SUPPLEMENTARY INFORMATION

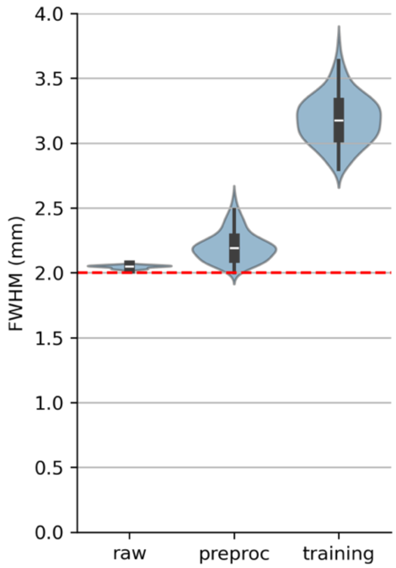
IMPACT OF FWHM RESOLUTION DIFFERENCES ON ICA-FIX DENOISING

Two resampling techniques were used during preprocessing of this dataset. The earliest approach used multi-shot resampling from the raw EPI image space to the final individual template space. This resulted in cumulative increases in effective spatial resolution following preprocessing steps such as head motion correction, distortion correction and registration to individual templates. The release version of the data employed a single-shot resampling from the raw EPI to final individual template space using functions provided by *fMRIPrep*. This approach mitigates increases in effective spatial resolution, estimated as a voxel FWHM, in the resampled data compared to multi-shot resampling.

Effective spatial resolution for raw BOLD EPI data, single-shot preprocessed (release) and multi-shot preprocessed data used for FIX training (training) was estimated from 4D BOLD EPI data using the AFNI 3dFWHMx command using spatial autocorrelation of residual noise. A combined FWHM value was estimated by geometric averaging of per-axis FWHM estimates for all EPI task runs.

All image acquisition parameters (TR, echo time, EPI echo spacing, etc) were identical between release and training datasets except for the resampling strategy which resulted in an estimated FWHM spatial resolutions of 2.2 mm and 3.2 mm respectively in the release and training data. The raw FWHM was close to the original sampled voxel resolution of 2 mm. The impact of using a FIX classifier trained with higher FWHM data on a lower FWHM (but otherwise identically preprocessed) dataset was determined by manual label classification into “retain” or “discard” groups by one rater (JMT) of a subset of 5 short tasks (5 minute trailer or movie repeats) and 5 long tasks (full movie segments) for each of the three subjects (15 task runs total). Component classifications followed the FIX convention, i.e. True indicates a discard component (noise or artifact) and False a retained component (signal of interest). Mean per-run accuracy statistics (accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value, NPV) and total unique explained variance (calculated by *FSL MELODIC)* for true positive (TP), true negative (TN), false positive (FP) and false negative (FN) component groups were calculated using the manual classifications as “ground truth”.

Estimated FWHM voxel resolution distributions, accuracy and explained variance statistics are summarized in Figure S1. Accuracy, sensitivity and specificity were high, between 94% and 96%, and PPV was close to 100% for both short and long tasks. NPV values were lower (67% for long tasks, 79% for long tasks). Since a True classification indicates a component that should be removed and manual labels were considered “ground truth”, this indicates that the FIX classifier trained on higher FWHM data is more conservative in discarding components than the human rater, identifying more components as signal of interest. Across all task runs, the FIX classifier mis-assigns as signal-of-interest approximately 4% to 9% of total unique explained variance. Both minimally preprocessed and FIX denoised BOLD data is provided in the OpenNeuro repository.



	Accuracy	Sensitivity	Specificity	PPV	NPV
Long	94%	94%	96%	99%	67%
Short	95%	96%	94%	99%	79%

	Mean Total Explained Variance			
	TP	TN	FP	FN
Long	85%	6%	0%	9%
Short	86%	9%	1%	4%

Figure S1: Estimated isotropic spatial resolution for un-preprocessed (raw) BOLD EPI data, single-shot preprocessed (preproc) data and multi-shot preprocessed FIX training data (training). Each distribution in represents FWHM estimates from 4D BOLD data over all tasks, sessions and subjects (141 images). The tables summarize accuracy and explained variance statistics between manual and FIX labeling of all independent components in subset of 18 task EPI runs spanning all three subjects. Reported values are mean percentages over all task runs.

Slab Coverage in MNI152 Space

Each subject's slab EPI data was registered to an individual AC-PC aligned template space. To provide a standard visualization of the extent of the slab coverage, a binary signal mask within the brain was calculated from the single-band reference EPI data for each task run (*nilearn* get_mask() function, lower threshold 1% of maximum intensity, cleaning selected), prior to averaging and mapping to the MNI152 2009c nonlinear asymmetric space using ANTs SyN non-linear registration (Figure S2

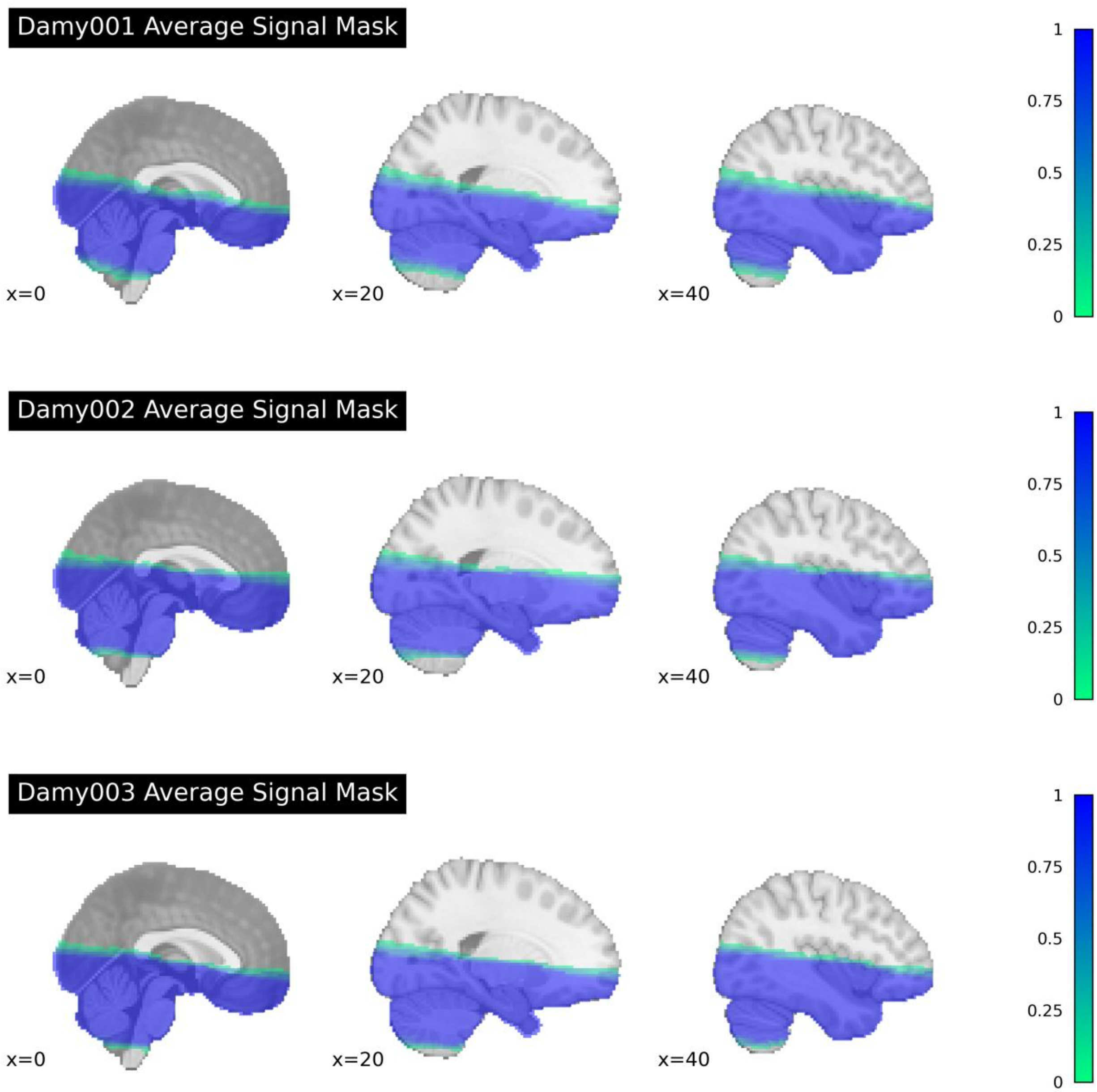


Figure S2: Sagittal sections through mean within-brain signal masks for each subject, mapped to the MNI152 2009c Nonlinear Asymmetric space (T1w brain template shown). Slab coverage in all subjects includes ventral temporal and frontal cortex, the medial temporal lobe and amygdala.