(1,37) = 2.74$, $p = 0.106$). Post-hoc analysis indicated that the main effect arose from lower means for Foot, Medial Visual, and Default Mode than other ROIs; and the Hemisphere * Condition interaction arose from higher means for "left hemisphere, RH-draw" than "right hemisphere, RH-draw." Therefore, COFFEE did not lead to systematic changes in signal magnitude. However, the meaning of "changes in signal magnitude" is not necessarily clear since this is real data with no known ground truth.

ROI			**Mean (across participants)**			***p***
Area	**Hemisphere**	**Condition**	**COFFEE**	**FSL-only**	**Increase (COFFEE-FSL)**	
Hand	Left	LH	0.067	0.012	0.055	0.356
Hand	Right	LH	0.905	1.129	-0.225	0.022
Hand	Left	RH	1.138	1.455	-0.317	0.012
Hand	Right	RH	-0.121	-0.303	0.181	0.001
Lip	Left	LH	0.472	0.524	-0.052	0.602
Lip	Right	LH	0.142	0.084	0.058	0.406
Lip	Left	RH	0.420	0.557	-0.137	0.185

Lip	Right	RH	-0.028	-0.183	0.155	0.009
Foot	Left	LH	-0.115	-0.622	0.507	0.011
Foot	Right	LH	-0.121	-0.946	0.826	0.008
Foot	Left	RH	-0.068	-0.678	0.609	0.001
Foot	Right	RH	-0.165	-0.872	0.706	0.005
Med Visual	Left	LH	-0.265	-0.554	0.289	0.000
Med Visual	Right	LH	-0.237	-0.499	0.262	0.000
Med Visual	Left	RH	-0.202	-0.395	0.192	0.001
Med Visual	Right	RH	-0.255	-0.529	0.274	0.000
Sensorimotor	Left	LH	0.197	0.162	0.034	0.572
Sensorimotor	Right	LH	0.262	0.274	-0.012	0.874
Sensorimotor	Left	RH	0.319	0.397	-0.079	0.123
Sensorimotor	Right	RH	0.100	0.036	0.063	0.135
Drsl Attention	Left	LH	0.308	0.393	-0.084	0.320
Drsl Attention	Right	LH	0.348	0.466	-0.118	0.193
Drsl Attention	Left	RH	0.349	0.552	-0.203	0.015
Drsl Attention	Right	RH	0.392	0.636	-0.244	0.002
Vtrl Attention	Left	LH	0.130	0.123	0.007	0.898
Vtrl Attention	Right	LH	0.174	0.124	0.051	0.426
Vtrl Attention	Left	RH	0.141	0.143	-0.002	0.961
Vtrl Attention	Right	RH	0.146	0.171	-0.025	0.544
Frontoparietal	Left	LH	0.001	-0.128	0.128	0.086
Frontoparietal	Right	LH	0.051	-0.027	0.079	0.226
Frontoparietal	Left	RH	0.008	-0.110	0.118	0.035
Frontoparietal	Right	RH	0.085	0.052	0.033	0.533
DMN	Left	LH	-0.162	-0.425	0.263	0.000
DMN	Right	LH	-0.104	-0.316	0.212	0.000
DMN	Left	RH	-0.162	-0.447	0.286	0.000
DMN	Right	RH	-0.113	-0.328	0.215	0.000
Subcortical	Left	LH	-0.025	-0.001	-0.024	0.726
Subcortical	Right	LH	0.068	0.032	0.036	0.456
Subcortical	Left	RH	0.024	0.009	0.015	0.781
Subcortical	Right	RH	0.076	0.102	-0.026	0.636
Whole cortex	n/a	LH	0.075	0.010	0.064	0.273
Whole cortex	n/a	RH	0.086	0.056	0.030	0.523

Table 2. Task-related signal magnitude (BOLD % signal change during Task > Rest) was sometimes, but not always, higher for COFFEE analysis than FSL-only analysis. For details see Table 1.

COFFEE increased signal in whole-brain analyses, including in task-specific areas

To assess the effects of COFFEE analysis on the whole brain, we conducted a whole-brain repeated measures analysis, treating COFFEE analysis and FSL analysis as the two measures for each participant. For both LH-draw and RH-draw conditions, the COFFEE > FSL-only contrast showed significant effects in a small but distributed set of cortical networks, as shown in **Figure 2**. No areas were significant for FSL-only > COFFEE.

The cluster list in **Table 3** illustrates how the COFFEE > FSL-only activity includes deep structures that may represent noise (e.g. ventricles and white matter), but also includes areas with expected task-specific actiity (sensorimotor, parietal, and premotor areas contralateral to movement). Therefore, while COFFEE may increase spurious signals in areas with high noise susceptibility, it also enhances signal in task-specific areas that are likely of primary interest.

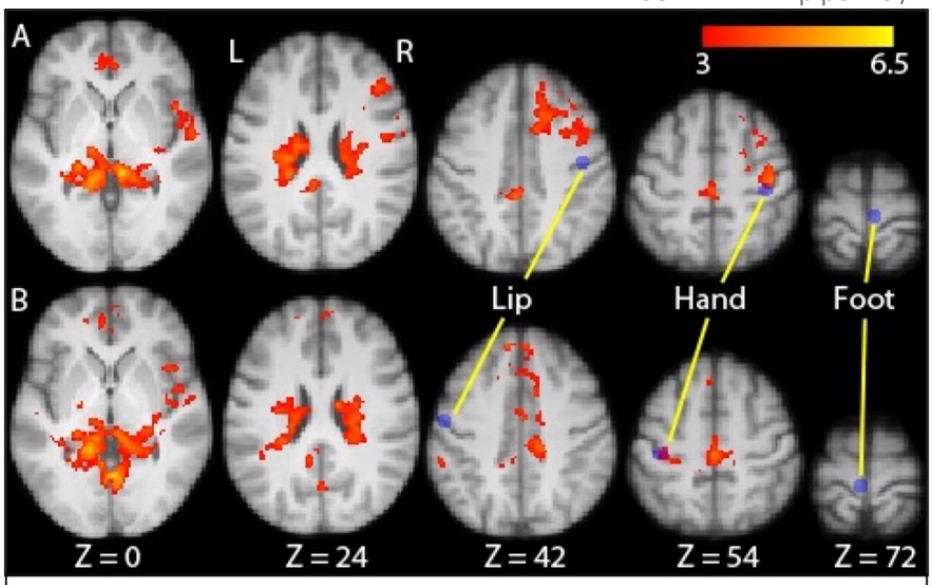

Figure 2. Significant differences in BOLD signal for COFFEE > FSL-only contrast. (No significant differences for FSL-only > COFFEE.) **A:** LH drawing condition. **B:** RH drawing condition.

Condition	Cluster area	Peak coordinates			Peak Z-score	# voxels
		x	y	z		
LH-draw	Cingulate gyrus, hippocampal subiculum, thalamus, lateral ventricles, basal ganglia, internal capsule, midbrain	12	-38	-8	6.42	7872
	Left cingulate gyrus	-4	-34	38	5.72	616
	Left subcallosal cortex (inferior prefrontal)	6	30	-12	4.97	174
	Right primary sensorimotor	42	-18	56	4.79	148
	Right superior frontal gyrus (premotor cortex)	24	4	60	3.9	77
RH-draw	Lingual gyrus, cingulate gyrus, thalamus, basal ganglia, internal capsule, lateral ventricles, midbrain, pons	8	-68	-4	5.73	8495
	Midline paracingulate gyrus	-2	56	8	4.61	1132
	Right frontal operculum	38	14	10	4.43	232
	Left precuneus & cuneal cortex	-12	-70	20	4.61	125
	Left inferior parietal lobule	-42	-68	18	4.34	117
	Left primary sensorimotor	-36	-24	52	4.57	96
	Left supramarginal gyrus (inferior parietal lobule)	-56	-38	40	3.77	83

Table 3. Significant clusters in COFFEE > FSL-only contrast. No significant clusters for FSL-only > COFFEE.

COFFEE identified smaller brain volumes, primarily through removal of non-brain voxels

To quantify the difference in brain extraction results between analysis methods, we compared the size of post-registration (i.e. standard space) brain masks. COFFEE masks were 235,222 ± 3,868 voxels (range 226,807-242,917), whereas FSL-only masks were 311,474 ± 8,072 voxels (range 297,666-329,132); COFFEE masks were smaller for all individuals, and this difference was statistically significant at the group level (paired-samples t-test: $p = 1.65 \times 10^{-35}$, $t = -114$). These smaller COFFEE masks largely reflected the successful removal of non-brain volume: for each participant, we evaluated the COFFEE-specific change (i.e., the area outside the COFFEE mask but within the FSL mask) by using the

MNI-152 standard brain mask to identify brain/non-brain voxels, and found that COFFEE removed over 15 times more non-brain voxels (72,249 ± 7,206, range 61,236-90,945) than brain voxels (4,746 ± 1,878, range 2,840-10,063). This ratio of non-brain/brain removal, while already high, reflects an over-estimation of removed brain voxels because the MNI brain mask includes non-brain areas such as the medial longitudinal fissure and inferior areas around the hypothalamus, as illustrated by our three example participant runs in **Figure 3**. Overall, these numbers show that while COFFEE (via its implementation of Freesurfer) did not produce perfect automated brain extractions, COFFEE generally provided superior brain extraction compared to FSL-only analysis.

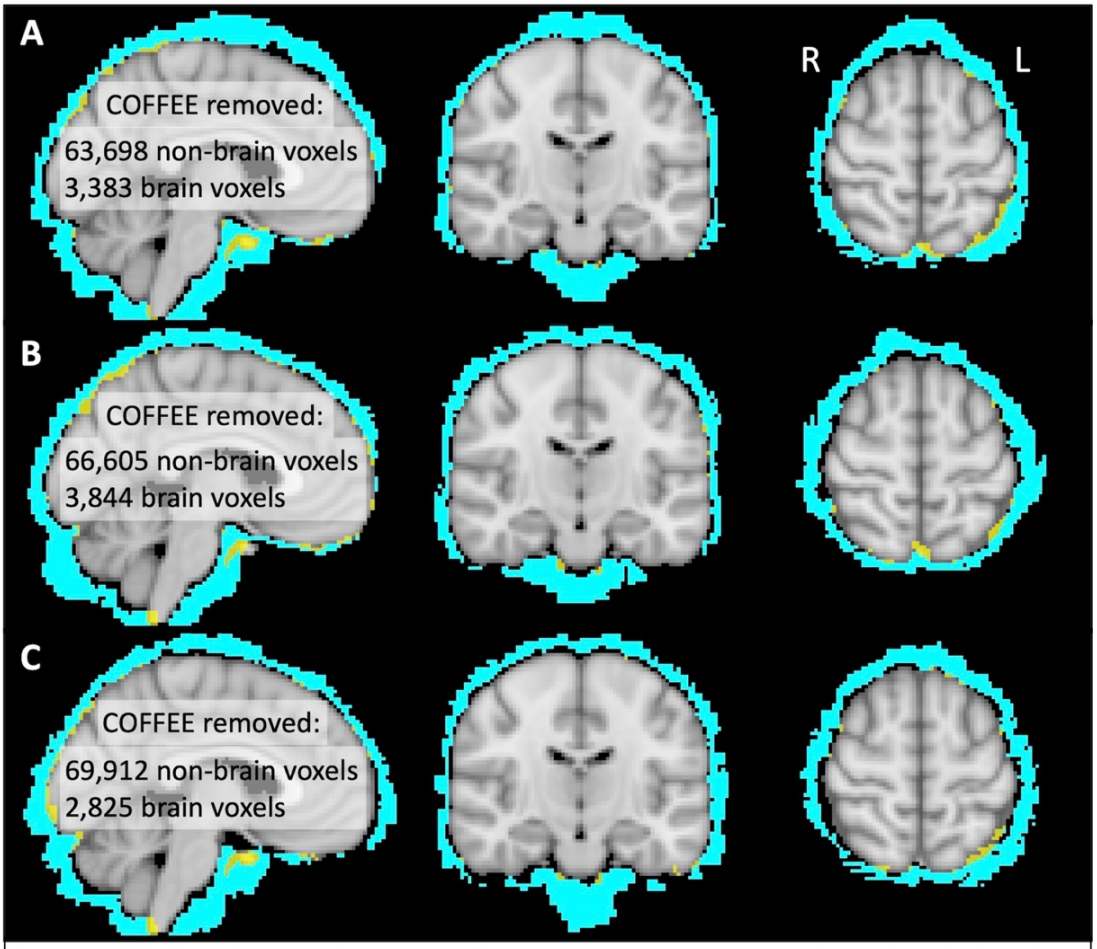

Figure 3. Brain extractions for 3 example runs, shown at x = -4, y = -15, z = 60. Text and colored area: voxels removed by COFFEE but not FSL. Brain voxels (yellow) and non-brain (blue) defined by MNI-152 standard brain mask. Participants chosen by change in signal magnitude across cortex, COFFEE vs. FSL-only. **A:** Participant with modest (25th percentile) change in signal magnitude. **B:** Participant with median change in signal magnitude. **C:** Participant with large (75th percentile) change in signal magnitude.

Differences between COFFEE and FSL-only analyses are broadly consistent across participants

This sample's participant group includes at least one major unmodeled source of variance: the distinction between typical adults and patients with peripheral nerve injury. This provides an advantage for methods testing, because it demonstrates COFFEE's robustness in the face of complex real-world data. However, we must rule out the possibility that our results were distorted by this sample choice. To address this, we first assessed whether participant group (patient vs. control) affected our measure of task-related BOLD signal (i.e. % BOLD signal change during Task > Rest). We performed a between-groups t-test to assess group differences in BOLD signal, and repeated this analysis separately for each ROI and analysis method (i.e. 82 comparisons). We found a between-group difference for only 1 comparison (Foot, COFFEE), and it was weak enough that would disappear under any amount of multiple comparison correction (t = 2.10, p = 0.0475). Therefore, peripheral nerve injury seemed to have no systematic effect on the measures tested here.

To further address the underlying concern about between-methods effects potentially being unevenly distributed across individuals, we examined the correlation between COFFEE signal intensities and FSL signal intensities (% BOLD signal change, Task > Rest) for each ROI. Correlations were high (Spearman's ρ 0.843 ± 0.090, range 0.533-0.954), with $\rho \geq 0.8$ in 33/42 (79%) of ROIs, indicating that the COFFEE:FSL relationship was broadly linear. Three sample ROIs are shown in **Figure 4**, illustrating the absence of within-ROI clusters of participants with different between-analysis relationships. Overall, the advantages of COFFEE analysis over FSL-only analysis did not seem to depend on subgroups or systematic

differences between participants.

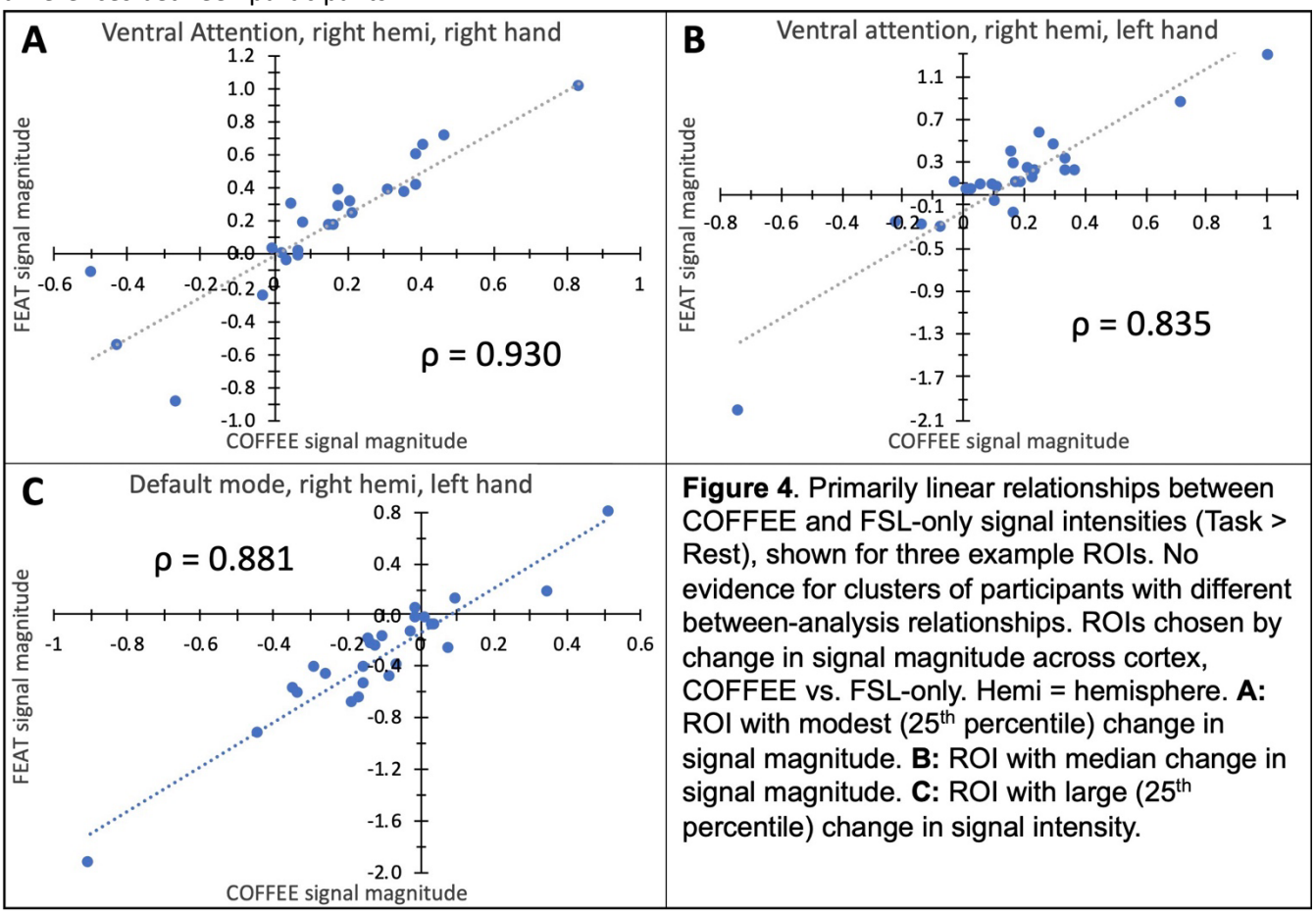

Figure 4. Primarily linear relationships between COFFEE and FSL-only signal intensities (Task > Rest), shown for three example ROIs. No evidence for clusters of participants with different between-analysis relationships. ROIs chosen by change in signal magnitude across cortex, COFFEE vs. FSL-only. Hemi = hemisphere. **A:** ROI with modest (25th percentile) change in signal magnitude. **B:** ROI with median change in signal magnitude. **C:** ROI with large (25th percentile) change in signal intensity.

DISCUSSION

We developed the new COFFEE analysis pipeline for HCP preprocessing of fMRI data for FSL analysis, and demonstrated its utility in analyzing task fMRI data. In our example dataset (precision drawing motor task, block design), COFFEE provided substantial advantages for data analysis. COFFEE reduced the number voxels incorrectly registered (to brain vs. non-brain), which led to changes in FEAT higher-level results consistent with increased signal in task-relevant areas. Critically, COFFEE pre-processing led to decreased inter-participant variability in BOLD signal magnitude across the brain.

Reduced inter-individual variability in functional analysis

COFFEE was designed as an alternative preprocessing pipeline for FSL FEAT analysis. We tested COFFEE against traditional FSL preprocessing using the same dataset and FEAT GLM for each analysis method. Our ultimate finding was that COFFEE led to significantly increased detection of BOLD signal in task-relevant areas of the brain (sensorimotor, parietal, and premotor areas contralateral to movement; **Figure 2**). This likely arises from reduced inter-individual variability in BOLD signal, which in turn may arise from increased precision in the process of registering individual data to standard space, as discussed further below.

COFFEE also led to a statistically significant increase in BOLD signal in deep structures that likely represent noise (e.g. white matter, ventricles). This may reflect amplification of real artifacts present in the data (e.g. movement-related). While amplification of artifacts is unhelpful, it is easily distinguishable from the expected cortical areas and networks involved in our task (precision drawing), and thus the heightened signal in deep structures should not present an obstacle to interpretation of task fMRI results.

Because we tested COFFEE with real data rather than simulated data, it is impossible to determine whether increased signal in task-relevant areas reflects a more-precise measurement of a strong signal, or a less-precise measurement of a

weak signal. However, the latter interpretation – artifactual inflation of BOLD signal strength – is incompatible with our finding that COFFEE's primary impact arose less from a change in individual magnitudes (**Table 2**) and more from decreased inter-individual differences (**Table 1**).

Overall, COFFEE's decreased inter-individual variability supports its value as an effective preprocessing pipeline for FSL FEAT analysis. Inter-individual variability is a longstanding challenge in fMRI research (Van Horn, Grafton and Miller, 2008; Dubois and Adolphs, 2016), because group averaging creates a tradeoff between statistical power and relevance to individual participants.

Increased precision in registration to standard space

COFFEE performs its registration and brain extraction using the established algorithms of Freesurfer software (e.g. Fischl and Dale, 2000; Fischl et al., 2002; Fischl et al., 2004), but the COFFEE pipeline allows a direct comparison to FSL's brain extraction. The COFFEE pipeline led to substantially smaller brain estimates, which were overwhelmingly caused by increased precision: for identification of non-brain voxels missed by FSL's BET tool, COFFEE/Freesurfer had a hit:false-positive ratio above 15:1. While this result is not "perfect," Freesurfer is known to be more robust than FSL for identifying and quantifying gray matter (Klauschen et al., 2009; Heinen et al., 2016). Improved precision in the brain extraction process may underpin most or all of COFFEE's improved fMRI results. Further improvement may be possible by adjusting the Freesurfer settings and manual editing of the Freesurfer output.

Use case of COFFEE

Numerous methods exist to tackle inter-individual variability in fMRI data (Napadow et al., 2008; Gordon et al., 2017; Michon et al., 2022), but not all methods are feasible for all research questions, groups, and protocols. The purpose of COFFEE is to provide a turnkey solution that integrates the established preprocessing methods from Human Connectome Project and Freesurfer, for use with with the popular, accessible fMRI analysis toolbox provided by FSL.

Limitations

Here we have tested COFFEE with one task fMRI dataset, using a sample of magnitude-based analysis methods. Many future directions present themselves: for example, COFFEE should be tested using similarity-based fMRI analyses (Nili et al., 2014), on task fMRI datasets with other tasks, on task-free fMRI, and on simulated data with known ground truth. Our initial test provides the foundation for more laboratories to join the effort of testing COFFEE's robustness across a broad range of analyses and data sets.

CONCLUSION

The COFFEE (Connectome Operations For FSL ExEcution) pipeline provides a turnkey alternative to FSL preprocessing based on established methods from the Human Connectome Project. In our sample dataset (real data from a fMRI hand motor task), COFFEE reduced inter-individual variability and increased signal in task-relevant regions. COFFEE is available online for further testing and use at: https://github.com/PhilipLab/COFFEE-pipeline

References

Andersson J, Jenkinson M, Smith S (2007) Non-linear registration, aka Spatial normalisation FMRIB technical report TR07JA2. FMRIB Analysis Group of the University of Oxford.

Boesen K, Rehm K, Schaper K, Stoltzner S, Woods R, Luders E, Rottenberg D (2004) Quantitative comparison of four brain extraction algorithms. Neuroimage 22:1255-1261.

Dubois J, Adolphs R (2016) Building a Science of Individual Differences from fMRI. Trends Cogn Sci 20:425-443.

Fischl B, Dale AM (2000) Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proceedings of the National Academy of Sciences 97:11050-11055.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, Van Der Kouwe A, Killiany R, Kennedy D, Klaveness S (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33:341-355.

Fischl B, Salat DH, van der Kouwe AJ, Makris N, Segonne F, Quinn BT, Dale AM (2004) Sequence-independent segmentation of magnetic resonance images. Neuroimage 23 Suppl 1:S69-84.

Fonov VS, Evans AC, McKinstry RC, Almli CR, Collins D (2009) Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage S102.

Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, Van Essen DC, Jenkinson M, Consortium WU-MH (2013) The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage 80:105-124.

Gordon EM, Laumann TO, Gilmore AW, Newbold DJ, Greene DJ, Berg JJ, Ortega M, Hoyt-Drazen C, Gratton C, Sun H, Hampton JM, Coalson RS, Nguyen AL, McDermott KB, Shimony JS, Snyder AZ, Schlaggar BL, Petersen SE, Nelson SM, Dosenbach NUF (2017) Precision Functional Mapping of Individual Human Brains. Neuron 95:791-807 e797.

Greve DN, Fischl B (2009) Accurate and robust brain image alignment using boundary-based registration. NeuroImage 48:63-72.

Heinen R, Bouvy WH, Mendrik AM, Viergever MA, Biessels GJ, de Bresser J (2016) Robustness of Automated Methods for Brain Volume Measurements across Different MRI Field Strengths. PLoS One 11:e0165719.

Jenkinson M (2018) BET User Guide. vol. 2023 Oxford UK: FMRIB ANalysis Group.

Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012) FSL. Neuroimage 62:782-790.

Jezzard P, Balaban RS (1995) Correction for geometric distortion in echo planar images from B0 field variations. Magnetic resonance in medicine 34:65-73.

Karimpoor M, Tam F, Strother SC, Fischer CE, Schweizer TA, Graham SJ (2015) A computerized tablet with visual feedback of hand position for functional magnetic resonance imaging. Frontiers in human neuroscience 9:150.

Klauschen F, Goldman A, Barra V, Meyer-Lindenberg A, Lundervold A (2009) Evaluation of automated brain MR image segmentation and volumetry methods. Hum Brain Mapp 30:1310-1327.

Klein A, Ghosh SS, Avants B, Yeo BT, Fischl B, Ardekani B, Gee JC, Mann JJ, Parsey RV (2010) Evaluation of volume-based and surface-based brain image registration methods. Neuroimage 51:214-220.

Michon KJ, Khammash D, Simmonite M, Hamlin AM, Polk TA (2022) Person-specific and precision neuroimaging: Current methods and future directions. Neuroimage 263:119589.

Mohapatra S, Gosai A, Schlaug G (2023) Brain extraction comparing segment anything model (sam) and fsl brain extraction tool. arXiv preprint arXiv:230404738.

Morgan W (1939) A test for the significance of the difference between the two variances in a sample from a normal bivariate population. Biometrika 31:13-19.

Napadow V, Dhond R, Conti G, Makris N, Brown EN, Barbieri R (2008) Brain correlates of autonomic modulation: combining heart rate variability with fMRI. Neuroimage 42:169-177.

Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N (2014) A toolbox for representational similarity analysis. PLoS computational biology 10:e1003553.

NITRC (2022) Resting State Networks cortical and subcortical atlas (rsnatlascxsubcx: Atlas RSN v1). vol. 2023: NeuroImaging Tools & Resources Collaboratory.

Philip BA, Frey SH (2014) Compensatory Changes Accompanying Chronic Forced Use of the Nondominant Hand by Unilateral Amputees. J Neurosci 34:3622-3631.

Philip BA, Li F, Hawkins-Chernof E, Swamidass V, Zwir I (2023) STEGA: Initial Validation of an iPad App to Measure Precision Drawing and Handwriting. American Journal of Occupational Therapy 77.

Quilis-Sancho J, Fernandez-Blazquez MA, Gomez-Ramirez J (2020) A comparative analysis of automated MRI brain segmentation in a large longitudinal dataset: Freesurfer vs. FSL. bioRxiv.

Siegel JS, Power JD, Dubis JW, Vogel AC, Church JA, Schlaggar BL, Petersen SE (2014) Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. Hum Brain Mapp 35:1981-1996.

Smith JC, Frey SH (2011) Use of independent component analysis to define regions of interest for fMRI studies. In: International Society for Magnetic Resonance in Medicine (ISMRM) Montreal, Canada.

Smith SM (2002) Fast robust automated brain extraction. Human brain mapping 17:143-155.

Smith SM, Brady JM (1997) SUSAN—a new approach to low level image processing. International journal of computer vision 23:45-78.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004) Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23 Suppl 1:S208-219.

Van Horn JD, Grafton ST, Miller MB (2008) Individual Variability in Brain Activity: A Nuisance or an Opportunity? Brain Imaging Behav 2:327-334.

Woolrich MW, Behrens TE, Beckmann CF, Jenkinson M, Smith SM (2004) Multilevel linear modelling for FMRI group analysis using Bayesian inference. Neuroimage 21:1732-1747.

Woolrich MW, Ripley BD, Brady M, Smith SM (2001) Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage 14:1370-1386.

Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zollei L, Polimeni JR, Fischl B, Liu H, Buckner RL (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol 106:1125-1165.

Zhang X, Feng Y, Chen W, Li X, Faria AV, Feng Q, Mori S (2019) Linear Registration of Brain MRI Using Knowledge-Based Multiple Intermediator Libraries. Front Neurosci 13:909.

Zhuang AH, Valentino DJ, Toga AW (2006) Skull-stripping magnetic resonance brain images using a model-based level set. Neuroimage 32:79-92.