Identification, Characterization, and Mitigation of a Simultaneous Multi-Slice fMRI Signal Artifact: Multiband Artifact Regression in Simultaneous Slices (MARSS)

Philip N. Tubiolo[1,2], John C. Williams[1,2], & Jared X. Van Snellenberg[1,2,3]

[1]Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY 11794
[2]Department of Psychiatry and Behavioral Health, Renaissance School of Medicine at Stony Brook University, Stony Brook, NY 11794
[3]Department of Psychology, Stony Brook University, Stony Brook, NY 11794

CORRESPONDING AUTHOR: Jared X. Van Snellenberg
Email: Jared.VanSnellenberg@stonybrookmedicine.edu
Address: 101 Nicolls Rd, Health Sciences Center 10-087J, Stony Brook, NY 11794

Abstract

Simultaneous multi-slice (multiband) acceleration in fMRI has become widespread, but may be affected by novel forms of signal artifact. Here, we demonstrate a previously unreported artifact manifesting as a shared signal between simultaneously acquired slices in all resting-state and task-based multiband fMRI datasets we investigated, including publicly available consortium data. We propose Multiband Artifact Regression in Simultaneous Slices (MARSS), a regression-based detection and correction technique that successfully mitigates this shared signal in unprocessed data. We demonstrate that the signal isolated by MARSS correction is likely non-neural, appearing stronger in neurovasculature than grey matter. We show that MARSS correction leads to study-wide increases in signal-to-noise ratio, decreases in cortical coefficient of variation, and mitigation of systematic artefactual spatial patterns in participant-level task betas. Finally, we demonstrate that MARSS correction has substantive effects on second-level t-statistics in analyses of task-evoked activation. We recommend that investigators apply MARSS to all multiband fMRI datasets.

1. Introduction

1.1 Principles and Advantages of Multiband fMRI

Simultaneous multi-slice (multiband; MB) fMRI has become a prevalent fMRI acquisition acceleration technique that partially mitigates the limitations posed by standard single-slice (or single-band) EPI acquisition[1,2,3]. In MB fMRI acquisition, composite radiofrequency pulses comprising central frequencies (with bandwidths that determine slice thickness) that match the precession frequency of hydrogen nuclei in several axial sections along the field of view are transmitted in order to acquire signal from several slices in a single readout, and several sets of slices within a single TR; the number of slices acquired simultaneously is denoted by the MB acceleration factor. During readout, simultaneously acquired slices are “stacked” on top of each other, phase-shifted along the phase-encoding direction in order to reduce the amount of overlapping signal between slices. Slices are then mathematically unmixed using matrix inversion techniques that utilize radiofrequency coil sensitivity maps that are acquired prior to image acquisition. Overall, MB acceleration allows for substantive increases in spatiotemporal resolution that reduce signal-to-noise ratio but improve the spatial precision of neural signal estimates[1]. These improvements in data quality have led to its widespread adoption in both individual research laboratories and in large consortiums that have produced publicly available datasets, such as the Human Connectome Project (HCP)[4,5], UK Biobank study[6], and Adolescent Brain Cognitive Development (ABCD) Study[7].

1.2 Previously Discovered Multiband-Specific Artifacts and Mitigation Methods

As MB has gained adoption, several sources of contamination stemming from MB acquisition have been identified. Todd and associates identified a source of shared signal among voxels in simultaneously acquired slices that were overlapped during readout, coined “Interslice leakage”[8]. It was determined that this artifact could be substantially reduced by using

Split-Slice GRAPPA[9] image reconstruction over other techniques, such as Slice-GRAPPA[10]. McNabb and associates subsequently identified that interslice leakage artifacts are exacerbated by eye movement, even when Split-splice GRAPPA reconstruction is used, by having participants forcefully blink during resting-state MB fMRI acquisition[11].

Additionally, MB fMRI acquisition has been found to pose unique challenges in the estimation and removal of artifacts due to in-scanner participant motion, as the higher sampling rate enabled by these sequences facilitates the capture of additional nuisance signals[12,13,14,15,16,17,18,19], including respiratory motion and pseudomotion, or factitious head motion observed in the phase-encode direction as a consequence of lung expansion distorting the B_0 field[14].

1.3 Discovery of a Shared Artifact Signal in Simultaneously Acquired Slices

In working with resting-state and task-based MB fMRI data in our own laboratory, we observed periodic banding patterns in customized carpet plots[20] that display mean slice-wise timeseries (i.e., a simple slice x time plot of signal intensity following data preprocessing; see **Figure S1**). In investigating the source of this banding pattern, we discovered that raw, unprocessed MB fMRI data display elevated Pearson correlations between the average timeseries in simultaneously acquired slices over the correlations between non-simultaneously acquired slices (see **Figure 1**). We speculated that this indicated a shared signal source between simultaneously acquired slices that warranted further investigation, given that there is no biological basis for true neural signal fluctuations to be shared between arbitrarily located, spatially disparate, axial slices. Therefore, we believe that these elevated correlations represent a BOLD signal artifact that is specifically observable in MB accelerated fMRI.

Consequently, we detail the identification of this novel artifact signal across several resting-state and task-based MB fMRI datasets, including data from two large, publicly available, datasets[4,5,7], and MRI scanners across three major manufacturers: Siemens (Munich,

Germany), General Electric (GE; Boston, MA), and Philips (Andover, MA). We propose a regression-based technique to reduce the presence of the shared signal between simultaneously acquired slices, via estimation and removal of the shared signal in unprocessed MB fMRI data files, which we term Multiband Artifact Regression in Simultaneous Slices (MARSS). This estimation method is based on the idea that, provided major sources of shared signal (i.e., global signal and motion-related signal) are separately accounted for, averaging together a sufficient number of voxels containing the same artifact signal (i.e., all voxels from all simultaneously-acquired slices) will converge on the artifact signal itself, as true neural signals from disparate brain regions will "average out."

Furthermore, we show that this artifact signal persists when previously-identified interslice leakage artifacts are removed via regression, indicating that this novel artifact is distinct from interslice leakage. We also investigate the spatial, temporal, and spectral properties of the isolated artifact signal. Finally, we perform a working memory task-based analysis in both processed and unprocessed MB fMRI to highlight the effects of MB artifact correction on spatial patterns of estimates of task-evoked activation. We predicted that the artifact signal may originate from a combination of fluid motion and periodic mechanical noise during acquisition. Further, we hypothesized that noise reduction due to artifact removal would result in increased magnitude of spatial t-statistics in between-participants analyses of task-based fMRI, most likely as a result of reduced variance due to removal of spurious artifact signals.

2. Materials and methods

2.1 Functional Magnetic Resonance Imaging Datasets

To comprehensively evaluate the presence of the artifact signal across a range of scanner types and scan parameters, we utilized phantom and in vivo MB fMRI data from several sources, including samples from two publicly available datasets (HCP and ABCD), and two in-

house samples from the New York State Psychiatry Institute (NYSPI) and Stony Brook University (SBU). A comprehensive data summary including scan acquisition parameters can be found in **Table 1**. Demographic information for each dataset can be found in **Table S1**.

All procedures described in this study that involve the acquisition, curation, and analysis of data from human participants at NYSPI and SBU were approved by the New York State Psychiatric Institute Institutional Review Board and Stony Brook University Institutional Review Board, respectively. All participants provided written consent prior to their participation in study procedures.

2.1.1 Resting-State In-Vivo Datasets

We first employed unprocessed resting-state MB fMRI data acquired at the New York State Psychiatric Institute (NYSPI-GE; n = 63), with between 1 and 4 runs per participant (1 participant had 1 run, 4 had 2 runs, 9 had 3 runs, and 49 had 4 runs). An additional resting-state dataset was acquired at Stony Brook University (SBU-Siemens; n = 10), with 4 or 6 runs per participant (1 participant had 6 runs). One notable acquisition difference was that NYSPI-GE data was collected using a Nova Medical (Wilmington, MA) 32-channel head coil, while SBU-Siemens data used a Siemens 64-channel head-and-neck coil.

We also used resting-state data from a subset of the Human Connectome Project (HCP-Siemens)[4,5] 500 Subjects Release (n = 25) and the Adolescent Brain Cognitive Development Study (ABCD-Siemens, ABCD-GE, and ABCD-Philips; n = 35)[7]. All participants in the HCP-Siemens dataset had 4 runs of resting-state data, while participants in the ABCD datasets had between 1 and 5 runs of resting-state data.

To evaluate the relationship between artifact presence and MB factor in vivo, one SBU-Siemens participant and one NYSPI-GE participant underwent 1 resting-state run for each MB factor of 3, 5, and 7.

2.1.2 Task-Based Datasets

Up to 4 runs of task-based, MB fMRI data were acquired from participants in the SBU-Siemens dataset (n = 10; 1 participant had 3 runs) and a subset of the NYSPI-GE dataset (n = 58; 8 participants had 1 run, 4 had 2 runs, 14 had 3 runs, and 32 had 4 runs) using identical acquisition parameters to the resting-state data. Participants performed the self-ordered working memory task (SOT; see *Supplemental Materials*). Additionally, each HCP-Siemens participant (n = 25) performed 2 runs of working memory task-based data[21] with identical acquisition parameters to HCP-Siemens resting-state data.

2.1.3 Phantom Datasets

Phantom data was acquired to investigate both the non-neural origin of the artifact signal and the relationship between artifact presence and MB factor. Three runs of a Siemens D165 Spherical Phantom containing a Nickel sulfate solution were acquired at SBU, and three runs of an agar-filled Functional Biomedical Informatics Research Network (FBIRN) spherical stability phantom[22] were acquired at NYSPI.

2.2 Identification of Artifact Signal in Simultaneously Acquired Slices

A shared artifact signal across simultaneously acquired slices was identified separately in each run of MB fMRI data across all datasets. For each run, the average timeseries in each axial slice was obtained. In order to remove the impact of participant head motion on raw slice timeseries data, each slice timeseries was then taken to be the residuals of a multiple linear regression modeling the average slice timeseries as a function of six rigid-body head motion parameters, their squares, derivatives, and squared derivatives. The presence of a shared artifact signal across simultaneously acquired slices was quantified as the difference in Pearson correlation between the timeseries in simultaneously acquired slices and all non-simultaneously acquired slices, averaged across all slices in a volume. All Pearson correlations are Fisher Z-transformed prior to averaging or subtracting and transformed back to Pearson *r* before being

reported throughout this manuscript. This procedure was performed before and after artifact correction techniques were used in order to quantify the magnitude of artifact and success of artifact removal. Statistical significance of elevated correlations between simultaneously acquired slices in uncorrected data and the reduction in these correlations in MARSS-corrected data were determined via Wilcoxon signed-rank tests[23,24].

2.3 Multiband Artifact Regression in Simultaneous Slices (MARSS)

2.3.1 MARSS Estimation of Artifact Signal

We developed a regression-based method, MARSS, to estimate the magnitude of, and remove, the artifact signal shared across simultaneously acquired slices for each run of MB fMRI data. First, given a run with N_s slices and a MB acceleration factor MB_f, the number of simultaneously acquired slices is given by:

$$N_{MB} = \frac{N_s}{MB_f} \tag{1}$$

Given a slice j from this run, we define the set of slices acquired simultaneous to j as:

$$U_j = \{u|\, u = m \cdot c_j \wedge u \leq N_s\} \tag{2}$$

where N_s is the total number of slices in the volume, $m \in \mathbb{N}^+$, and:

$$c_j = \begin{Bmatrix} mod(j, N_{MB}), & mod(j, N_{MB}) \neq 0 \\ N_{MB}, & mod(j, N_{MB}) = 0 \end{Bmatrix} \tag{3}$$

Next, we define the average timeseries across all voxels $d_{i,u}$ in slices simultaneously acquired to slice j, but excluding slice j, as:

$$\boldsymbol{s_j} = \frac{1}{N_v(MB_f - 1)} \sum_{u \in U_j, u \neq j} \sum_{i=1}^{N_v} \boldsymbol{d_{i,u}} \tag{4}$$

where N_v is the number of voxels in each slice. We also define the average global signal for all slices that were not acquired simultaneously to slice j as:

$$\boldsymbol{g}_{U_j} = \frac{1}{N_v(N_s - MB_f)} \sum_{j=1, j \notin U_j}^{N_s} \sum_{i=1}^{N_v} \boldsymbol{d}_{i,j} \tag{5}$$

In order to estimate the artifact signal via regression, we construct a design matrix:

$$\boldsymbol{X}_j = \begin{bmatrix} \mathbf{1} & \boldsymbol{g}_{U_j} & \boldsymbol{M} \end{bmatrix} \tag{6}$$

where $\mathbf{1}$ is a vector of 1s with length equal to the run timeseries (i.e., number of volumes), and $\boldsymbol{M}$ is a matrix of nuisance parameters, including six rigid-body head motion parameters and their squares, derivatives, and squared derivatives. Next, we regress $\boldsymbol{X}_j$ onto $\boldsymbol{s}_j$:

$$\boldsymbol{b}_j = \left(\boldsymbol{X}_j{}^T \boldsymbol{X}_j\right)^{-1} \boldsymbol{X}_j{}^T \boldsymbol{s}_j \tag{7}$$

and estimate the artifact timeseries for slice j as the residuals of that regression, or:

$$\boldsymbol{a}_j = \boldsymbol{s}_j - \boldsymbol{X}_j \boldsymbol{b}_j \tag{8}$$

2.3.2 MARSS Removal of Artifact Signal

Given $\boldsymbol{a}_j$, the artifact signal estimate for slice j, we construct a design matrix to estimate the artifact signal contribution to each voxel i in slice j:

$$\boldsymbol{W}_j = [\mathbf{1} \quad \boldsymbol{a}_j \quad \boldsymbol{g}_{U_j} \quad \boldsymbol{M}] \tag{9}$$

and regress it onto the voxel-wise timeseries $\boldsymbol{d}_{i,j}$:

$$\boldsymbol{b}_{i,j} = \left(\boldsymbol{W}_j{}^T \boldsymbol{W}_j\right)^{-1} \boldsymbol{W}_j{}^T \boldsymbol{d}_{i,j} \tag{10}$$

The voxel-wise artifact signal estimate $\boldsymbol{a}_{i,j}$ can be defined as the product of the columns of $\boldsymbol{b}_{i,j}$ and $\boldsymbol{W}_j$ that correspond to the artifact estimate, $\boldsymbol{a}_j$ (the second column):

$$\boldsymbol{a}_{i,j} = \boldsymbol{b}_{i,j_{a_j}} * \boldsymbol{W}_{j_{a_j}} \tag{11}$$

Lastly, we obtain the corrected timeseries $\boldsymbol{d}'_{i,j}$ of voxel i in slice j as:

$$\boldsymbol{d}'_{i,j} = \boldsymbol{d}_{i,j} - \boldsymbol{a}_{i,j} \tag{12}$$

Following MARSS artifact removal, a new fMRI timeseries was saved as a NIFTI image, using the $\boldsymbol{d}'_{i,j}$ timeseries at each voxel in place of the original data. In addition, the spatial distribution of the artifact estimated by MARSS, $\boldsymbol{A}$, was determined by taking the mean (over timepoints, t) of the absolute value of the artifact signal at each voxel:

$$\boldsymbol{A}_{i,j} = \frac{\boldsymbol{1}}{\boldsymbol{T}} \sum_{\boldsymbol{t=1}}^{\boldsymbol{T}} \boldsymbol{abs}\left(\boldsymbol{a}_{i,j,t}\right) \tag{13}$$

2.3.3 Alternative Artifact Estimation and Removal Methods

In a modification to the technique described above, artifact signal was estimated using only background signal. This was performed by masking out the brain prior to artifact estimation using whole-brain masks generated by SynthStrip [25] included in FreeSurfer version 7.3.0[26]. Masks were dilated once prior to use. MARSS Artifact estimation and removal remained otherwise unchanged in this modification.

2.4 Characterization of Isolated Artifact Signal

Following MARSS artifact estimation and removal, we sought to investigate the spectral, temporal, and spatial properties of the isolated artifact signal to provide evidence for its potential source.

2.4.1 Frequency Spectral Analysis

For each run in a given dataset, the average, whole-image, isolated artifact timeseries was mean-centered. All runs were then concatenated to form a single vector, with zero-padding of the length of the timeseries included after each sample. Welch's power spectral density PSD;

[27] was calculated on the concatenated timeseries with a Hanning window applied to each individual timeseries and 50% overlap between samples. In order to quantify the average relationship between PSD and frequency in the roughly linear portion of the spectrum (with a lower frequency bound of -2 log[Hz]), a linear regression was performed on log-log transformed PSD that had been weighted by the inverse kernel density of the spectrum. The slope of this regression is useful in estimating colored noise characteristics of the artifact signal.

2.4.2 Autocorrelation

In order to investigate the potential periodicity of the isolated artifact signal, we calculated its autocorrelation as a function of lag, with each lag representing a temporal shift in the signal of one TR. In each dataset, the isolated artifact timeseries, obtained from Equation 11, above, was spatially averaged at each timepoint was averaged to produce a one-dimensional artifact timeseries. This timeseries was then mean-centered and linearly detrended before calculating the autocorrelation up to 20 lags (or a shift of 20 TRs) to sufficiently account for the period of physiological signals that may be components of the artifact signal. Autocorrelations were then averaged across runs in each dataset.

2.4.3 Interslice Leakage Artifact Analysis

Interslice leakage artifacts have been previously shown to contaminate estimates of neural signal in MB fMRI [11]. They arise as a consequence of the imperfect “unmixing” of signals in simultaneously acquired slices that are originally stacked and phase-shifted by a set fraction of the field of view (FOV) immediately preceding reconstruction. Interslice leakage artifacts, however, typically comprise a shared signal between two voxels that were overlaid during acquisition, rather than entire slices. We sought to demonstrate that the artifact detected and removed by MARSS is not a product of interslice leakage.

In resting-state MB fMRI data in the SBU-Siemens dataset, it was known that simultaneously acquired slices were phase-shifted by multiples of $\frac{FOV}{3}$ voxels (that is, multiples of 34 voxels) relative to each other prior to reconstruction. For each voxel, the contribution to its signal due to the voxels corresponding to the phase shift was estimated and removed via regression. The presence of the shared artifact in simultaneously acquired slices that remained after removing any potential interslice leakage was then determined using prior methods (see above, *2.2 Identification of Artifact Signal in Simultaneously Acquired Slices*).

2.4.4 Analysis of Artifact Spatial Distribution

We assessed the average spatial distribution of the isolated artifact signal in resting-state data in NYSPI-GE and SBU-Siemens. For each run, the MARSS estimate of the 3D spatial distribution of artifact was obtained using Equation 13, above. Each run's spatial distribution was warped and resampled into MNI152Nlin6Asym template space[28] by the single-step resampling procedure within the HCP Minimal Preprocessing Pipeline (OneStepResampling.sh)[29]. This was performed after fMRI preprocessing and utilized previously calculated gradient distortion corrections and affine transformations calculated during that process (see below, *2.6.3 fMRI Preprocessing*). In order to determine how the magnitude of signal fluctuations due to artifact was impacted by the underlying signal intensity (e.g., brain matter versus empty space), this process was repeated after rescaling each unprocessed average artifact distribution to reflect the percentage of artifact signal relative to mean signal intensity on a per-voxel basis.

2.4.5 Temporal Signal-to-Noise Ratio

Following artifact removal, we assessed the resulting change in whole-brain temporal signal to noise ratio (tSNR) in both unprocessed and preprocessed resting-state and task-based data. For each run, tSNR in voxel i of slice j was calculated as:

$$tSNR_{i,j} = \frac{\overline{\boldsymbol{s_{i,j}}}}{\sigma_{i,j}} \tag{14}$$

where $\overline{\boldsymbol{s_{ij}}}$ is the mean signal and σ_{ij} is the standard deviation of the signal[30,31]. In unprocessed data, voxel-wise tSNR estimates were averaged within dilated, whole-brain masks generated for each run by SynthStrip[25] and compared between corrected and uncorrected data. In preprocessed data, mean tSNR difference maps were generated separately for resting-state and task-based data in NYSPI-GE and SBU-Siemens. Significance of the elevation in tSNR across each dataset was determined via Wilcoxon signed-rank tests.

2.5 Task-Based Analysis

We sought to determine the effects of MB artifact correction on the spatial distribution of statistical maps resulting from task-based modeling (see *Supplemental Materials*). We first performed three within-participants analyses on unprocessed, working memory task-based MB fMRI data in NYSPI-GE (n=58), SBU-Siemens (n=10), and HCP-Siemens (n = 25) before and after MARSS artifact correction. We then performed separate between-participants analyses on preprocessed SBU-Siemens and NYSPI-GE data to highlight the effects of artifact correction on cortical, group-level t-statistics.

2.5.1 Within-Participants Task Modeling

Prior to modeling, each fMRI run was normalized to reflect percent signal change on a per-voxel basis, $\boldsymbol{n_{i,j}} = 100 * \boldsymbol{s_{i,j}}/\overline{\boldsymbol{s_{i,j}}}$, where $\boldsymbol{s_{i,j}}$ is the timeseries and $\overline{\boldsymbol{s_{i,j}}}$ is the mean signal in voxel i in slice j. Each task was modeled using a single-parameter canonical hemodynamic response function (HRF) in SPM12[32,33]. For the SOT, regressors of interest were those reflecting correct trials in the control task or each of the first seven steps of the SOT, modeled separately. For the visual n-back task, a block design was utilized to model all 0-back and 2-back conditions. Trial timings were extracted from event files ("EVs") distributed with HCP task data. Nuisance parameters included six motion parameters, their squares, derivatives, and

squared derivatives, as well as spike regressors to remove high motion volumes, without being convolved with the HRF. Spike regressors were identified using run-adaptive, generalized extreme value, low-pass filtered DVARS (GEV-DV) thresholds and chosen to identify approximately 3% of total volumes across each dataset[16]. The GEV-DV parameters used for each dataset were 28 in NYSPI-GE, 21 in SBU-Siemens, and 42 in HCP-Siemens.

Following model estimation, whole-brain mean squared residuals were obtained for each participant as a measure of within-participant model error and compared in uncorrected and MARSS-corrected data via Wilcoxon signed-rank test. In each participant, greyordinate-wise t-statistics were calculated for the control condition and each of the seven steps of the SOT as

$$\boldsymbol{t} = \frac{\beta}{\sqrt{\boldsymbol{resMS}}} \tag{15}$$

where β is the task beta for each task step and $resMS$ is the mean squared residual of the within-participants linear model. T-statistics were then averaged across steps for each participant and compared in corrected and uncorrected data via Wilcoxon signed-rank test.

2.5.2 Spatial Frequency Analysis of Unprocessed Task fMRI

We performed a spatial frequency analysis along the slice direction in each task dataset to evaluate the presence of periodic patterns of task betas corresponding to simultaneous slice spacing both before and after artifact removal. This was performed separately for task and control conditions in each task. One participant in SBU-Siemens and 7 participants in NYSPI-GE were excluded from this analysis (and subsequent analyses of preprocessed data; see below, *2.5.4 Between-Participants Analysis*) for not sufficiently performing all task conditions. Betas for each condition were averaged across the x- and y- dimensions, resulting in a single value per slice, and mean-centered. Frequency spectral analysis was then performed using previously described methods (see above, *2.4.1 Frequency Spectral Analysis*).

2.5.3 fMRI Preprocessing

Corrected and uncorrected data in SBU-Siemens and NYSPI-GE were preprocessed using the HCP Minimal Preprocessing Pipeline (MPP)[4,29] Version 4.2.0. Briefly, data underwent processing through the *PreFreeSurfer, FreeSurfer, PostFreeSurfer, fMRIVolume,* and *fMRISurface* pipelines. A critical step in the HCP preprocessing is its ribbon-constrained volume to surface mapping algorithm[29]. Notably, this step chooses to exclude voxels with locally high coefficients of variation (COV) from the mapping process. We quantified the mean, voxel-wise change in COV before and after MARSS artifact correction within the cortical ribbon isolated for each participant during anatomical volume preprocessing in order to determine whether MARSS increases the probability that cortical voxels will be included in volume to surface mapping.

Using the preprocessed outputs in CIFTI format (*dtseries.nii*), cortical surface GIFTI files and subcortical NIFTI files were generated, smoothed with a 4 mm full width at half maximum Gaussian kernel, and normalized to reflect percent signal change on a per-greyordinate basis.

2.5.4 Between-Participants Analysis

First level modeling was performed for the SOT as previously described (see above, *2.5.1 Within-Participants Task Modeling*), with the exception of using cortical surfaces and subcortical volumes rather than pure volumetric data. Task contrasts for each participant were calculated as the average beta estimate from the first seven steps of the SOT minus the control condition.

Contrasts from corrected and uncorrected data were analyzed separately in between-participants analyses using Permutation Analysis of Linear Models (PALM)[34,35,36,37] alpha119. Analyses in PALM were performed using 1024 and 20000 sign flips in SBU-Siemens and NYSPI-GE, respectively, with built-in family-wise error rate (FWER) correction[38] both with and without threshold-free cluster enhancement (TFCE)[39]. T-statistic result images with and without TFCE were subtracted to obtain differences in t-statistics resulting from artifact removal. This subtraction was also repeated after taking the absolute value of t-statistics in corrected and uncorrected data where uncorrected data had negative t-statistics. This yielded difference maps

reflecting the magnitude of the change in t-statistics (e.g, so that a decrease in signal in an area showing deactivation would appear as an *increase* in t-statistic magnitude), rather just the direction of the change.

In addition to this, we quantified the change in t-statistic within an a priori region of interest (ROI), the left dorsolateral prefrontal cortex (lDLPFC). The lDLPFC ROI (see **Figure S2**) was derived from regions of significant activation in response to SOT task demands from previous work[40,41]. The ROI was first resampled from volume to surface space, then dilated to fill holes, and finally eroded by 4mm. This analysis was performed on between-participant cortical t-statistic maps generated both with and without the use of TFCE.

3. Results

3.1 Artifact Identification and Removal Across Multiple Datasets

The presence of elevated Pearson correlations in simultaneously acquired slices over non-simultaneously acquired slices across multiple MB factors in vivo and in phantom data can be seen in correlation matrices in **Figure 1** (see **Table S2**)**.** In each matrix, the artifact manifests as a diagonal striped pattern that varies with MB factor. Additionally, striped patterns of elevated Pearson correlations were observed in all in vivo datasets explored regardless of scanner or task type (see **Figures S3, S4, Table S3**). Following artifact removal with MARSS, correlations in simultaneously acquired slices were significantly reduced to values closer to the mean correlation between all non-simultaneously acquired slices (see **Table S4** for statistics).

Results of alternative artifact removal procedures tested in SBU-Siemens RSFC data are shown in **Figures S5** and **S6**. When artifact magnitude was estimated using background alone, the difference in correlation between simultaneously acquired slices and non-simultaneously acquired slices was reduced from 0.463 to 0.361 ($W(42) = 903$, $p = 1.648 \times 10^{-8}$; see **Figure S5**). This decrease in correlation, however, was much less pronounced than that produced by MARSS artifact correction (see **Table S3**). Additionally, regression-based interslice

leakage produced a very slight increase in the elevation of correlations between simultaneously acquired over non-simultaneously acquired slices, from 0.463 to 0.466 ($W(42) = 638$, $p = 0.020$; see **Figure S6**). This indicates that no level of artifact correction was achieved by this method.

3.2 Characterization of Artifact Signal

3.2.1 Spatial Artifact Distribution

The spatial distributions of the average MARSS-estimated artifact signal across resting-state runs in NYSPI-GE and SBU-Siemens are shown in **Figure 2.** Images are shown on both the scale of the raw EPI data and as a percentage of the mean voxel-wise EPI signal intensity. A video depicting a 360-degree view of the rendered NYSPI-GE spatial distribution is included in the *Supplementary Information*. In both datasets, artifact signal appears strongest in vascular structures, notably the superior sagittal sinus, inferior sagittal sinuses, transverse sinuses, middle cerebral arteries, Circle of Willis, and superficial temporal arteries. Artifact signal is also notably strong in the eyes. Further, MARSS removes a larger percentage of voxel signal from both the image background and vasculature than from grey and white matter.

The spatial distribution of the MARSS-estimated artifact signal in a single phantom run from SBU-Siemens acquired with a MB factor of 6 is shown in **Figure S7**. Artifact signal is strongest near the posterior edge of the phantom, where it contacted the scanner table.

3.2.2 Spectral and Autocorrelative Properties

PSD estimates and autocorrelations of the MARSS-estimated artifact signal timeseries across all in vivo, task-based fMRI datasets are shown in **Figure 3**. All PSD estimates were linearly fit with slopes between -0.716 and -0.934 (see top row of **Figure 3**). Autocorrelations in each dataset show varying degrees of long-range dependence from approximately 5 to 15 lags (see bottom row of **Figure 3**). These spectral and autocorrelative characteristics roughly correspond to the properties expected of 1/f colored noise.

Additional PSD estimates from all in vivo, resting-state fMRI datasets are shown in **Figure S8**. Further, the PSD estimate from phantom data acquired with a MB factor of 6 at SBU is shown in **Figure S9**. Autocorrelation of the artifact timeseries in in-vivo, resting-state data as a function of lag is shown in **Figure S10.**

3.2.3 Temporal Signal-to-Noise Ratio

Across all unprocessed, in vivo datasets, 100% of runs showed significant increases in mean, whole-brain tSNR (all $p<0.001$; see **Table S5**) following MARSS artifact removal. In preprocessed, task-based data, 69.91% of voxels in NYSPI-GE and 95.16% of voxels in SBU-Siemens experienced mean increases in tSNR (see top row of **Figure 4**). Similarly, in preprocessed, resting-state fMRI data, 77.79% of voxels in NYSPI-GE and 97.45% of voxels in SBU-Siemens experienced mean increases in tSNR across participants (see **Figure S11**).

3.3 Within-Participants Analysis of Unprocessed Task fMRI

Single-participant examples of the difference in within-participant betas in unprocessed data from the task-on condition before and after MARSS correction are shown in the top row of **Figure 5.** In all datasets, the spatial pattern of beta differences is systematic and occurs at both the fundamental spatial frequency of simultaneously acquired slice spacing and its corresponding harmonic spatial frequencies. As shown in the bottom two rows of **Figure 5,** there is a reduction in PSD at these spatial frequencies in all datasets following MARSS.

3.4 Within- and Between-Participants Analysis of Preprocessed Task fMRI

3.4.1 Artifact Correction Effects on Volume to Surface Mapping

Decreases in cortical coefficient of variation were seen in 94.62% of task-based runs in NYSPI-GE ($W(186) = 534$, $p = 1.247 \times 10^{-28}$) and 100% of task-based runs in SBU-Siemens ($W(39) = 0$, $p = 5.255 \times 10^{-8}$) following MARSS, with 80.77% of voxels in NYSPI-GE and 97.0%

of voxels in SBU-Siemens experiencing mean decreases across participants (see bottom row of **Figure 4**).

3.4.2 Artifact Correction Effects on Within-Participants Modeling

Following MARSS, we observe a mean decrease in within-participants mean squared residuals of -0.045 (-2.16%) in NYSPI-GE ($W(52) = 0$, $p = 3.504 \times 10^{-10}$) and -0.062 (-5.3%) in SBU-Siemens ($W(10) = 0$, $p = 0.002$). Conversely, mean single-participant t-statistics increase by 0.003 in NYSPI-GE ($W(52) = 1002$, $p = 0.004$) and by 0.010 in SBU-Siemens ($W(10) = 52$, $p = 0.010$).

3.4.3 Between-Participants Analysis

Figure 6 shows the difference in between-participants (second level) t-statistics obtained without TFCE before and after MARSS in NYSPI-GE and SBU-Siemens task-based data. Striped patterns in t-statistic differences can be observed in both datasets, although these are masked by non-linear distortions in the slice pattern of the artifact, which also differ across participants, due to preprocessing. T-statistic maps reflecting the change in *magnitude* of t-statistics can be seen in **Figure S12** (that is, if a negative t-statistic became more negative, it would be shown as a decrease in Figure 6 but as an increase in Figure S12). In addition, the difference in t-statistics, as well as their magnitude difference, in between-participants analyses performed with TFCE are shown in **Figure S13** and **Figure S14**, respectively. For all between-participants analyses, the deciles of absolute value of t-statistic change were calculated and are shown in **Table 2**. In NYSPI-GE, 50% of t-statistics changed by a magnitude of 0.171 or more (11.243 with TFCE), while in SBU-Siemens 50% changed by a magnitude of 0.169 or more (3.041 with TFCE). Within the left DLPFC ROI (see **Figure S2**), the mean t-statistic increased from 5.325 to 5.376 (+0.958%) without TFCE and from 1,172.00 to 1,240.00 (+5.80%) with

TFCE in NYSPI-GE. In SBU-Siemens, the mean left DLPFC t-statistic increased from 2.440 to 2.720 (+11.44%) without TFCE and from 68.231 to 83.512 (+22.40%) with TFCE.

4. Discussion

The results presented here demonstrate the successful mitigation of a shared artifactual signal between simultaneously acquired slices via MARSS correction. We have provided evidence that signal removed via MARSS is likely non-neural in nature. MARSS-corrected data exhibits higher overall tSNR and lower cortical coefficients of variation than uncorrected data. MARSS results in decreased participant-level task model error and removes spatial patterns of task betas at spatial frequencies corresponding to the distance between simultaneously acquired slices. Between participants, MARSS causes substantive changes in group-level t-statistics that tend to follow striped patterns associated with slice acquisition, even after spatial normalization is performed. All of the above results indicate significant improvements in data quality following MARSS correction.

4.1 MARSS Correction Reduces Simultaneous Slice Correlations

We have shown the presence of a shared artifactual signal between simultaneously acquired slices of multiband fMRI data acquired across six datasets, including in two widely used, publicly available datasets[5,7]. This signal was originally identified as elevated correlations between time series in simultaneously acquired slices relative to those with non-simultaneously acquired slices, in both phantom and *in vivo* data, with elevations as high as a difference in Pearson *r* correlation coefficients of 0.3 or 0.4 in many datasets. Further, the spatial pattern of elevated correlations follows a clear, systematic relationship with the MB factor used during acquisition. As there is no physiological justification for BOLD signal in arbitrary axial slices to be more related in simultaneously acquired slices than in any other given pair of slices (including neighboring slices), we developed a regression-based detection and correction

technique, MARSS, that estimates the shared signal across simultaneously-acquired slices in unprocessed MB-fMRI data and removes the estimated signal from each voxel in an fMRI run. This technique successfully reduces the magnitude of the difference in simultaneous slice correlations between simultaneously acquired and non-simultaneously-acquired slices. Although MARSS also reduced simultaneous slice correlations to levels slightly below that of the average correlation between non-simultaneously acquired slices, it does so to a much lesser degree than the magnitude of the original elevated correlations. As shown in **Table S2**, the maximum negative difference in correlation between simultaneously acquired slices and non-simultaneously acquired slices induced by artifact correction in any *in vivo* dataset is -0.101, as opposed to a correlation of 0.434 in uncorrected data. Although we collected only minimal data at lower MB factors (e.g., 3 or 4), the data in **Table S1** raise the possibility that higher MB factors are associated with smaller magnitude reductions in simultaneous-slice correlations after MARSS correction. This is consistent with the fact that lower MB factors should, a priori, produce poorer estimates of the shared signal across slices due to having fewer slices to average together to arrive at an estimate of the artifact.

4.2 MARSS Correction Improves SNR and Removes Non-Neural Signal

Critically, we also show that MB artifact correction results in the removal of image noise rather than true neural signal. First, an increase in mean, whole-brain tSNR in 100% of unprocessed runs across all datasets after MARSS correction suggests improved data quality. In preprocessed data, the largest mean increases in tSNR occurred in cortical regions, while many decreases occurred at tissue boundaries where BOLD signal most likely contained greater contamination from fluid motion or vascular signal (see top row of **Figure 4**, **Figure S11**). Although tSNR carries no information on activation strength in task-based fMRI data, it is a useful measure of signal stability in studies of resting-state fMRI[42].

One major indication that MARSS predominantly removes non-neural signal lies in the spatial distribution of the estimated artifact in **Figure 2**, where the largest amount of signal is removed from large vasculature, and a higher percentage of mean signal is removed from the background of the image. The significant vascular involvement suggests that artifact correction is isolating signal related to vasomotion or pulsatile blood flow whose contribution elevates signal correlations between simultaneously acquired slices. It is plausible that this signal originates from systemic low-frequency oscillations (sLFOs), which are considered non-neural in nature and account for a significant portion of the variance in signal in the range of approximately 0.01-0.15 Hz[43,44]. Notably, isolated artifact timeseries show autocorrelations up to lags that approximately correspond with this frequency range (see **Figures 3** and **S10**).

The frequency band occupied by sLFOS is the same as that which is occupied by fluctuations in neural activity that are the basis of resting-state functional connectivity. Thus, traditional spectral filtering techniques cannot preferentially remove these signals without compromising true signal from neuronal fluctuations. Tong and colleagues have demonstrated that sLFOs can confound estimates of resting-state functional connectivity, going as far as mimicking resting-state networks isolated via independent components analysis (ICA)[45]. Additionally, they have identified major vascular fMRI signal components that appear in the spatial distribution of the artifact signal, such as the superior sagittal sinus, to be highly correlated with simultaneously-acquired near infrared spectroscopy estimates of blood flow[46].

Slightly complicating this explanation of the artifact source is its presence in phantom data. The artifact signal intensity in phantom data is strongest near the point of contact with the scanner table (see **Figure S7**). Due to this being a liquid-filled phantom, it is possible that scanner table vibrations caused periodic fluid motion that resembles pulsatile blood flow. While inconclusive at this time, the possibility that MARSS artifact estimation is partially or completely identifying signal originating from the vasculature bears further investigation.

4.3 MARSS Correction Reduces Cortical Coefficient of Variation

The removal of noise from unprocessed, volumetric data due to artifact correction is beneficial to volume-to-surface mapping in the HCP MPP[29]. For context, the mapping algorithm chooses to exclude voxels with coefficients of variation that are 0.5 standard deviations above the mean local coefficient of variation in a 5mm Gaussian neighborhood. Both the NYSPI-GE and SBU-Siemens datasets experience decreases in coefficient of variation in 97.00% and 80.77% of voxels, respectively. This nearly global decrease suggests an increase in probability of most cortical voxels to be included in volume to surface mapping. Notably, several runs of NYSPI-GE data were discovered to have striped patterns of high coefficients of variation (see **Figure S15**) that is mitigated after MARSS artifact correction. Upon inspection, data from these runs show no discernable quality control concerns, suggesting that multiband acquisition can sometimes cause spurious, yet systematic noise patterns that could potentially surface timeseries extraction in otherwise usable data.

4.4 MARSS Correction Improves Spatial Statistics Within and Between Participants

Critically, we have shown that MARSS data correction for this shared signal artifact has substantive implications for spatial statistics in task-based analyses, both within and between participants. When comparing within-participants analyses in unprocessed data prior to and following artifact correction, we observe a striping pattern in task beta differences across all task-based datasets (see top row of **Figure 5**). We analyzed the spatial frequency of these patterns along the slice direction and demonstrated that systematic patterns in task betas occur at the fundamental frequency and harmonic frequencies of the simultaneously acquired slice pattern in uncorrected data (see bottom two rows of **Figure 5**), and that the spectral power at these spatial frequencies is reduced by MARSS. While there is no ground truth expectation as to the magnitude of task betas, it is reasonable to conclude that true estimates of task-evoked

neuronal activation should not exhibit a spatial frequency dependent on the spacing between simultaneously-acquired slices. The fact that MARSS correction reduces the spectral power of these patterns indicates that this spatial frequency is being removed from the data by our method, rather than being added.

In within-participants analyses of preprocessed, task-based data, we observe statistically significant decreases in mean-squared residuals, indicating lower model error and suggesting that task-irrelevant, and potentially non-neural, signals are being removed by MARSS. In addition, we observe increases in single-participant t-statistics, showing that the reduction of noise due to MARSS correction outweighs any potential decrease in the magnitude of task-evoked activation. At the group level, we observe striping patterns in the difference in t-statistics between analyses performed before and after artifact correction (see **Figure 6**). Spatial frequency analyses demonstrated that the systematic pattern is present first in unprocessed, uncorrected data and persists in the data even after spatial normalization and volume-to-surface resampling distort the slice acquisition pattern to different degrees in each participant. MARSS correction led to substantial changes in between-participants t-statistics, with 10% of greyordinates changing by at least 0.458 and 0.490 in NYSPI-GE and SBU-Siemens, respectively (see **Table 2**).

Since there is no ground truth expectation for estimations of whole-brain, task-evoked activation, we then constrained our analysis to within the left DLPFC, which is known to show robust activation in response to working memory task demands[47,48,49]. Within the chosen ROI, which was derived from previous group-level analysis of the SOT in a separate dataset[40,41], t-statistics increase in both NYSPI-GE and SBU-Siemens with and without the use of TFCE. The larger percentage increase in t-statistic that results from using TFCE suggests that removal of noise due to artifact correction enhances the spatial homogeneity of patterns of true task-

evoked activation, leading to clearer identification and amplification of significant regions by the TFCE algorithm.

4.5 Differences in Results Between Datasets

One theme in the results presented here is the differences in results between NYSPI-GE and SBU-Siemens data, where NYSPI-GE appears to show less improvement in between-participants analyses following artifact correction. There are several hardware differences between these datasets that may influence this, such as scanner manufacturer and coil configuration. Further, the slightly longer TR used in NYSPI-GE data could result in diminished temporal resolution and less adequate sampling of artifactual periodic signals identified by artifact estimation. Despite these differences, the spatial distribution of isolated artifact signal (see **Figure 2**) in NYSPI-GE is highly similar to that of SBU-Siemens. Additionally, artifact correction succeeds in significantly reducing or removing artificially elevated PSD of task betas at fundamental and harmonic frequencies related to the simultaneous slice acquisition in both datasets, indicating improvement in task modeling regardless of the resulting change in between-participants t-statistics.

4.6 Conclusion

In summary, we have demonstrated the presence and successful correction of an artifactual, non-neural signal that is shared between simultaneously acquired slices in multiple datasets. Removal of this artifact with the MARSS method developed here results in substantive improvements in data quality, including nearly global increases in tSNR and improvements to the extraction of the cortical surface during preprocessing. The artifact appears strongest in the neurovasculature, but has clearly demonstrable impacts throughout grey matter voxels and greyordinates. Critically, artifact correction leads to a reduction of systematic spatial patterns in estimates of task-evoked activation related to slice acquisition patterns, both within and between

participants, which ultimately leads to large changes in group-level t-statistics throughout the brain. We expect that application of our artifact correction method will lead to an improvement in both the quality and reproducibility of all multiband fMRI datasets, and recommend that investigators implement this technique prior to functional data preprocessing.

5. Acknowledgements and Disclosures

Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health (NIH) (Award Nos. R01MH120293 and K01MH107763 [to JXVS], and F30MH122136 [to JCW]). JCW was also supported by the Stony Brook University Medical Scientist Training Program (Award No. T32GM008444; Principal Investigator: Dr. Michael A. Frohman). PNT was supported by a Stony Brook University Department of Biomedical Engineering Graduate Assistance in Areas of National Need Fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

The authors would like to thank Jack Grinband for comments and suggestions on preliminary findings in this work. We thank Yash Patel, Jaeyop Jeong, Sam R. Luceno, and Sameera Abeykoon for their contributions to fMRI data preprocessing and Mahika Gupta for assistance with preparing MARSS code for public release. We would also like to thank Guarav Patel and Matthew Riddle at the New York State Psychiatric Institute, and Turhan Canli and Kim Burke at Stony Brook University for their assistance in acquiring phantom data. Finally we would like to acknowledge the substantial computing resources and technical assistance provided by Stony Brook Medicine Research Computing, with substantial support from Allen Zawada and James Xikis, as well as Stony Brook Research Computing and Cyberinfrastructure and the Institute for Advanced Computational Science at Stony Brook University for access to the high-performance SeaWulf computing system (National Science Foundation Award Nos. 1531492

and 2215987, and matching funds from the Empire State Development's Division of Science, Technology and Innovation program contract C210148), with notable support from Fırat Coşkun, Daniel Wood, and David Carlson.

The authors report no biomedical financial interests or potential conflicts of interest.

6. Data Availability

Data used in this manuscript can be made available by the authors upon request. Publicly available HCP data can be obtained from ConnectomeDB (https://db.humanconnectome.org/)[50]. ABCD study data can be obtained from the National Institute of Mental Health National Data Archive (NIMH NDA).

7. Code Availability

MARSS is available as an open-source MATLAB software package on GitHub (https://github.com/CNaP-Lab/MARSS) and the MathWorks File Exchange (https://www.mathworks.com/matlabcentral/fileexchange/156782-multiband-artifact-regression-in-simultaneous-slices-marss). Code to replicate all analyses in this manuscript can be made available by the authors upon request.

References

1. Feinberg DA*, et al.* Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *PLoS One* **5**, e15710 (2010).

2. Moeller S*, et al.* Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn Reson Med* **63**, 1144-1153 (2010).

3. Xu J*, et al.* Evaluation of slice accelerations using multiband echo planar imaging at 3 T. *Neuroimage* **83**, 991-1001 (2013).

4. Ugurbil K*, et al.* Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *Neuroimage* **80**, 80-104 (2013).

5. Van Essen DC*, et al.* The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, 62-79 (2013).

6. Miller KL*, et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* **19**, 1523-1536 (2016).

7. Casey BJ*, et al.* The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev Cogn Neurosci* **32**, 43-54 (2018).

8. Todd N, Moeller S, Auerbach EJ, Yacoub E, Flandin G, Weiskopf N. Evaluation of 2D multiband EPI imaging for high-resolution, whole-brain, task-based fMRI studies at 3T: Sensitivity and slice leakage artifacts. *Neuroimage* **124**, 32-42 (2016).

9. Cauley SF, Polimeni JR, Bhat H, Wald LL, Setsompop K. Interslice leakage artifact reduction technique for simultaneous multislice acquisitions. *Magnetic Resonance in Medicine* **72**, 93-102 (2014).

10. Griswold MA*, et al.* Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine* **47**, 1202-1210 (2002).

11. McNabb CB, Lindner M, Shen S, Burgess LG, Murayama K, Johnstone T. Inter-slice leakage and intra-slice aliasing in simultaneous multi-slice echo-planar images. *Brain Struct Funct* **225**, 1153-1158 (2020).

12. Burgess GC*, et al.* Evaluation of Denoising Strategies to Address Motion-Correlated Artifacts in Resting-State Functional Magnetic Resonance Imaging Data from the Human Connectome Project. *Brain Connectivity* **6**, 669-680 (2016).

13. Glasser MF*, et al.* Using temporal ICA to selectively remove global noise while preserving global signal in functional MRI data. *NeuroImage* **181**, 692-717 (2018).

14. Power JD, Lynch CJ, Silver BM, Dubin MJ, Martin A, Jones RM. Distinctions among real and apparent respiratory motions in human fMRI data. *NeuroImage* **201**, (2019).

15. Fair DA*, et al.* Correction of respiratory artifacts in MRI head motion estimates. *NeuroImage* **208**, (2020).

16. Williams JC, Tubiolo PN, Luceno JR, Van Snellenberg JX. Advancing motion denoising of multiband resting-state functional connectivity fMRI data. *Neuroimage* **249**, 118907 (2022).

17. Power JD, Lynch CJ, Silver BM, Dubin MJ, Martin A, Jones RM. Distinctions among real and apparent respiratory motions in human fMRI data. *Neuroimage* **201**, 116041 (2019).

18. Fair DA*, et al.* Correction of respiratory artifacts in MRI head motion estimates. *Neuroimage* **208**, 116400 (2020).

19. Burgess GC*, et al.* Evaluation of Denoising Strategies to Address Motion-Correlated Artifacts in Resting-State Functional Magnetic Resonance Imaging Data from the Human Connectome Project. *Brain Connect* **6**, 669-680 (2016).

20. Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage* **84**, 320-341 (2014).

21. Barch DM*, et al.* Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage* **80**, 169-189 (2013).

22. Friedman L, Glover GH. Report on a multicenter fMRI quality assurance protocol. *Journal of Magnetic Resonance Imaging* **23**, 827-839 (2006).

23. Gibbons JD, Chakraborti S. *Nonparametric statistical inference*, 6th edition. edn. CRC Press (2021).

24. Hollander M, Wolfe DA, Chicken E. *Nonparametric statistical methods*, Third edition / edn. John Wiley & Sons, Inc. (2014).

25. Hoopes A, Mora JS, Dalca AV, Fischl B, Hoffmann M. SynthStrip: skull-stripping for any brain image. *NeuroImage* **260**, (2022).

26. Fischl B. FreeSurfer. *Neuroimage* **62**, 774-781 (2012).

27. Welch P. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics* **15**, 70-73 (1967).

28. Grabner G, Janke AL, Budge MM, Smith D, Pruessner J, Collins DL. Symmetric Atlasing and Model Based Segmentation: An Application to the Hippocampus in Older Adults. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*) (2006).

29. Glasser MF*, et al.* The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80**, 105-124 (2013).

30. Murphy K, Bodurka J, Bandettini PA. How long to scan? The relationship between fMRI temporal signal to noise ratio and necessary scan duration. *NeuroImage* **34**, 565-574 (2007).

31. Welvaert M, Rosseel Y. On the Definition of Signal-To-Noise Ratio and Contrast-To-Noise Ratio for fMRI Data. *PLoS ONE* **8**, (2013).

32. Friston KJ. *Statistical parametric mapping : the analysis of funtional brain images*, 1st edn. Elsevier/Academic Press (2007).

33. Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping* **2**, 189-210 (2004).

34. Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. *NeuroImage* **92**, 381-397 (2014).

35. Winkler AM, Ridgway GR, Douaud G, Nichols TE, Smith SM. Faster permutation inference in brain imaging. *NeuroImage* **141**, 502-516 (2016).

36. Winkler AM, Webster MA, Brooks JC, Tracey I, Smith SM, Nichols TE. Non-parametric combination and related permutation tests for neuroimaging. *Human Brain Mapping* **37**, 1486-1511 (2016).

37. Winkler AM, Webster MA, Vidaurre D, Nichols TE, Smith SM. Multi-level block permutation. *Neuroimage* **123**, 253-268 (2015).

38. Holmes AP, Blair RC, Watson JDG, Ford I. Nonparametric Analysis of Statistic Images from Functional Mapping Experiments. *Journal of Cerebral Blood Flow & Metabolism* **16**, 7-22 (2016).

39. Smith S, Nichols T. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage* **44**, 83-98 (2009).

40. Van Snellenberg JX*, et al.* Dynamic shifts in brain network activation during supracapacity working memory task performance. *Hum Brain Mapp* **36**, 1245-1264 (2015).

41. Van Snellenberg JX*, et al.* Mechanisms of Working Memory Impairment in Schizophrenia. *Biol Psychiatry* **80**, 617-626 (2016).

42. De Blasi B*, et al.* Noise removal in resting-state and task fMRI: functional connectivity and activation maps. *Journal of Neural Engineering* **17**, (2020).

43. Frederick Bd, Nickerson LD, Tong Y. Physiological denoising of BOLD fMRI data using Regressor Interpolation at Progressive Time Delays (RIPTiDe) processing of concurrent fMRI and near-infrared spectroscopy (NIRS). *NeuroImage* **60**, 1913-1923 (2012).

44. Tong Y, Hocke LM, Frederick BB. Low Frequency Systemic Hemodynamic “Noise” in Resting State BOLD fMRI: Characteristics, Causes, Implications, Mitigation Strategies, and Applications. *Frontiers in Neuroscience* **13**, (2019).

45. Tong Y, Hocke LM, Fan X, Janes AC, Frederick Bd. Can apparent resting state connectivity arise from systemic fluctuations? *Frontiers in Human Neuroscience* **9**, (2015).

46. Tong Y, Hocke LM, Nickerson LD, Licata SC, Lindsey KP, Frederick Bd. Evaluating the effects of systemic low frequency oscillations measured in the periphery on the independent component analysis results of resting state networks. *NeuroImage* **76**, 202-215 (2013).

47. Barbey AK, Koenigs M, Grafman J. Dorsolateral prefrontal contributions to human working memory. *Cortex* **49**, 1195-1205 (2013).

48. Funahashi S, Takeda K, Watanabe Y. Neural mechanisms of spatial working memory: Contributions of the dorsolateral prefrontal cortex and the thalamic mediodorsal nucleus. *Cognitive, Affective, & Behavioral Neuroscience* **4**, 409-420 (2004).

49. Hampson M, Zhang G, Yao L, Zhang H, Long Z, Zhao X. Improved Working Memory Performance through Self-Regulation of Dorsal Lateral Prefrontal Cortex Activation Using Real-Time fMRI. *PLoS ONE* **8**, (2013).

50. Hodge MR*, et al.* ConnectomeDB—Sharing human brain connectivity data. *NeuroImage* **124**, 1102-1107 (2016).

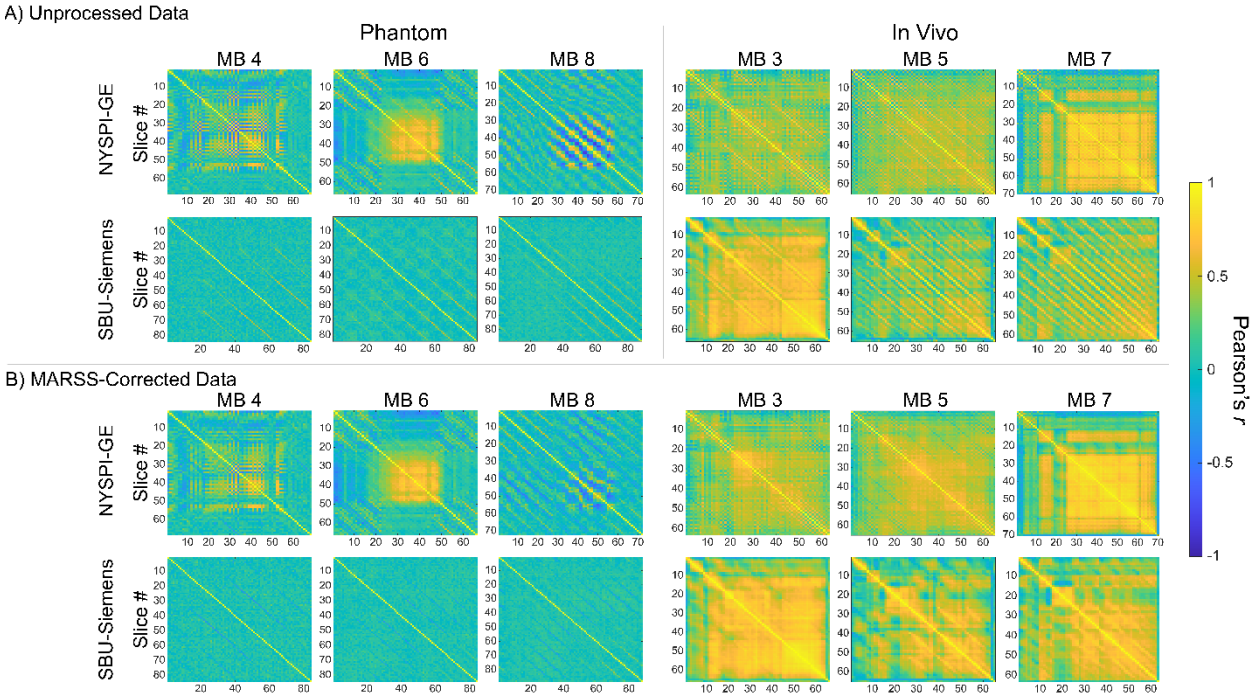

Figure 1. Pearson correlation matrices between average signal in all slice pairs in A) unprocessed data and B) MARSS-artifact corrected data, performed in phantom and in vivo at varying multiband (MB) acceleration factors.

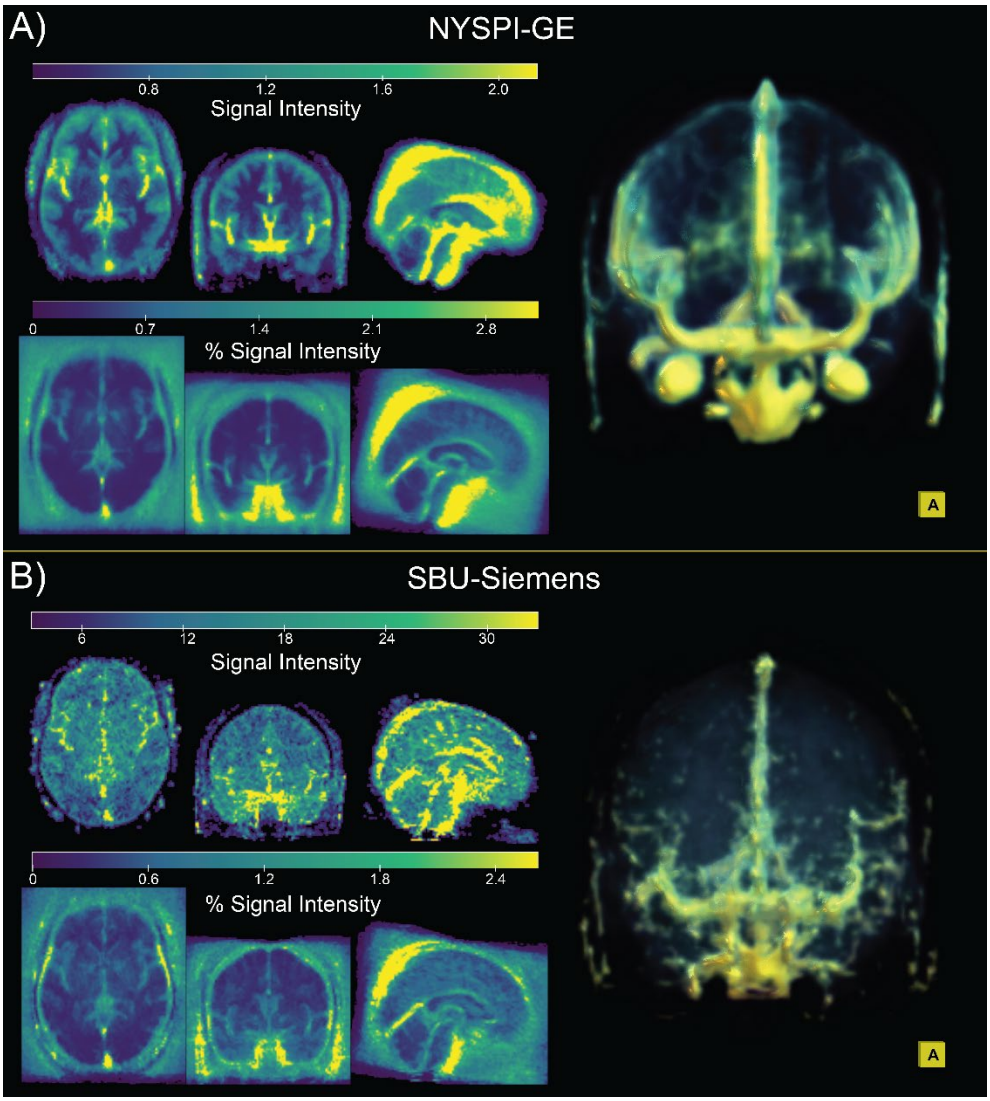

Figure 2. Average spatial distribution of the isolated artifact signal in resting-state data from A) NYSPI-GE and B) SBU-Siemens. Artifact distribution is shown as both raw EPI signal intensity (top left of each panel) and as a percentage of the mean voxelwise EPI intensity (bottom left of each panel). Three-dimensional renders (right of each panel) display anterior view of the raw signal intensity artifact distribution.

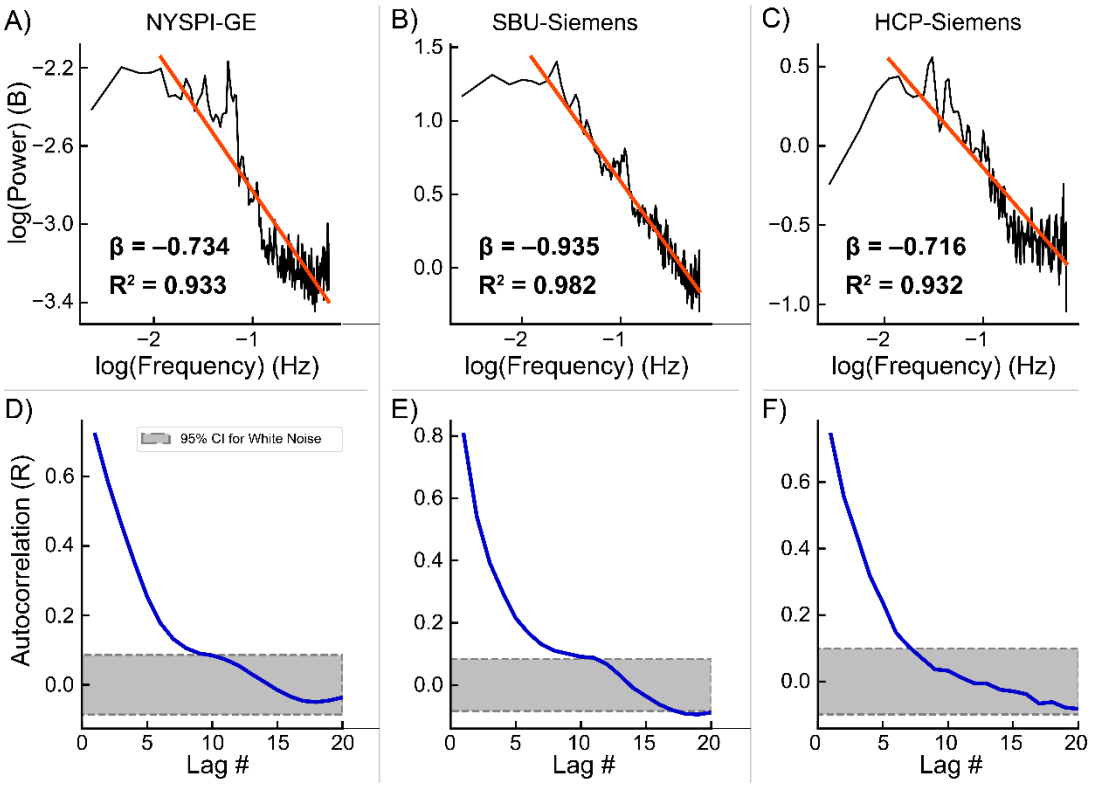

Figure 3. Log-log transformed power spectral density (PSD) of the isolated artifact timeseries concatenated across participants (top row) and average autocorrelation (bottom row) in task-based fMRI data from NYSPI-GE (A,D), SBU-Siemens (B,E), and HCP-Siemens (C,F). PSD plot annotations denote the slope (β) and R^2 of each linear fit (red line). Grey areas with dotted borders in autocorrelation plots denote the 95% confidence interval for significant autocorrelation over and above that of white noise.

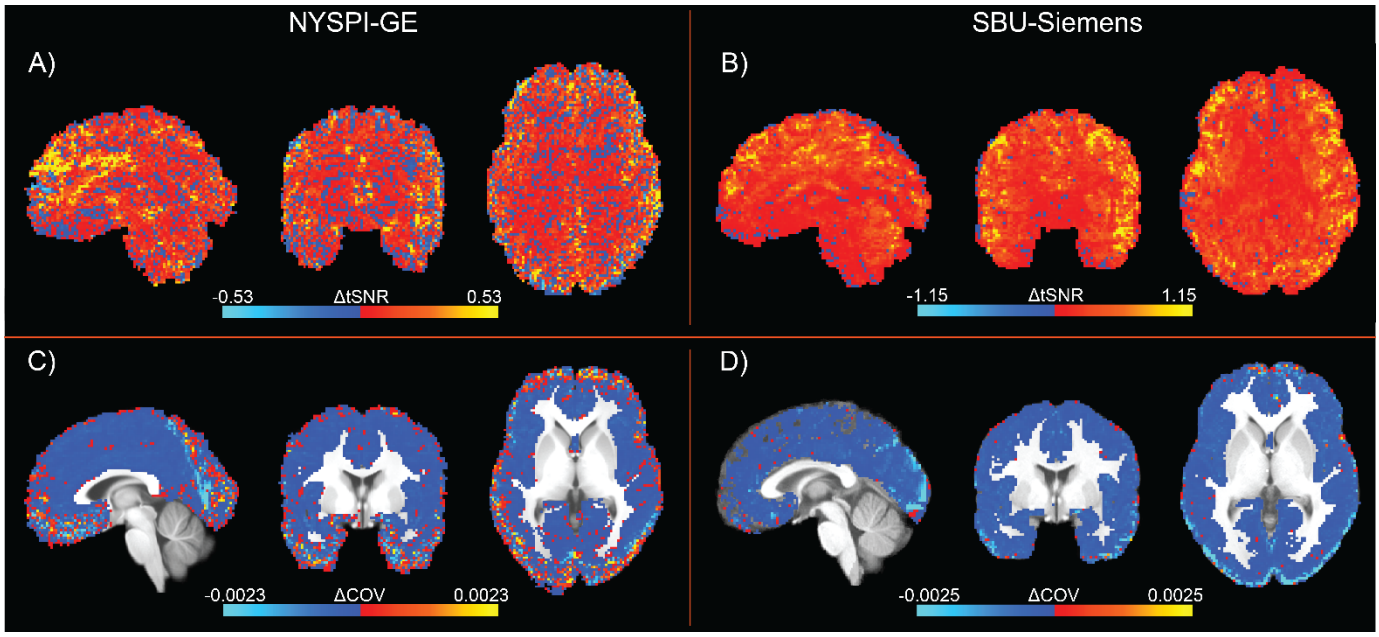

Figure 4. Mean, voxel-wise change in temporal signal-to-noise ratio (tSNR; A,B) and Coefficient of Variation (COV, C,D) as a result of MARSS artifact correction in preprocessed, task-based fMRI data from NYSPI-GE (A,C) and SBU-Siemens (B,D).

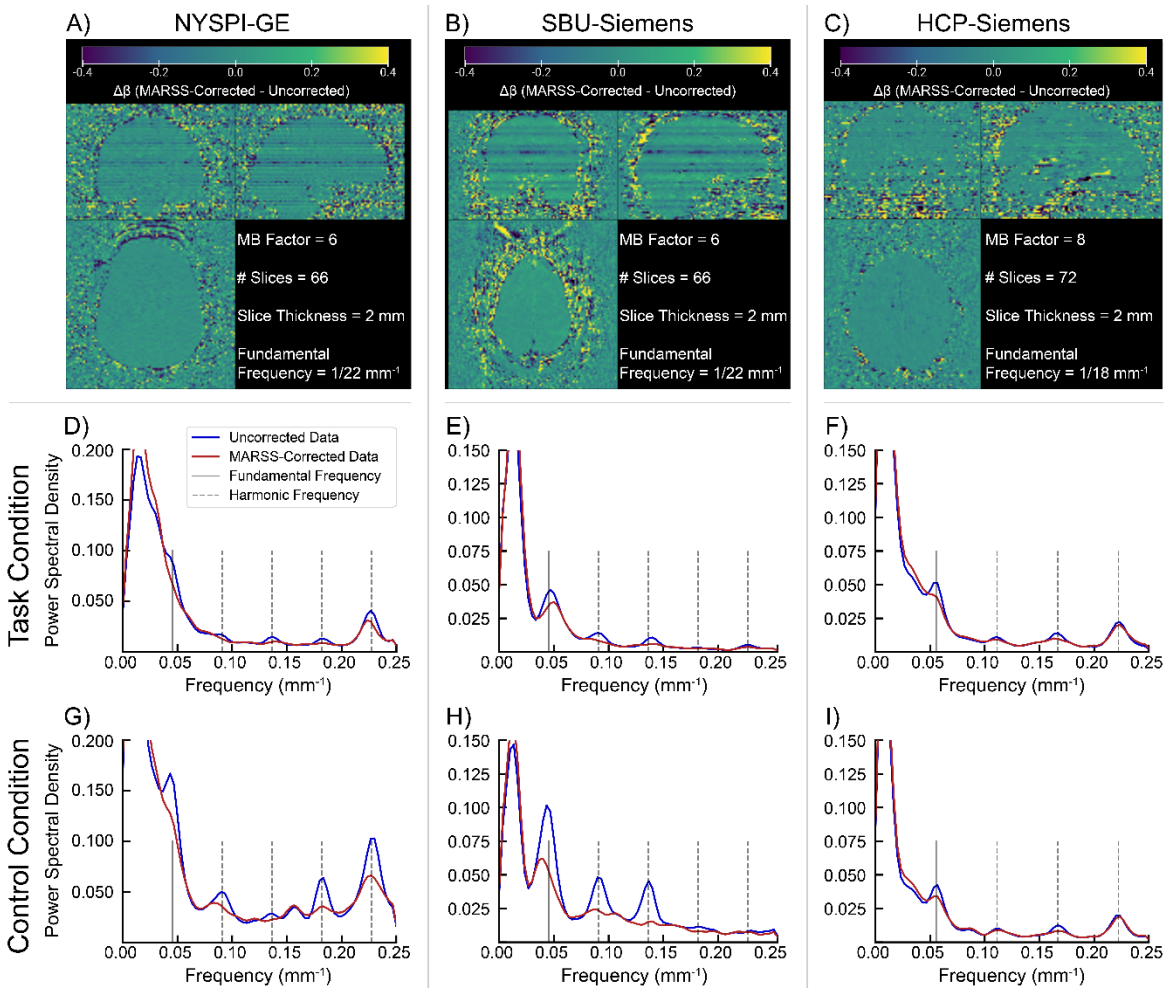

Figure 5. Single-participant examples of the spatial distribution of differences in task betas between MARSS-corrected and uncorrected data (A-C), as well as power spectral density (PSD) of average within-participants betas as a function of spatial frequency (D-I) in uncorrected data (blue line) and corrected data (red line) in data from NYSPI-GE (A, D, G), SBU-Siemens (B, E, H), and HCP-Siemens (F, I). In PSD plots, vertical lines denote the fundamental spatial frequency (solid grey line) and harmonic spatial frequencies (dotted grey lines) associated with the simultaneous slice pattern in each dataset. PSD is calculated separately for the task-on condition (D-F) and control condition (G-I) of each task.

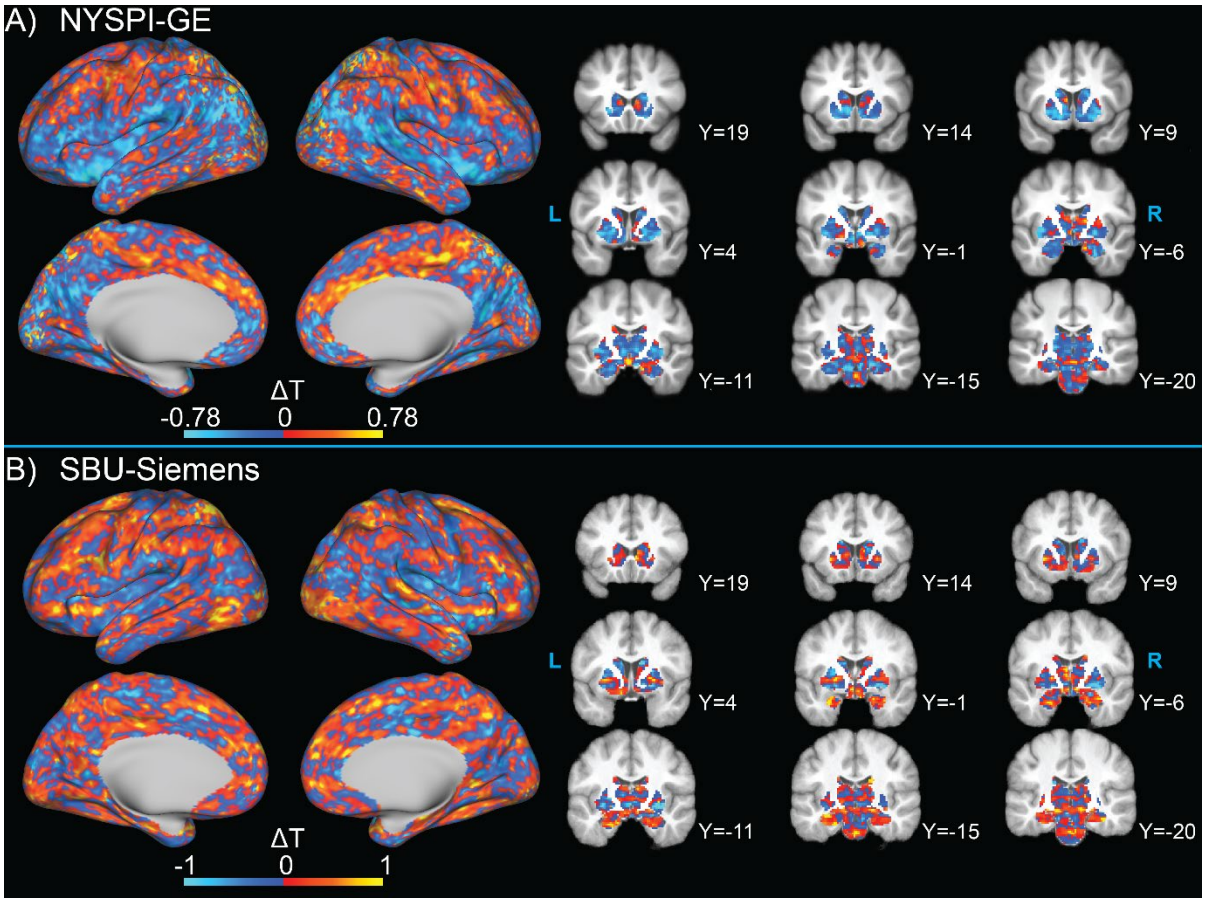

Figure 6. Spatial distribution of the difference in t-statistics in between-participants analyses performed in MARSS-corrected and uncorrected data from A) NYSPI-GE and B) SBU-Siemens. T-statistics were obtained without the use of threshold-free cluster enhancement (TFCE).

Table 1. Datasets used in this study. Datasets are referred to by their Dataset ID throughout the manuscript.

Dataset ID	Scanner	# Subjects	Phantom/ In Vivo	MB Factor	TR (ms)	TE (ms)	FOV (mm)	Matrix Size	Flip Angle (°)	Partial Fourier Factor	Voxel Size	Tasks
NYSPI-GE	GE MR 750 3T	N = 66	In Vivo	6	850	25	204x204	102x102	60	17/24	2 mm isotropic	RS (N=63), SOT (N=58)
		N = 1	In Vivo	3	1325	25	204x204	102x102	60	17/24	2 mm isotropic	RS
				5	850		204x204	102x102				
				7	800		192x192	96x96				
		N = 1	Phantom	4	1300	25	204x204	102x102	60	17/24	2 mm isotropic	RS
				6	850		204x204	102x102				
				8	850		192x192	96x96				
SBU-Siemens	Siemens Prisma 3T	N = 10	In Vivo	6	800	25	204x204	102x102	60	6/8	2 mm isotropic	RS (N=10), SOT (N=10)
		N = 1	In Vivo	3	1580	25	204x204	102x102	60	6/8	2 mm isotropic	RS
				5	939							
				7	657							
		N = 1	Phantom	4	1505	25	204x204	102x102	60	6/8	2 mm isotropic	RS
				6	1010							
				8	798							
ABCD-Siemens	Siemens Prisma 3T	N = 25	In Vivo	6	800	30	216x216	90x90	52	N/A	2.4 mm isotropic	RS
ABCD-GE	GE MR750 3T	N = 3	In Vivo	6	800	30	216x216	90x90	52	N/A	2.4 mm isotropic	RS
ABCD-Philips	Philips Achieva dStream 3T	N = 7	In Vivo	6	800	30	216x216	90x90	52	N/A	2.4 mm isotropic	RS
HCP-Siemens	Siemens Connectome Skyra 3T	N = 25	In Vivo	8	720	33.1	208x180	104x90	52	N/A	2 mm isotropic	RS (N=25), WM (N=25)

SBU = Stony Brook University; NYSPI = New York State Psychiatric Institute; ABCD = Adolescent Brain Cognitive Development; HCP = Human Connectome Project; MB = Multiband; TR = Repetition Time; TE = Echo Time; FOV = Field of View; RS = Resting-State; SOT = Self-Ordered Working Memory Task; WM = Working Memory Task

Table 2. Deciles of the greyordinate-wise, absolute change in T-statistic resulting from MARSS artifact correction in between-participants analyses performed both with and without Threshold-free Cluster Enhancement (TFCE).

		Deciles of Absolute Change in T-statistic ("x% of T-stats change by at least _")								
Dataset	**TFCE**	**90%**	**80%**	**70%**	**60%**	**50%**	**40%**	**30%**	**20%**	**10%**
NYSPI-GE	No TFCE	0.031	0.063	0.096	0.132	0.171	0.216	0.270	0.341	0.458
	TFCE	0.188	1.090	3.120	6.421	11.243	18.328	28.561	45.046	83.804
SBU-Siemens	No TFCE	0.030	0.062	0.094	0.129	0.169	0.217	0.275	0.353	0.490
	TFCE	0.101	0.420	1.015	1.893	3.041	4.747	7.415	12.175	23.532

SBU = Stony Brook University; NYSPI = New York State Psychiatric Institute; TFCE = Threshold-free Cluster Enhancement