A Comparison of fMRI Data-Derived and Physiological Data-Derived Methods for Physiological Noise Correction

Martin Krentz[1], Rayyan Tutunji[1], Nikos Kogias[1], Hariharan Murali Mahadevan[2], Zala C. Reppmann[3], Florian Krause[1] and Erno J. Hermans[1]

[1]Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands.

[2]Max Planck Institute for Biological Intelligence, Am Klopferspitz 18, 82152 Planegg, Germany

[3]Department of Psychiatry and Neurosciences - CCM, Research Division of Mind and Brain, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Charitéplatz 1, 10117 Berlin, Germany.

We have no conflicts of interest to disclose.

Correspondence: Martin Krentz (martin.krentz@donders.ru.nl)

Abstract

Physiological noise has been shown to have a large impact on the quality of functional MRI data, especially in areas close to fluid-filled cavities and arteries, such as the brainstem. Commonly, physiological recordings during scanning are transformed with methods such as RETROICOR and used as nuisance regressors in general linear models to remove variance associated with cardiac and respiratory cycles from the data. In contrast, modern pre-processing pipelines such as fMRIPrep, have created easy access to streamlined data-driven noise reduction methods such as aCompCor and ICA-AROMA. In combination, these methods have shown efficacy in correcting for motion, scanner as well as physiological artifacts. Given the ease of usability, it has to be questioned, whether there is any added benefit to applying logistically demanding methods such as RETROICOR. To answer this question, we applied RETROICOR, ICA-AROMA and aCompCor to a resting-state data set and compared variance explained by the respective methods and improvements in temporal signal-to-noise ratio throughout different regions of interest in the brain. In line with previous literature, RETROICOR significantly explains variance throughout the brain with peaks around areas of strong cardiac pulsations. ICA-AROMA and aCompCor largely account for the same variance. Nonetheless, RETROICOR retains unique explanatory power in individual participants. Further analysis points towards a pattern of unreliability of ICA-AROMA and aCompCor to consistently remove physiological noise across recordings, which is compensated by RETROICOR. While some of this inconsistency could be attributed to misclassifications in the noise selection models of ICA-AROMA, most is likely the consequence of secondary factors such as fMRI sequence parameters (e.g. long TR) limiting the efficiency of aCompCor and ICA-AROMA. Thus, it is advisable to additionally apply RETROICOR, especially when assuming regionally high levels of physiological noise.

1. Introduction

1.1. Rising interest in brain stem imaging

Functional imaging of brainstem regions has become increasingly important for research in a variety of neuroscience sub-fields, ranging from investigations of task-related activity in brain stem nuclei to functional connectivity with higher-order brain networks (Chu et al., 2021; Colizoli et al., 2022; Henderson & Macefield, 2013). For instance, the Locus Coeruleus (LC) has seen a surge of interest, because of the crucial role it plays in a variety of processes as the sole source of cortical norepinephrine (Elman et al., 2017; Schwarz & Luo, 2015). Despite the interest, it is largely unexplored in its functional involvement in neuroimaging due to its' high susceptibility to noise interference, primarily physiological noise. While availability of higher-field strength MRI is growing and allowing for higher resolution imaging and progress in usage of novel sequences to structurally identify different brainstem nuclei (Singh et al., 2022), the functional acquisition of their respective activity remains notoriously difficult. Over the past two decades a variety of tools have been proposed to reduce the impact of physiological noise on fMRI data quality (Brooks et al., 2013; Caballero-Gaudes & Reynolds, 2017; Murphy et al., 2013). However, as the available solutions differ in accessibility, effort and ease of use, questions about their relative theoretical benefits have to be weighed against practical considerations. With a needed push for reproducibility and ease of access, the question arises whether standard processing pipelines, such as fMRIPrep (Esteban et al., 2018), streamlining pre-processing steps and accompanying noise correction methods such as ICA-AROMA (Pruim, Mennes, van Rooij, et al., 2015) and aCompCor (Behzadi et al., 2007), could substitute more labor and expertise extensive methods with external physiological recordings even for imaging of brainstem areas (Glover et al., 2000).

1.2. Sources of physiological noise

Classically, physiological noise describes fMRI signal variation attributed to physiological processes, such as cardiac and respiratory activity, which are unrelated to neural signal of interest (i.e. task or rest BOLD effects). These undesired signal variations are diverse in their temporal and spatial characteristics, as well as their discernability and acting mechanisms.

1.2.1. Non-BOLD effects of physiological noise

One of the most prominent components, often associated with cyclic cardiac artifacts, are cardiac pulsations, leading to voxel displacement and partial-volume effects prominently in areas proximal to large arteries or CSF areas such as 3rd or 4th ventricle (Dagli et al., 1999; Glover et al., 2000; T. T. Liu, 2016). Similarly, cyclic respiratory activity shows indirect BOLD effects through breathing induced organ movement and susceptibility changes, resulting in magnetic field change (B0) as

well as modulation of intravoxel dephasing (Raj et al., 2001) and venous pressure (Windischberger et al., 2002). In difference to cardiac cyclic artifacts, the localization of respiratory effects appears stronger along the edges of the brain and ventricles (van de Moortele et al., 2002; Windischberger et al., 2002).

1.2.2. BOLD effects of physiological noise

Beyond effects on the quality of data acquisition itself and limitations in accuracy of the BOLD assessment, low frequency physiological processes such as heart rate frequency changes (Shmueli et al., 2007) or changes in CO2 as a potential vasodilator (Birn et al., 2006, 2008; Wise et al., 2004) can directly affect the BOLD signal. Furthermore, an interdependency between cardiac and respiratory activity exists, reflecting itself in respiratory cycle dependent heart rate changes (sinus arrhythmia, Berntson et al., 1993) as well as CSF flow, hypothesized to be largely driven by changes in negative intrathoracic venous pressure as a consequence of inhalation (Dreha-Kulaczewski et al., 2015; Kollmeier et al., 2022).

Since measured BOLD is dependent on cerebral blood flow (CBF) as well as cerebral blood volume (CBV), heart rate or respiration induced blood pressure and CO2 changes directly affect the derived signal (Das et al., 2021; T. T. Liu, 2016). These changes in difference to artifacts created by the cyclic higher frequency cardiac (~ 1.3Hz) and respiratory (~ 0.5Hz) dynamics, act in low frequency bands (~ 0.1Hz) and thus are especially detrimental for methodology such as resting-state functional connectivity which relies on synced low frequency changes in activity (Kelly et al., 2012; Zuo et al., 2019).

While the described physiological noise has a varying impact on all BOLD-based brain functional assessments, brainstem regions are specifically affected due to their aforementioned high susceptibility to arterial and CSF pulsatility and respiratory magnetization effects (Brooks et al., 2013; Krüger & Glover, 2001; T. T. Liu, 2016). This difficulty is amplified by the general nature of assessed brainstem nuclei due to their small size compared to cortical gray matter regions, close proximity to other brainstem regions with distinct functionality and direct edge proximity to areas of non-interest such as the 4th ventricle. Some of these downsides could be compensated with high-resolution fMRI assessments using high field-strength (i.e. 7T) and improved receiver coil positioning and sensitivity. However, while high-field fMRI imaging delivers on higher resolution and spatial discernability it is disproportionally affected by physiological noise and thus makes further compensation of physiological noise paramount (Hutton et al., 2011; Krüger & Glover, 2001; Triantafyllou et al., 2005)

1.3. Correcting for physiological noise

Over the last decade a wide range of improvements has been made in regards to tackling noise in resting-state fMRI. Not only in regard to the quality of noise correction, but also its ease in applicability. With a push for reproducible science and lower barrier of access, software packages such as 'fMRIPrep' (Esteban et al., 2018) have been growing in popularity; making access to certain noise correction pipelines less dependent on high levels of localized expertise.

1.3.1. Sequence based approaches

Due to the complexity of physiological noise dynamics, a wide variety of methods have been proposed to tackle different aspects of the problem at different stages of acquisition and analysis (Caballero-Gaudes & Reynolds, 2017; Murphy et al., 2013). One way to account for physiological noise, is to do so pre-emptively by choosing sequences such as multi-echo sequences to distinguish between inflow and BOLD-like signal changes (Glover et al., 1996; Kundu et al., 2012) or sequences using cardiac gating (Amoore & Ridgway, 1989), a method binding volume triggers to the cardiac rhythm. While effective, these methods do only partially limit the impact of complex physiological noise and can lead to difficulties both in post-processing and data interpretability.

1.3.2. Using external recordings for physiology clean-up

More prominent approaches of physiological noise correction focus on the post-hoc removal of physiologically induced artifacts. One possibility is to account for physiological noise by acquiring cardiac and respiratory measures during fMRI assessments. The most common approach – and often treated as the gold standard in physiological noise correction - is RETROICOR (Glover et al., 2000), a Fourier series of cardiac and respiratory signal to model the phase of the respective cycles. This phase information is typically used for nuisance regression by being included into planned general linear models. Inclusion of phase information specifically aims to reduce effects of cardiac pulsatility and motion-like respiratory effects onto the main magnetic field. In line with the higher susceptibility to physiological noise, areas in or proximal to the brainstem – especially areas surrounding main cerebral blood flow i.e. Circle of Willis – have shown the biggest improvement, despite overall efficacy in reducing physiological noise across the cortex (Glover et al., 2000; van Buuren et al., 2009).

While only suffering from the minimal downside of loss of temporal degrees of freedom (tDOF), this retrospective approach puts extra strain onto the logistics of an experimental set-up. It necessitates the application of extra equipment leading to increased preparation time and sources of potential malfunctions or experimental inconsistencies.

1.3.3. Data-driven approaches of physiological clean-up

An alternative approach for noise removal is to employ temporo-spatial decomposition of the data using independent or principal component analysis. This approach has been used to great success in identifying dynamics of functional-brain networks in resting-state (Beckmann et al., 2005; Beckmann & Smith, 2004). Since it can be used for identification of systematic temporo-spatial signal patterns, it similarly can be used to identify components of non-interest, such as head motion- and scanner artifacts, as well as some aspects of physiological noise (Pruim, Mennes, van Rooij, et al., 2015). While powerful, the main difficulty of using an ICA-approach is the necessary prior knowledge for correct classification of independent components as noise. In an effort to create an automated component selection, independent of rater expertise, Pruim and colleagues (2015) suggest an automatic tool for removal of noise artifacts (ICA-AROMA; Pruim, Mennes, Buitelaar, et al., 2015). Different to other solutions, such as ICA-FIX (Salimi-Khorshidi et al., 2014) which needs a training data-set to run a classifier and subsequent component selection, ICA-AROMA uses a pre-trained classifier based on three decision criteria for each component: (1) whether it contains high-frequency content unlikely to be brain signal, (2) whether its spatial layout shows high overlap with brain edges and (3) whether its spatial layout strongly overlaps with CSF regions. While this approach was originally designed to account for head motion artifacts, ICA-AROMA assesses features that are equally crucial in detecting physiological noise. The approach of using a shrunk brain masks to calculate edge fraction allows for possible identification of respiratory bulk head motion or effects visible in venous drainage in the sagittal sinus. More importantly, however, the CSF fraction approach classifies based on IC component spatial distribution along edges of CSF within the brain – typically regions most affected by cardiac pulsatility. Thus, while ICA-AROMA does not aim for physiological noise correction, its mechanisms make detection likely. ICA-AROMA can be furthermore combined with aCompCor (anatomical Component Based Noise Correction), a method utilizing principal components of white matter- and CSF masks as nuisance variables (Behzadi et al., 2007), which is often used for physiological noise correction. Similar to ICA-AROMA, it can be derived post-hoc from pre-processed data and has shown promise in approaching the efficacy of externally recorded methods such as RETROICOR (Behzadi et al., 2007; Muschelli et al., 2014). The joint integration in fMRIPrep makes it a ready-made-approach of correcting for physiological noise.

1.4. Commonalities and differences between noise correction methods

In this article we assess whether a combination of ICA-AROMA and aCompCor has equal potential to classical RETROICOR in removing physiological noise, which might allow easier access to brain stem assessment without sacrificing logistic feasibility and a priori considerations.

The main goal of this study is to closer investigate the commonalities and differences between the externally recorded RETROICOR and functional data-driven combination of ICA-AROMA and aCompCor in regards to explaining signal-of-non interest in resting state fMRI. To test group-level differences we assess the shared and unique contributions to improvement of temporal signal to noise ratio (tSNR) as a consequence of the different correction methods for separate regions of interest. The regions are chosen for different levels of specificity (whole-brain, gray matter, brain stem and locus coeruleus) in regard to proximity and likelihood of high physiological noise. By testing the unique contribution of the respective approaches to image variance, we hope to identify common components of noise signal and their consistency across a variety of subjects. Furthermore, by using the information about spatial patterns gained from cardiac and respiratory regressors, we investigate the efficacy of ICA-AROMA and aCompCor in regards to their consistency and stability. By answering these questions, we hope to shed light on the efficacy of physiological noise correction in brainstem imaging, and evaluate the efficiency of ICA-AROMA and aCompCor compared to RETROICOR as an easily accessible lower-threshold post-hoc alternative to classical physiology correction.

2. Methods

2.1. Participants

Thirty healthy volunteers (15 females, 15 males) between 18 and 55 years were recruited for this experiment via the 'Radboud Research Participaction System' and compensated 60€ for their participation. Participants were right-handed, healthy, and without a history of neurological or psychiatric illness. Additionally, MRI contraindications, such as ferromagnetic implants or a history of claustrophobia, were an exclusion criterion. Before completion of the experimental sessions, participants received instructions and information about the experimental procedures. The data used in this article is part of a larger study involving multiple sessions and manipulations of no relevance for the here described approach. Two participants did not complete all sessions and scans and were thus excluded from further analysis. Due to a technical problem with the heart-rate acquisition, one participant had to be excluded from the current analysis. Thus, a total of 27 subjects were used for reported analyses (12 females, 15 males; mean age = 21.45, SD = 3.6). An error in locus coeruleus imaging led to a missing mask for one subject, reducing the number of subjects included for the LC ROI analysis of LC to 26. The study was approved by the local ethics committee, and participants gave their written informed consent before the procedure.

2.2. Design

The study consisted of three separate sessions including behavioral measures not subject to this study, as well as structural and functional MRI acquisitions. In an initial session, structural images were acquired including a T1-weighted image as well as a combination of turbo spin-echo (TSE) sequences for identification of the Locus Coeruleus (Sasaki et al., 2006). Sessions two and three containing functional resting state runs were identical, except for a counter-balanced stress-induction procedure administered before the initial start of scanning without further relevance for the current analysis.

2.2.1. Procedure

Before entering the MR scanner, participants were attached to all physiological recording devices and while seated on the scanner bed. Subsequently, participants were placed into the head coil with firm foam cushioning around their head to reduce head movement. Additionally, a piece of tape was placed over their forehead to further reduce motion (Krause et al., 2019).

2.2.2. MRI data acquisition

All MR images were recorded on a Siemens Prisma-Fit 3T MR System (Siemens, Erlangen, Germany) using a 32-channel receiver head coil. Structural images were acquired using a T1-weighted sequence with using a 3D resolution of 0.8x0.8x0.8 mm (repetition time = 2.2s, echo

time = 26.4ms, flip angle = 11, field of view 256x256mm, slices = 320. Functional images were acquired with a BOLD contrast sensitive EPI sequence using a multiband acceleration factor of 3. For higher sensitivity in brain stem resolution, we chose a slice thickness of 1.5 mm (no gap) with an in-plane resolution of 1.5x1.5 mm (repetition time = 2020ms, echo time = 28.4ms, flip angle = 80, field of view = 210x210 mm, slices = 87) and a generalized auto calibrating partial parallel acquisition (GRAPPA) factor of 2. For structural assessment of the locus coeruleus (LC) and delineation as a functional mask, we used a T1-weighted turbo spin-echo sequence with slice thickness of 3mm and an in-plane resolution of 0.24x024mm (repetition time = 1870ms, echo time = 12ms, flip angle = 120, slices = 16, field of view = 220x200mm) (Sasaki et al., 2006). Detailed processing mechanisms for mask creation are presented in Appendix A.

2.3. MRI data preprocessing

2.3.1. Anatomical data preprocessing

Results included in this manuscript partially come from preprocessing performed using fMRIPrep (Esteban et al., 2018), which is based on Nipype 1.2.0 (Gorgolewski et al., 2011).Further processing was done using custom python scripts. The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al. 2010), distributed with ANTs 2.2.0 (Avants et al. 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as the target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter (WM), and gray matter (GM) were performed on the brain-extracted T1w using fast (FSL 5.0.9, RRID:SCR_002823, Zhang, Brady, and Smith 2001). Volume-based spatial normalization to one standard space (MNI152NLin6Asym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: FSL MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic. For LC imaging adjustments were made to the MR sequence parameters used by Sasaki et al. (2006) based on in-house testing results showing a maximization of contrast. An in-house pipeline was used (Supplement A) to delineate a usable LC mask for functional signal extraction on an individual level.

2.3.2. Functional image processing

For each individual resting state run the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. A deformation field to correct for susceptibility distortions was estimated based on a field map that

was co-registered to the BOLD reference, using a custom workflow of fMRIPrep derived from D. Greve\u2019s epidewarp.fsl script and further improvements of HCP Pipelines (Glasser et al., 2013). Based on the estimated susceptibility distortion, an unwarped BOLD reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9, Jenkinson et al., 2002). BOLD runs were slice-time corrected using 3dTshift from AFNI 20160207 (Cox, 1996, RRID:SCR_005927). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. The BOLD time-series were resampled into standard space, generating a preprocessed BOLD run in [\u2018MNI152NLin6Asym\u2019] space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Automatic removal of motion artifacts using independent component analysis (ICA-AROMA, Pruim, Mennes, van Rooij, et al., 2015) was performed on the preprocessed BOLD on MNI space time-series after removal of non-steady-state volumes and spatial smoothing with an isotropic, Gaussian kernel of 6mm FWHM (full-width half-maximum). \u201caggressive\u201d noise-regressors were collected and placed in the corresponding confounds file. All resampling was performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resampling was performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels. Components classified as noise after running MELODIC-ICA (FSL), were concatenated per subject to create a regressors matrix of AROMA noise regressors. The AROMA noise regressors are of variable length, as both the number of ICA components as well as classifications as noise varies across participants as a function of motion and other noise sources. These concatenated regressors will further be called ‘AROMA regressors’. Furthermore, aCompCor components were extracted from the nuisance regressor file created by fmriprep. For the analysis the first 5 PCA components derived from degraded white matter and CSF masks as implemented in aCompCor were used. These 5 components were concatenated and are further called ‘aCompCor regressors’.

2.3.3. Physiological data acquisition

Physiological data were recorded using a finger pulse sensor placed on the index finger (Brain Products, Gliching, Germany) connected to a BrainAmp ExG MR (Brain Products, Gliching, Germany). An abdominal respiration belt (attached to a pneumatic sensor, Brain Products; Gliching, Germany) was used to assess respiratory activity. Acquired physiological data were subsequently processed using an in-house tool for manual noise rejection, peak detection, and quality control (https://github.com/can-lab/hera). The outputs were a corrected time series for the cardiac acquisition. Respiratory data was transformed using another in house tool and concurrently cleaned for possible artifacts (https://github.com/can-lab/brainampconverter).

Subsequently, the processed physiological data was input into an inhouse MATLAB-based implementation of RETROICOR (v2.1; https://github.com/can-lab/RETROICORplus) combining trigger information and physiology to create multiple regressors modeling the phase of both cardiac and respiratory rate (Glover et al., 2000). The final model was reduced to include a total of 6 cardiac phase regressors and 8 respiratory phase regressors, and a multiplication term of the first cardiac and respiratory phase regressors was calculated to account for a possible interaction effect (Harvey et al., 2008). This resulted in a total of 18 regressors further called 'RETROICOR regressors'. For a detailed view of recording data quality see Appendix B.

2.4. Pipeline for correctional method comparison

2.4.1. General linear models integrating physiological noise correction

Regressors resulting from ICA-AROMA, aCompCor, and RETROICOR are typically used as nuisance variables in general linear models (GLMs) to remove variance attributed to various sources of noise from signal of interest. To directly assess the impact of the different noise correction procedures on an individual level we chose to compute GLMs treating these commonly used nuisance variables as variables of interest to directly assess the variance explained by the respective cleaning procedures. Furthermore, to assess how RETROICORs externally calculated nuisance regressors compare to the data-derived regressors of AROMA and aCompCor, combinatory GLMs were created. In total, six GLMs were computed using Nilearn (https://github.com/nilearn; Abraham et al., 2014). The contrasts computed for all GLMs were z-transformed and FDR-thresholded ($p < .05$) to ensure the detection of spatial patterns. A GLM was computed for each set of nuisance regressors respectively. To match the total number of regressor of the models described below we added a shuffled run- and subject-specific number of the respective untested method regressors to the model, in an effort to maintain within-subject comparability and temporal degrees of freedom (tDoF). (1-3) Separate GLMs were computed for

all RETROICOR, all ICA-AROMA and all aCompCor regressors respectively, as variables of interest. Consequently, an F-contrast was computed across all regressors for each method respectively. (4) A GLM with both aCompCor regressors as well as AROMA regressors was computed. The goal was to visualize whether there is a discernable benefit of running the PCA-based aCompCor method on top of ICA-AROMA. An F-contrasts was calculated across all aCompCor regressors to attribute all variance unique to aCompCor nuisance regression. (5) To assess the unique contribution of RETROICOR to ICA-AROMA we computed a GLM using RETROICOR as well as AROMA regressors. (6) To conclude, a GLM was run including RETROICOR, aCompCor, and AROMA regressors to assess RETROICOR's unique contribution beyond a combination of both data-driven methods. An F-contrast was computed across all RETROICOR regressors. All models were run without further nuisance regression, except for high-pass filtering (cut-off frequency = 120 Hz), as implemented in Nilearn. No second-level statistics were calculated based on these models, as the main focus was an inter-individual assessment of stability and individual impact of the chosen respective noise correction methods as well as statistical limitations of using F-contrasts for second-level statistics.

2.4.2. Calculation and usage of temporal signal-to-noise ratio

To gauge the potential impact of nuisance regression on data metrics on a group level we analyzed the development of temporal signal-to-noise ratio (tSNR) as a consequence of noise regression using either RETROICOR, aCompCor, and AROMA. These calculations were made with a custom Python (v3.6.2) script (including image smoothing using a 6mm Gaussian FWHM-kernel) utilizing the functionality of Nilearn and NiBabel (https://github.com/nipy/nibabel; DOI: 10.5281/zonodo.4295521).

TSNR was calculated using the formula

$$TSNR = \frac{\mu}{\sigma} = \frac{\mu}{\sqrt{1/n \sum_{i=1}^{N} (x_i - \mu)^2}} \quad (1)$$

where N is the number of time points in a voxel time-series, μ is the mean of the time series and σ is its standard deviation. Calculation of voxel-wise tSNR was done for the entire functional data matrix, resulting in a tSNR matrix for the functional run. To test tSNR improvement as a result of noise regression, we used the residuals of the previously described GLMs 1 to 3 with the respective cleaning procedures for the tSNR calculation. Furthermore, to allow for improved relative comparability, we computed the within-subject increase of tSNR as a consequence of noise correction over no correction in percent. To retain the number of temporal degrees of

freedom, one more GLM was computed, using shuffled versions of RETROICOR, aCompCor, and AROMA regressors per subject, and used the residuals of this regression as 'uncleaned' data. This approach resulted in percent change TSNR maps for each method with RETROICOR, AROMA, and aCompCor noise regression respectively. Given that aCompCor and AROMA are presented for conjoint use, an additional map was computed assessing the unique improvement of tSNR of aCompCor when tested together with AROMA (GLM 4). To calculate the unique improvement of tSNR as a consequence of RETROICOR when combined with AROMA, we calculated a voxel-wise difference score in tSNR between a combined version of RETROICOR and AROMA and the previously calculated implementations (GLM 5). Finally, the approach as shown above was then further extended to create a tSNR-map showing voxel-wise unique RETROICOR improvement when added with aCompCor and AROMA (GLM 6).

Inferential testing of TSNR improvement was done by means of one-sample t-tests, with FDR-correction for multiple comparisons. These comparisons were made separately for four different regions of interest to account for a different level of abstractions and proximities to physiological noise sources: (1) Whole brain, (2) cortical grey matter, (3) brainstem and (4) locus coeruleus.

2.5. Identifying inconsistency in ICA-AROMA processing

A remaining question is, whether deviations in explanatory effects between RETROICOR and ICA-AROMA are associated with inconsistencies in individual identification of physiological noise when using ICA-AROMA. We calculated a correlation between the tSNR improvement as a consequence of ICA-AROMA nuisance regression and unique tSNR improvement as a consequence of RETROICOR when tested with ICA-AROMA in the model. On a group level, this correlation could indicate the dynamic relationship between successful noise removal of RETROICOR, depending on the success of ICA-AROMA.

2.6. Identifying potential misclassifications in ICA-AROMA

One aspect of possible individual variation in physiological nuisance regression using AROMA is the potential of misclassification of physiological noise components as signals in the AROMA algorithm following ICA-MELODIC. To control whether effects of RETROICOR are caused by component misclassifications in ICA-AROMA we deployed a comparison assessing whether spatial similarity in unique RETROICOR effects as implemented in GLM 5 and melodic components classified as a signal, might increase AROMA classification accuracy.

We ran regression models for each MELODIC component on a per-subject basis to create a spatial back projection of a component. Subsequently, a 'Goodness of Fit' score was calculated

between each melodic component and a masked binarized version of all suprathreshold voxels uniquely explaining variance across all RETROICOR regressors in GLM 5. The score was calculated, by comparing the ratio of inside versus outside voxels of the MELODIC components with respect to the RETROICOR mask. All components across subjects that fulfilled the conditions (1) goodness of fit higher than 0.75 and (2) a component being classified as signal, were then taken for further analysis.

Following overlap identification, MELODIC components were added to the respective AROMA regression models (GLM 5) to assess whether addition of these potential misclassification reduces the variance uniquely explained by RETROICOR. To quantify this change, all suprathreshold voxels of unique RETROICOR regression and after addition of any newly identified MELODIC component were compared and an average z-change was calculated between the two respective maps.

3. Results

3.1. tSNR changes and differences resulting from noise regression

To assess the impact of RETROICOR in comparison to the other two noise correction methods we first visually assessed the spatial distribution of tSNR changes as a consequence of the respective noise removal application. These images allow to assess the spatial layout of observed tSNR changes resulting from nuisance regression, both with and without the presence of other cleaning methods. Furthermore, it has to be noted that improvement is measured in percent and thus scales are variable for visualization purposes and not directly comparable between methods.

3.1.1. Spatial layout of tSNR improvements

Looking at the spatial distribution of tSNR improvement as a consequence of RETROICOR (Figure 1a), it is apparent that the relatively strongest tSNR changes are spatially clustered around arterial vascular areas. As expected, strong changes are visible along the basilar artery and cerebral arteries as well as ventricles. Furthermore, some sagittal sinus related changes can also be detected in posterior parts of the brain. Comparatively, tSNR change as a consequence of AROMA cleaning as visible in Figure 1b shows a much more complex spatial distribution. In line with its intended use for motion correction, high tSNR changes can be seen in prefrontal and occipital cortex areas typically associated with head motion (Pruim, Mennes, Buitelaar, et al., 2015; Pruim, Mennes, van Rooij, et al., 2015). However, substantial tSNR changes also occur in most inferior brainstem areas which are strongly affected by magnetization and are consequently most likely filtered out via the frequency-based component selection in ICA-AROMA (Pruim, Mennes, van Rooij, et al., 2015). It is apparent that despite its intentionality of motion correction AROMA also strongly improved tSNR in areas associated with RETROICOR. Along the vascular systems and ventricles AROMA shows substantial changes in tSNR. Similarly, aCompCor shows a wide-spread pattern of tSNR improvement, with however strong similarity to RETROICOR. While tSNR improvement is more homogenous the clear peaks mimic RETROICOR tSNR improvement by being centered around the basilar artery, Circle of Willis and ventricular system. Comparing the spatial tSNR improvement pattern of the two data-driven methods in Figure 1d, it becomes clear that existing differences occur mostly along grey matter areas along the cortex with a focus on pre-frontal cortex and posterior occipital lobe. A distinct spatial pattern of added tSNR impact surrounding the areas visible for aCompCor use in Figure 1c is not visible.

When visualizing tSNR improvement of RETROICOR in presence of AROMA, overall relative improvement is low yet centered around the vascular and ventricular areas hallmark to RETROICOR. Despite a strong visible improvement of AROMA to tSNR improvement in these

areas as visible in Figure 1b, the small effect remains on average in these areas when applying RETROICOR. It has to additionally be noted that the posterior sagittal sinus venous drainage effect also remains visible beyond AROMA cleaning regression. A similar picture of spatial distribution is retained when adding aCompCor to the AROMA-based regression. While the impact visually appears further reduced the spatial pattern around vascular and ventricular areas remains. Note that no thresholding was implemented for the creation of these images.

3.1.2. Group effects within ROIs

To test the observed group effects on tSNR and the variety in the spatial layout of tSNR changes, we compared the tSNR change based on four different ROIs to see where the impact might be strongest These regions were chosen to be progressively smaller and in higher proximity to areas affected by physiological noise. As well as showing different levels of abstraction to account for varying degrees of sensitivity to potential physiological noise. We assessed tSNR change for whole-brain (1), grey matter (2), brainstem (3) and the locus coeruleus close to the 4th ventricle (4) as visible in Figure 2.

First, as a sanity check, we tested whether the different noise correction methods had a significant impact on tSNR across the four different abstraction levels as visualized in Figure 2. AROMA shows a consistent strong improvement of tSNR in line with the multi-facetted approach it takes. Similarly, aCompCor and RETROICOR show significant difference from 0 across all ROIs. Respective comparisons between methods were not computed here, as (1) samples across ROIs are not independent and partially redundant and (2) the singular use of one method and its impact on tSNR is of little overall value in respect to the questions. TSNR improvement – accounting for tDoF – was consistently impactful across all regions of interest. With RETROICOR increasing tSNR significantly for whole brain (1); $t(26) = 9.095$, $p < .001$, $d = 1.783$; cortical grey matter (2); $t(26) = 7.725$, $p < .001$, $d = 1.453$; brainstem (3); $t(26) = 7.409$, $p < .001$, $d = 1.151$; as well as locus coeruleus (4); $t(25) = 5.479$, $p < .001$, $d = 1.074$. AROMA regression likewise lead to significant tSNR improvement across regions of interest for whole brain (1); $t(26) = 5.998$, $p < .001$, $d = 1.174$; cortical grey matter (2); $t(26) = 5.667$, $p < .001$, $d = 1.321$; brainstem (3); $t(26) = 6.739$, $p < .001$, $d = 1.111$; and locus coeruleus (4); $t(25) = 8.513$, $p < .001$, $d = 1.669$. Furthermore, aCompCor does significantly explain variance for whole brain (1); $t(26) = 6.409$, $p < .001$, $d = 1.257$; cortical grey matter (2); $t(26) = 6.77$, $p < .001$, $d = 1.66$; as well as brainstem; $t(26) = 8.466$, $p < .001$, $d = 1.327$; and locus coeruleus (4); $t(26) = 6.68$, $p < .001$. $d = 1.345$. Testing the unique contribution of aCompCor to tSNR improvement when in combination with AROMA, tSNR was significantly improved for whole brain (1); $t(26) = 7.025$, $p < .001$, $d = 1.377$;

as well as cortical gray matter (2); $t(26) = 8.0455$, $p < .001$, $d = 1.577$; brain stem; $t(26) = 4.012$, $p < .001$, $d = 0.786$; and locus coeruleus; $t(26) = 4.345$, $p < .001$, $d = 0.852$.

Figure 3 displays the tSNR improvement unique to an addition of RETROICOR either alone or in combination with the other two methods. Testing the difference shows that RETROICOR uniquely improves tSNR over AROMA, showing effects across all ROIs for whole brain (1); $t(26) = 9.702$, $p < .001$, $d = 1.902$; cortical gray matter; $t(26) = 7.791$, $p < .001$, $d = 1.528$; brain stem; $t(26) = 10.359$, $p < .001$, $d = 2.032$; and locus coeruleus; $t(26) = 3.628$, $p = .0013$, $d = 0.711$. Additionally, the unique contribution of RETROICOR to a joint application of AROMA and aCompCor remained significant across ROIS for whole brain (1); $t(26) = 9.448$, $p < .001$, $d = 1.853$; cortical gray matter (2); $t(26) = 8.873$, $p < .001$, $d = 1.74$; brain stem(3); $t(26) = 9.824$, $p < .001$, $d = 1.923$; as well as locus coeruleus (4); $t(26) = 3.6$, $p = .0013$, $d = .706$.

3.2. Statistical improvement for selection of noise regression methods

To directly assess the individual specificity of the different noise correction procedures, the computed GLMS were chosen as means to test the respective variance explained by the different noise correction methods in a resting state run for RETROICOR, ICA-AROMA, and aCompCor respectively. The calculations were made for all methods separately in comparison to uncleaned data (Appendix C1-C3). Spatial patterns are generally consistent with tSNR maps as seen in Figure Appendix C1-C3, yet generally differ in the amount of variance explained across methods.

3.2.1. tSNR improvement due to RETROICOR over ICA-AROMA

As can be seen in Figure 4, the unique explained variance of RETROICOR when tested with AROMA in the same model varies strongly across participants. While it is visible that subjects previously showing almost no discernable pattern of activation (010 and 012) show no variance explained in addition to AROMA, some subjects do show a significant amount of unique explained variance (014, 015, 017) that, similar to aCompCor, does not seem to be specific to the brainstem. For subjects that generally show a low amount of unique RETROICOR variance, the variance explained appears more focused around the brainstem and the Circle of Willis (018, 022, 027, 028, 030). A unique case poses subject 021, with a pattern not exclusively linked to the physiologically explained variance.

3.2.2. tSNR improvement due to RETROICOR over combined ICA-AROMA and aCompCor

Investigating the unique variance of RETROICOR when testing with a combination of AROMA and aCompCor a pattern appears showing high inter-subject variance in the impact of RETROICOR cleaning as visible in Figure 5. As visible on subjects 004, 006, 020, 022, 024 and

025, AROMA in combination with aCompCor does well in explaining physiological noise. However, in contrast, more impacted subjects (002, 003, 007, 008, 016, 017, 023, 027, 029), retain a significant amount of uniquely explained variance by RETROICOR. While the impact of RETROICOR on an individual level is comparatively significantly decreased, especially around the brainstem as well as third and fourth ventricles the combination of data-driven approaches does not appear to completely account for physiological noise captured with external recordings.

3.3. Inter-subject inter-run variability in ICA-AROMA processing

As visible in Figure 6, there is a significant negative correlation between unique tSNR improvements as a consequence of RETROICOR cleaning and tSNR improvements as a consequence of ICA-AROMA cleaning for both whole brain as well as grey matter. Similarly, while not significant, there is a visible continuation of the pattern for both brainstem and LC. Together, this hints towards an inconsistency in ICA-AROMA noise removal which is compensated in RETROICOR. It has to be noted that a relative reduced number of data point for brainstem and LC might have affected strength of statistical inferences for these ROIs.

3.4. Misclassifications in ICA-AROMA

Another aspect of ICA-AROMA is the classification of components using a pre-specified set of rules for noise identification. It is possible that variability in the consistency of this classification process, due to it not being tailored for physiological noise specifically, accounts for some of the RETROICOR effects. Consequently, we tested whether the spatial maps of either our RETROICOR approach or the aCompCor approach correspond to spatial maps of ICA components classified as signal by ICA-AROMA. As can be seen in Figure 7, the addition of the misclassified components severely reduces the variance explained by RETROICOR. The approach here however, can only serve exemplary as other components with high overlap and strong z-change would visually be identified as signal components, despite reducing the effects of RETROICOR. It can be argued, that the long TR and relatively short run led to clustering of noise signal in these components. As is visible across subjects the correction for misclassifications either implicitly by adding aCompCor to AROMA or by reclassifying MELODIC components as noise, does not completely explain RETROICOR results as some degree of additionally explained variance is retained.

4. Discussion

The need and acceptance for streamlined pre-processing pipelines has been growing. The now commonly used pre-processing tool fMRIPrep can automatically output regressors to be included in GLM analysis of functional MRI data with ICA-AROMA as well as aCompCor to remove noise attributed to a variety of sources (Esteban et al., 2018). This approach is compelling in comparison to classical physiological noise correction methods utilizing external measurements, such as RETROICOR (Glover et al., 2000). RETROICOR requires additional processing steps, equipment, and preparation during the actual execution of an experiment. The question arises how applicable the combination of data driven methods is to account for physiological noise in noise sensitive settings such as brain stem imaging. As expected, all tested methods (i.e. RETROICOR, ICA-AROMA and aCompCor) show a significant impact on tDOF as well as explained variance (Behzadi et al., 2007; Krüger & Glover, 2001; Pruim, Mennes, van Rooij, et al., 2015). While aCompCor and ICA-AROMA show a generalized high amount of explained variance, RETROICOR shows a consistent impact on explained variance, especially around fluid cavities such as ventricles and arteries surrounding the brain stem. On a group level, the difference between the data-derived methods and RETROICOR seems minute, yet a look at individual expressions reveals a consistent impact of RETROICOR with inconsistencies in ICA-AROMA and aCompCor predictive ability.

4.1. Effects of noise nuisance regression on group level

All tested methods show unique improvements in tSNR in comparison to a baseline that differs both in magnitude as well as spatial profile. In line with prior findings around RETROICOR, a large part of removed variance can be attributed to both the cardiac and the respiratory cycle. These tSNR improvements are largely explained by the underlying removal of non-BOLD physiological noise arising from cardiac pulsations. These motion-like pulsations result in changes mostly visibly in fluid-filled cavities such as the 4th and 3rd ventricle, as well as arteries and the Circle of Willis (Glover et al., 2000; Krüger & Glover, 2001). Beyond common pulsatility effects, there were few discernable spatial patterns. In contrast, ICA-AROMA reveals a more widespread improvement of tSNR at a higher magnitude in line with its broader removal of both physiological and motion noise (Pruim, Mennes, Buitelaar, et al., 2015; Pruim, Mennes, van Rooij, et al., 2015). While less specific to pulsatility-related noise as RETROICOR, it yet shows high improvement of tSNR in areas also impacted by RETROICOR cleaning. Peak improvement is centered on most inferior slices, likely heavily impacted by high-frequency noise due to both coil insensitivity as well as susceptibility artifacts, which thus is likely captured by ICA-AROMAs high-frequency content selection criteria (Pruim, Mennes, van Rooij, et al., 2015). Similarly, it is dominated in pre-frontal

and occipital locations related to bulkhead movement as well as susceptibility artifacts (Pruim, Mennes, Buitelaar, et al., 2015). aCompCor in comparison shows a similar focus on CSF and cardiac areas as RETROICOR (Muschelli et al., 2014). The PCA-based approach shows overall less specificity to high pulsation areas, yet nonetheless resembles in its peaks the previously shown pattern of RETROICOR (i.e. regions surrounding 3rd and 4th ventricle as well as Circle of Willis). The overall improvement of tSNR centered around gray matter is furthermore an expected consequence of the relatively broad approach of white matter nuisance regression (Muschelli et al., 2014). While the low TR of 2.02s leads to aliasing of actual cardiac frequency, it appears as if spatial coherence of the PCA components can compensate to some degree (T. T. Liu, 2016; Muschelli et al., 2014). Aliasing in this regard appears to be strongest for long TRs in regards to cardiac pulsations over respiration induced pulsatility (Huotari et al., 2019), yet long TR in this study does not negate the general detectability of pulsation related signal changes.

When adding ICA-AROMA and aCompCor to a GLM, the unique tSNR improvement of RETROICOR is generally strongly reduced. However, while large parts of its value are captured by these data-derived methods, small improvements remain, specifically centered around before mentioned fluid cavities affected by high cardiac pulsatility as occipital effects related to venous drainage. As expected, RETROICOR retains a high degree of improvement the closer to areas of pulsation, such as the LC, directly at the edge of the 4th ventricle. Yet, comparatively shows larger improvement effects on grey matter and brainstem as a whole. While mechanistically there is reason to expect a difference in effectiveness and an upside to direct nuisance regression over data-derived components, however, we first investigated whether variation in individual specificity account for possible differences.

4.2. Effects of noise nuisance regression on individual level

For all methods a pattern consistent with expected spatial features of variance explained was visible when testing in absence of other correction methods (Appendix C1-C3), though some features appeared more consistent than others hinting at variability in individual effectiveness. This picture of individual variability in the effectiveness of noise removal becomes more apparent when comparing RETROICOR to both ICA-AROMA as well as a combined approach of ICA-AROMA and aCompCor. For some subjects ICA-AROMA explains large amounts of variance in areas sensitive to cardiac pulsations, and thus almost entirely negates the effect of RETROICOR confound removal. However, in other subjects, AROMA fails to account for the same type of variance uniquely identified by RETROICOR. This trend of inconsistency can be clearly seen in the negative correlation between unique RETROICOR tSNR improvement and ICA-AROMA

improvements. This potentially means that RETROICOR performs better if ICA-AROMA failed to distinctly capture the physiological components. Thus, it can be inferred that while effective in removing physiological noise, inter-individual inconsistency creates a problem – one that could be detrimental when trying to assess areas, such as the brainstem, strongly susceptible to cardiac and respiration-induced pulsations. There are multiple possible explanations for these individual inconsistencies.

One aspect of ICA-AROMA inconsistencies could be the potential misclassification of components as signals that are attributable to physiological noise. In its design, ICA-AROMA is not tuned for physiological noise, though we and others have shown that based on its working mechanisms it shows promise of doing so. As shown here, while rare, misclassification of ICA components could substantially alter the interpretability of functional data. One alternative is to use methods, such as ICA-FIX to train and select noise components with higher accuracy and consistency. However, ICA-AROMA provides a degree of reproducibility when continuously implemented as part of a common pre-processing pipeline and does not require the training of a classifier and accompanying expertise in doing so. Consequently, adjustments could be made, using similar methods as used in this article when identifying misclassifications, to create an additional selection criterion based on overlap with known physiological noise components (Brooks et al., 2013; T. T. Liu, 2016).

Another issue of inconsistency could be related to the nature of the sequence used in this study. Brain stem imaging requires compromises to achieve relevant resolution and consistency. As the original focus was the imaging of brain stem nuclei under externally perturbated manipulation, the long TR with a relatively short runtime, hurts the formation of consistent strong ICA components (Pruim, Mennes, Buitelaar, et al., 2015; van Schuerbeek et al., 2022). As ICA component number rises as a consequence of volume number increases, it has been shown that smaller functional brain networks only separate into unique components under ideal circumstances (Cohen et al., 2021). Therefore, it could be the case here that aspects of physiological noise are inconsistently included in large components comprising mostly signals of interest.

Similar problems arise for aCompCor in this setting as well as the chosen mode of component selection. The here chosen methods of including the first five principal components created by aCompCor are detached from the explanatory power of these components and do not account for the actual total amount of variance explained. It can thus be, that these components have varying explanatory impacts between subjects. This aspect can play a large factor in the individual variation of explanatory power. An alternative method, often referred to as aCompCor50 similarly

calculated by default in fMRIPrep, includes all components up to 50% of the total explained variance (Behzadi et al., 2007; Muschelli et al., 2014). When under the assumption of white matter and CSF signal as noise this would include the most coherent temporal and spatial constructs, it would also lead to an enormous decrease in tDOF on average. When dealing with a low number of volumes as a consequence of high TR (TR = 2.02s) this could have a further detrimental impact on the ability to draw statistical inference. Other criteria, such as manual cut-off based on leveling plotted explanatory value, does both add increased labor and an undesired increase in manual input, but similarly also relies on TR, as the quality of component formation has been shown to increase with lower TRs and higher volume number (Cohen et al., 2021). While methods have been developed to create a more integrated combination of ICA-AROMA and aCompCor (van Schuerbeek et al., 2022), there has similarly been doubt cast on the assumption of white matter signal being completely devoid of grey matter signal of interest (Gawryluk et al., 2014). Based on these limitations, it is clear that ICA-AROMA and aCompCor offer an easy and convenient array of tools to account for different types of noise. However, when tackling regions, such as brain stem nuclei, that are highly sensitive to cardiac pulsations, the need for accuracy and consistency should justify the addition of RETROICOR. When choosing higher volume numbers and lower repetition times the specificity might increase, yet at a potential detrimental trade-off for usability in brain-stem imaging. Nonetheless, the relatively strong retained unique RETROICOR effect for the grey matter ROI is generally surprising (see Figure 3). However, Figure 5 shows that the retained brain wide effect of RETROICOR for some subjects, could explain the difference. Thus, while the brainstem ROI effect might be driven by expected localized high effectiveness of RETROICOR, the grey matter effect could be a consequence of overall discussed inter-subject inconsistency of ICA-AROMA and aCompCor. A conclusion also visible in higher inter-subject variation for grey matter.

However, in their working mechanisms the here tested methods are mostly focused on non-BOLD and motion-like artifacts. Except for aCompCor, which due to its potential for slow frequency component derived from white matter, ICA-AROMA based on its selection criteria and RETROICOR based on its regressors is limited in its ability to describe other types of less spatially specific physiological noise. Some of these low-frequency effects have been modeled using heart rate frequency (Shmueli et al., 2007) as well as respiratory volume (Birn et al., 2008). The evaluation of the impact of these methods is however beyond the scope of this work.

4.3. Limitations of the current study

A large aspect of possible limitations in ICA-AROMA and aCompCor inconsistency stems from the comparably high TR (2.02s) as well as a low number of volumes. While displaying the strength of external measurements as agnostic to these factors, the actual accuracy of the data-derived methods was not assessed in a competitive setting. This article only looks at impact of physiological noise in a resting state setting and thus ignores problems arising from combinations of tasks and nuisance variables. Classically used ICA-AROMA is able to remove noise, yet does so irrespective of task regressors, if not introduced into GLMs as nuisance variables. We did not use variation of a data proxy such as functional connectivity to gauge the impact of physiological noise correction. Thus, for brain stem regions it is unclear how the used correction methods affect the usability of derived data, whether being meaningful signal or dominated by noise. While it does not diminish the results of this study, it emphasizes that the statements here are aiming towards feasibility and not usability. An evaluation of possible interaction between e.g. ICA-AROMAs inconsistency in removal and formation of brain networks is beyond the scope of this article. Furthermore, within this study, we did not assess multiple sessions and thus draw no inference about inter-session stability of RETROICOR, ICA-AROMA or aCompCor nuisance regression. The conclusions drawn here are thus only applicable to a context that was based on a trade-off scenario for brain-stem imaging.

4.4. Conclusion

Overall, we show the effectiveness of RETROICOR, ICA-AROMA, and aCompCor to increase tSNR as a consequence of accounting for noise. While easily accessible methods such as ICA AROMA and aCompCor can account for most of the RETROICORs explanatory value, a small subset of information remains, mostly arising from higher intra-group consistency of noise detection. Apparent instability in ICA-AROMAs ability to detect physiological noise in this study points towards desired additional reliance on RETROICOR, even when combined with tools to tackle physiological noise such as aCompCor. Hence, adding RETROICOR can increase confidence in noise removal stability and can furthermore impact data quality when assessing brain stem areas or areas in proximity to fluid cavities.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*(FEB), 14. https://doi.org/10.3389/FNINF.2014.00014/BIBTEX

Amoore, J. N., & Ridgway, J. P. (1989). A system for cardiac and respiratory gating of a magnetic resonance imager. *Clinical Physics and Physiological Measurement : An Official Journal of the Hospital Physicists' Association, Deutsche Gesellschaft Fur Medizinische Physik and the European Federation of Organisations for Medical Physics*, *10*(3), 283–286. https://doi.org/10.1088/0143-0815/10/3/009

Beckmann, C. F., DeLuca, M., Devlin, J. T., & Smith, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1457), 1001–1013. https://doi.org/10.1098/RSTB.2005.1634

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging. *IEEE Transactions on Medical Imaging*, *23*(2), 137–152. https://doi.org/10.1109/TMI.2003.822821

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A Component Based Noise Correction Method (CompCor) for BOLD and Perfusion Based fMRI. *NeuroImage*, *37*(1), 90. https://doi.org/10.1016/J.NEUROIMAGE.2007.04.042

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1993). Respiratory sinus arrhythmia: Autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology*, *30*(2), 183–196. https://doi.org/10.1111/J.1469-8986.1993.TB01731.X

Birn, R. M., Diamond, J. B., Smith, M. A., & Bandettini, P. A. (2006). Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. *NeuroImage*, *31*(4), 1536–1548. https://doi.org/10.1016/J.NEUROIMAGE.2006.02.048

Birn, R. M., Murphy, K., & Bandettini, P. A. (2008). The effect of respiration variations on independent component analysis results of resting state functional connectivity. *Human Brain Mapping*, *29*(7), 740. https://doi.org/10.1002/HBM.20577

Brooks, J. C. W., Faull, O. K., Pattinson, K. T. S., & Jenkinson, M. (2013). Physiological Noise in Brainstem fMRI. *Frontiers in Human Neuroscience*, *0*(OCT), 623. https://doi.org/10.3389/FNHUM.2013.00623

Caballero-Gaudes, C., & Reynolds, R. C. (2017). Methods for cleaning the BOLD fMRI signal. *NeuroImage*, *154*, 128–149. https://doi.org/10.1016/J.NEUROIMAGE.2016.12.018

Chu, W. T., Mitchell, T., Foote, K. D., Coombes, S. A., & Vaillancourt, D. E. (2021). Functional imaging of the brainstem during visually-guided motor control reveals visuomotor regions in the pons and midbrain. *NeuroImage*, *226*, 117627. https://doi.org/10.1016/J.NEUROIMAGE.2020.117627

Cohen, A. D., Yang, B., Fernandez, B., Banerjee, S., & Wang, Y. (2021). Improved resting state functional connectivity sensitivity and reproducibility using a multiband multi-echo acquisition. *NeuroImage*, *225*, 117461. https://doi.org/10.1016/J.NEUROIMAGE.2020.117461

Colizoli, O., de Gee, J. W., van der Zwaag, W., & Donner, T. H. (2022). Functional magnetic resonance imaging responses during perceptual decision-making at 3 and 7 T in human cortex, striatum, and brainstem. *Human Brain Mapping*, *43*(4), 1265–1279. https://doi.org/10.1002/HBM.25719

Cox, R. W. (1996). AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173. https://doi.org/10.1006/CBMR.1996.0014

Dagli, M. S., Ingeholm, J. E., & Haxby, J. v. (1999). Localization of cardiac-induced signal change in fMRI. *NeuroImage*, *9*(4), 407–415. https://doi.org/10.1006/NIMG.1998.0424

Das, A., Murphy, K., & Drew, P. J. (2021). Rude mechanicals in brain haemodynamics: non-neural actors that influence blood flow. *Philosophical Transactions of the Royal Society B*, *376*(1815), 20190635. https://doi.org/10.1098/RSTB.2019.0635

Dreha-Kulaczewski, S., Joseph, A. A., Merboldt, K. D., Ludwig, H. C., Gärtner, J., & Frahm, J. (2015). Inspiration Is the Major Regulator of Human CSF Flow. *Journal of Neuroscience*, *35*(6), 2485–2491. https://doi.org/10.1523/JNEUROSCI.3246-14.2015

Elman, J. A., Panizzon, M. S., Hagler, D. J., Jr., Eyler, L. T., Granholm, E. L., Fennema-Notestine, C., Lyons, M. J., McEvoy, L. K., Franz, C. E., Dale, A. M., & Kremen, W. S.

(2017). Task-evoked pupil dilation and BOLD variance as indicators of locus coeruleus dysfunction. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *97*, 60. https://doi.org/10.1016/J.CORTEX.2017.09.025

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2018). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods 2018 16:1*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Gawryluk, J. R., Mazerolle, E. L., & D'Arcy, R. C. N. (2014). Does functional MRI detect activation in white matter? A review of emerging evidence, issues, and future directions. *Frontiers in Neuroscience*, *0*(8 JUL), 239. https://doi.org/10.3389/FNINS.2014.00239/ABSTRACT

German, D. C., Walker, B. S., Manaye, K., Smith, W. K., Woodward, D. J., & North, A. J. (1988). The human locus coeruleus: computer reconstruction of cellular distribution. *Journal of Neuroscience*, *8*(5), 1776–1788. https://doi.org/10.1523/JNEUROSCI.08-05-01776.1988

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., van Essen, D. C., & Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124. https://doi.org/10.1016/J.NEUROIMAGE.2013.04.127

Glover, G. H., Lemieux, S. K., Drangova, M., & Pauly, J. M. (1996). Decomposition of inflow and blood oxygen level-dependent (BOLD) effects with dual-echo spiral gradient-recalled echo (GRE) fMRI. *Magnetic Resonance in Medicine*, *35*(3), 299–308. https://doi.org/10.1002/MRM.1910350306

Glover, G. H., Li, T.-Q., & Ress, D. (2000). *Image-Based Method for Retrospective Correction of Physiological Motion Effects in fMRI: RETROICOR*. https://doi.org/10.1002/1522-2594

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, *5*, 13. https://doi.org/10.3389/FNINF.2011.00013/ABSTRACT

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/J.NEUROIMAGE.2009.06.060

Harvey, A. K., Pattinson, K. T. S., Brooks, J. C. W., Mayhew, S. D., Jenkinson, M., & Wise, R. G. (2008). Brainstem functional magnetic resonance imaging: Disentangling signal from physiological noise. *Journal of Magnetic Resonance Imaging*, *28*(6), 1337–1344. https://doi.org/10.1002/JMRI.21623

Henderson, L. A., & Macefield, V. G. (2013). Functional imaging of the human brainstem during somatosensory input and autonomic output. *Frontiers in Human Neuroscience*, *0*(SEP), 569. https://doi.org/10.3389/FNHUM.2013.00569/BIBTEX

Huotari, N., Raitamaa, L., Helakari, H., Kananen, J., Raatikainen, V., Rasila, A., Tuovinen, T., Kantola, J., Borchardt, V., Kiviniemi, V. J., & Korhonen, V. O. (2019). Sampling rate effects on resting state fMRI metrics. *Frontiers in Neuroscience*, *13*(APR), 279. https://doi.org/10.3389/FNINS.2019.00279/FULL

Hutton, C., Josephs, O., Stadler, J., Featherstone, E., Reid, A., Speck, O., Bernarding, J., & Weiskopf, N. (2011). The impact of physiological noise correction on fMRI at 7T. *NeuroImage*, *57*(1), 101–112. https://doi.org/10.1016/J.NEUROIMAGE.2011.04.018

Iyriboz, Y., Powers, S., Morrow, J., Ayers, D., & Landry, G. (1991). Accuracy of pulse oximeters in estimating heart rate at rest and during exercise. *British Journal of Sports Medicine*, *25*(3), 162. https://doi.org/10.1136/BJSM.25.3.162

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/NIMG.2002.1132

Kelly, C., Biswal, B. B., Craddock, R. C., Castellanos, F. X., & Milham, M. P. (2012). Characterizing variation in the functional connectome: promise and pitfalls. *Trends in Cognitive Sciences*, *16*(3), 181–188. https://doi.org/10.1016/J.TICS.2012.02.001

Khan, M., Pretty, C. G., Amies, A. C., Elliott, R., Shaw, G. M., & Chase, J. G. (2015). Investigating the Effects of Temperature on Photoplethysmography. *IFAC-PapersOnLine*, *48*(20), 360–365. https://doi.org/10.1016/J.IFACOL.2015.10.166

Kollmeier, J. M., Gürbüz-Reiss, L., Sahoo, P., Badura, S., Ellebracht, B., Keck, M., Gärtner, J., Ludwig, H. C., Frahm, J., & Dreha-Kulaczewski, S. (2022). Deep breathing couples CSF and venous flow dynamics. *Scientific Reports*, *12*(1). https://doi.org/10.1038/S41598-022-06361-X

Krause, F., Benjamins, C., Eck, J., Lührs, M., van Hoof, R., & Goebel, R. (2019). Active head motion reduction in magnetic resonance imaging using tactile feedback. *Human Brain Mapping*, *40*(14), 4026–4037. https://doi.org/10.1002/HBM.24683

Krüger, G., & Glover, G. H. (2001). Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magnetic Resonance in Medicine*, *46*(4), 631–637. https://doi.org/10.1002/MRM.1240

Kundu, P., Inati, S. J., Evans, J. W., Luh, W. M., & Bandettini, P. A. (2012). Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage*, *60*(3), 1759–1770. https://doi.org/10.1016/J.NEUROIMAGE.2011.12.028

Liu, T. T. (2016). Noise contributions to the fMRI signal: An overview. *NeuroImage*, *143*, 141–151. https://doi.org/10.1016/J.NEUROIMAGE.2016.09.008

Murphy, K., Birn, R. M., & Bandettini, P. A. (2013). Resting-state fMRI confounds and cleanup. *NeuroImage*, *80*, 349–359. https://doi.org/10.1016/J.NEUROIMAGE.2013.04.001

Muschelli, J., Nebel, M. B., Caffo, B. S., Barber, A. D., Pekar, J. J., & Mostofsky, S. H. (2014). Reduction of motion-related artifacts in resting state fMRI using aCompCor. *NeuroImage*, *96*, 22–35. https://doi.org/10.1016/J.NEUROIMAGE.2014.03.028

Pruim, R. H. R., Mennes, M., Buitelaar, J. K., & Beckmann, C. F. (2015). Evaluation of ICA-AROMA and alternative strategies for motion artifact removal in resting state fMRI. *NeuroImage*, *112*, 278–287. https://doi.org/10.1016/J.NEUROIMAGE.2015.02.063

Pruim, R. H. R., Mennes, M. D., van Rooij, D., Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*, *112*, 267–277. https://doi.org/10.1016/J.NEUROIMAGE.2015.02.064

Raj, D., Anderson, A. W., & Gore, J. C. (2001). Respiratory effects in human functional magnetic resonance imaging due to bulk susceptibility changes. *Physics in Medicine and Biology*, *46*(12), 3331–3340. https://doi.org/10.1088/0031-9155/46/12/318

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., & Smith, S. M. (2014). Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, *90*, 449–468. https://doi.org/10.1016/J.NEUROIMAGE.2013.11.046

Sasaki, M., Shibata, E., Tohyama, K., Takahashi, J., Otsuka, K., Tsuchiya, K., Takahashi, S., Ehara, S., Terayama, Y., & Sakai, A. (2006). Neuromelanin magnetic resonance imaging of locus ceruleus and substantia nigra in Parkinson's disease. *Neuroreport*, *17*(11), 1215–1218. https://doi.org/10.1097/01.WNR.0000227984.84927.A7

Schwarz, L. A., & Luo, L. (2015). Organization of the Locus Coeruleus-Norepinephrine System. *Current Biology*, *25*(21), R1051–R1056. https://doi.org/10.1016/J.CUB.2015.09.039

Shmueli, K., van Gelderen, P., de Zwart, J. A., Horovitz, S. G., Fukunaga, M., Jansma, J. M., & Duyn, J. H. (2007). Low Frequency Fluctuations in the Cardiac Rate as a Source of Variance in the Resting-State fMRI BOLD Signal. *NeuroImage*, *38*(2), 306. https://doi.org/10.1016/J.NEUROIMAGE.2007.07.037

Singh, K., García-Gomar, M. G., Cauzzo, S., Staab, J. P., Indovina, I., & Bianciardi, M. (2022). Structural connectivity of autonomic, pain, limbic, and sensory brainstem nuclei in living humans based on 7 Tesla and 3 Tesla MRI. *Human Brain Mapping*, *43*(10), 3086–3112. https://doi.org/10.1002/HBM.25836

Tona, K.-D., Keuken, M. C., de Rover, M., Lakke, E., Forstmann, B. U., Nieuwenhuis, S., & van Osch, M. J. P. (2017). In vivo visualization of the locus coeruleus in humans: quantifying the test?retest reliability. *Brain Structure and Function*. https://doi.org/10.1007/s00429-017-1464-5

Triantafyllou, C., Hoge, R. D., Krueger, G., Wiggins, C. J., Potthast, A., Wiggins, G. C., & Wald, L. L. (2005). Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. *NeuroImage*, *26*(1), 243–250. https://doi.org/10.1016/J.NEUROIMAGE.2005.01.007

van Buuren, M., Gladwin, T. E., Zandbelt, B. B., van den Heuvel, M., Ramsey, N. F., Kahn, R. S., & Vink, M. (2009). Cardiorespiratory effects on default-mode network activity as measured with fMRI. *Human Brain Mapping*, *30*(9), 3031–3042. https://doi.org/10.1002/HBM.20729

van de Moortele, P. F., Pfeuffer, J., Glover, G. H., Ugurbil, K., & Hu, X. (2002). Respiration-induced B0 fluctuations and their spatial distribution in the human brain at 7 Tesla. *Magnetic Resonance in Medicine*, *47*(5), 888–895. https://doi.org/10.1002/MRM.10145

van Schuerbeek, P., de Wandel, L., & Baeken, C. (2022). The optimized combination of aCompCor and ICA-AROMA to reduce motion and physiologic noise in task fMRI data. *Biomedical Physics & Engineering Express*, *8*(5). https://doi.org/10.1088/2057-1976/AC63F0

Windischberger, C., Langenberger, H., Sycha, T., Tschernko, E. M., Fuchsjäger-Mayerl, G., Schmetterer, L., & Moser, E. (2002). On the origin of respiratory artifacts in BOLD-EPI of the human brain. *Magnetic Resonance Imaging*, *20*(8), 575–582. https://doi.org/10.1016/S0730-725X(02)00563-5

Wise, R. G., Ide, K., Poulin, M. J., & Tracey, I. (2004). Resting fluctuations in arterial carbon dioxide induce significant low frequency variations in BOLD signal. *NeuroImage*, *21*(4), 1652–1664. https://doi.org/10.1016/J.NEUROIMAGE.2003.11.025

Zuo, X. N., Biswal, B. B., & Poldrack, R. A. (2019). Editorial: Reliability and reproducibility in functional connectomics. *Frontiers in Neuroscience*, *13*(FEB), 117. https://doi.org/10.3389/FNINS.2019.00117/BIBTEX

Figure 1

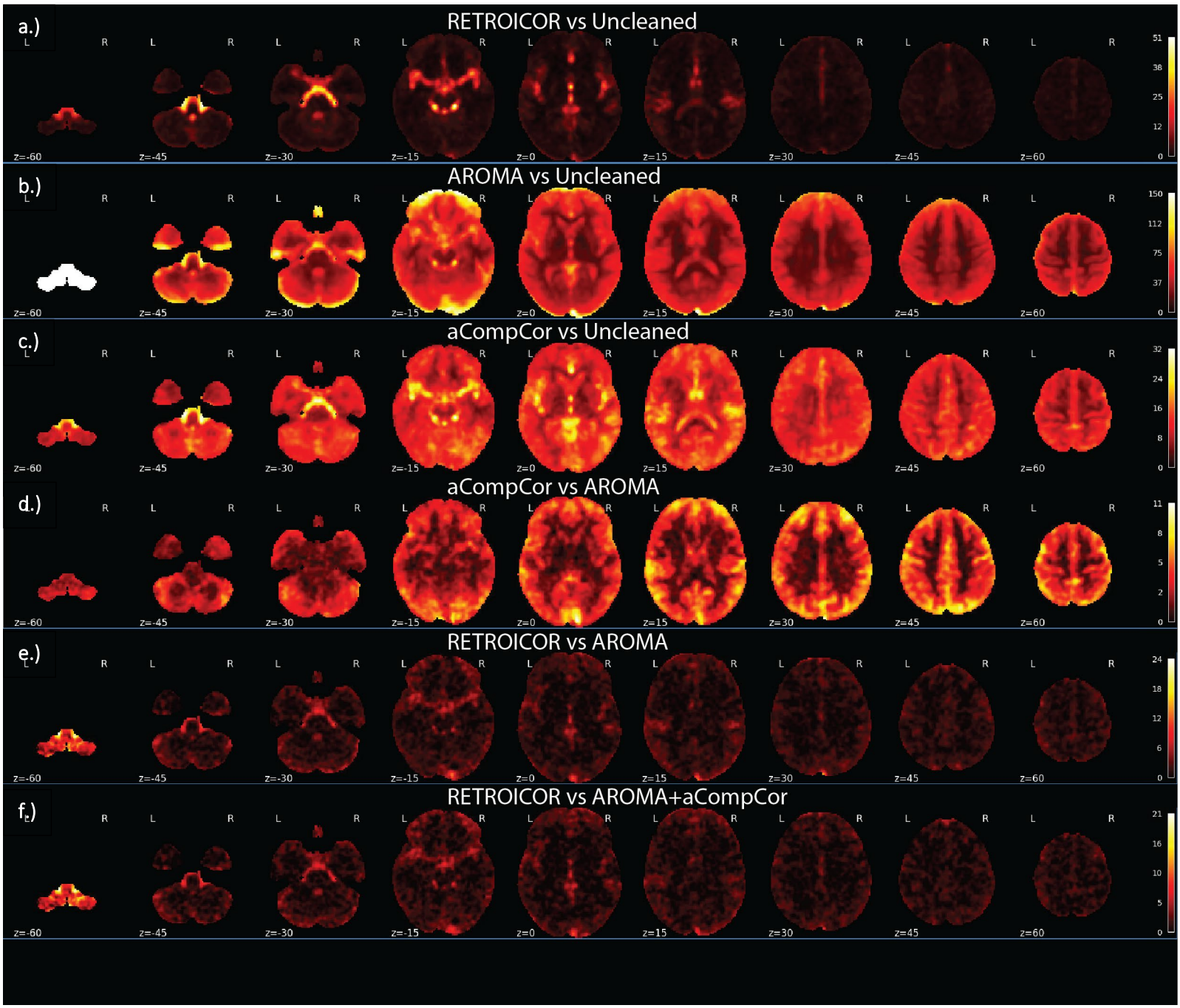

Figure 1. Mean tSNR maps visualizing relative improvement of tSNR compared to uncleaned data for all relevant comparisons. In section b.) the threshold was capped at 150, to allow for pattern visibility. Inferior slices exhibited otherwise extreme improvement due to correction of magnetization effects.

Figure 2

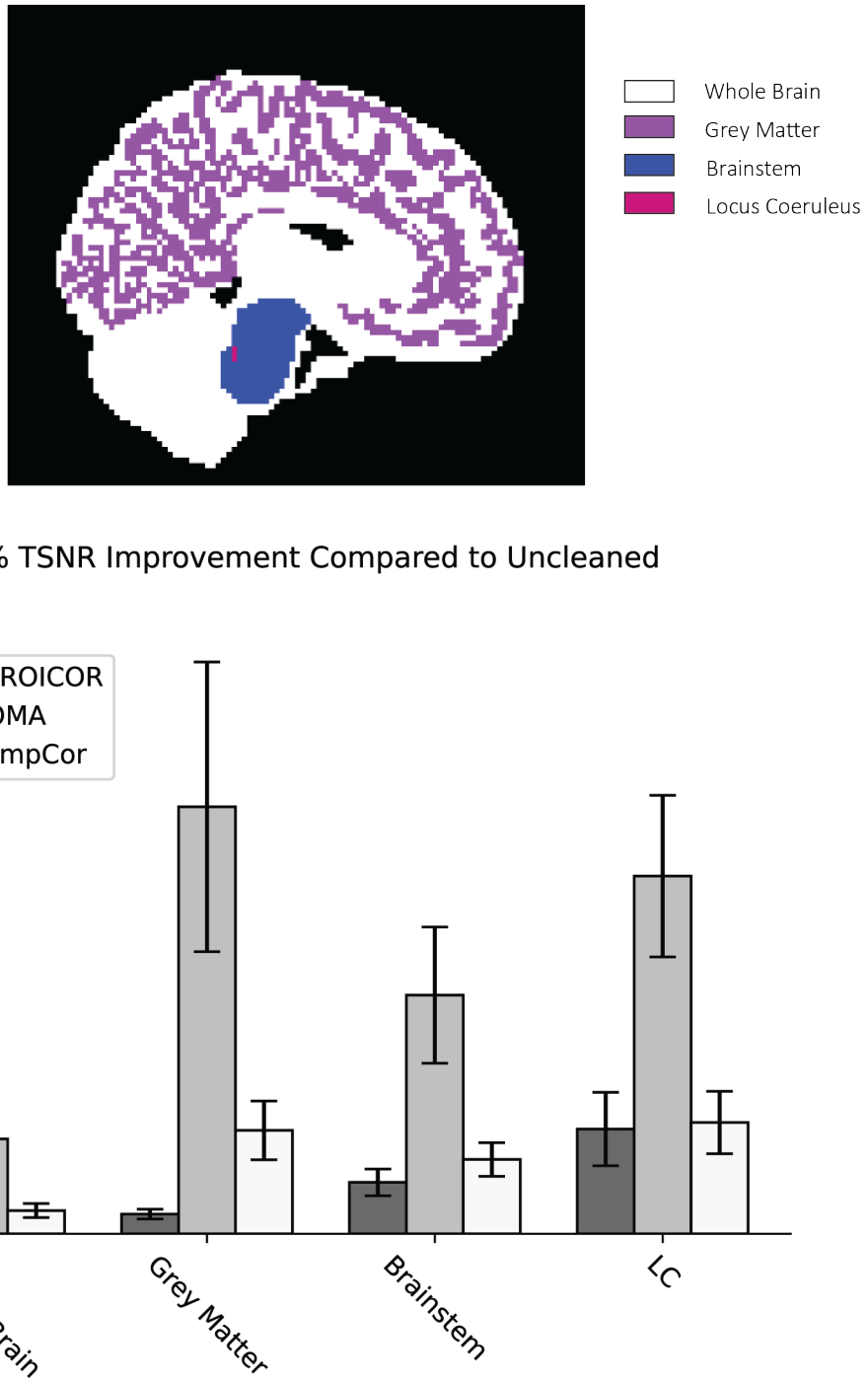

Figure 2. TSNR improvement as a consequence of nuisance regression of RETROICOR, AROMA, and aCompCor in percent. Error bars represent 95% confidence intervals.

Figure 3

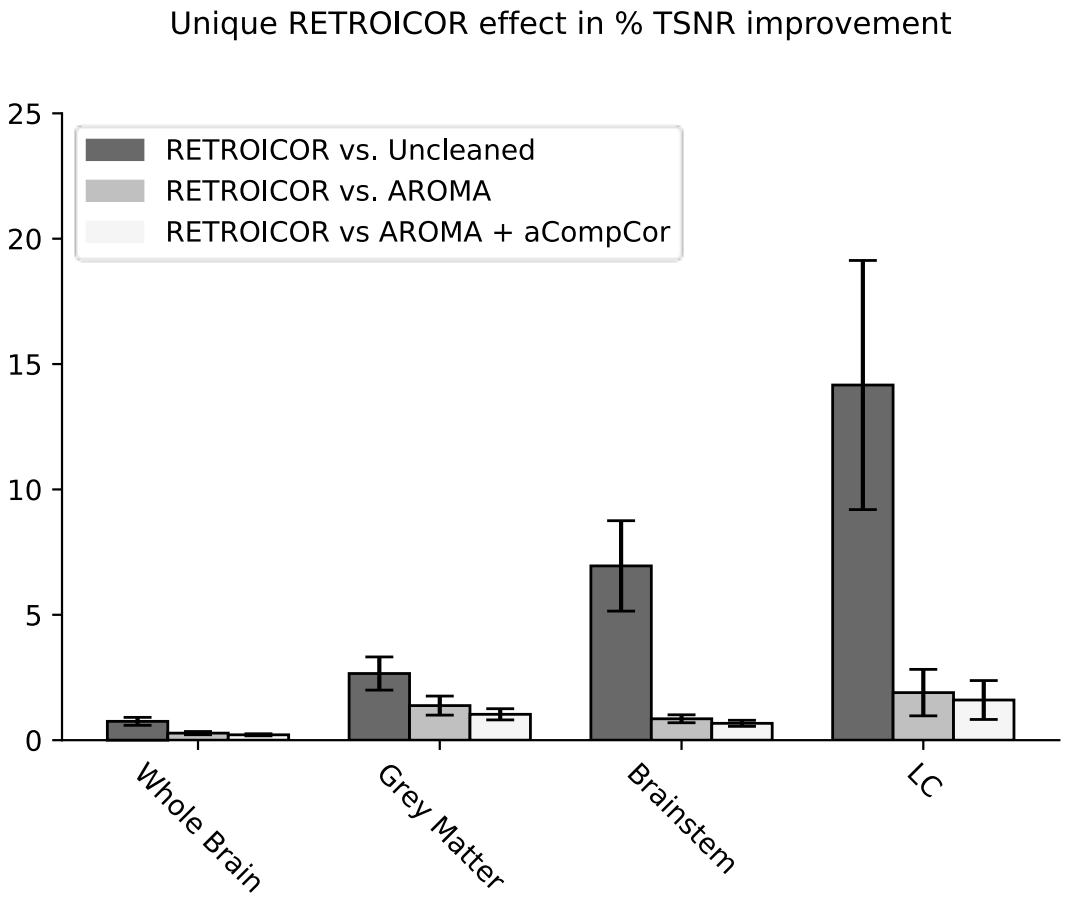

Figure 3. TSNR improvement of RETROICOR nuisance regression without other models, with AROMA and with AROMA + aCompCor in Percent. Error bars represent 95% confidence intervals.

Figure 4

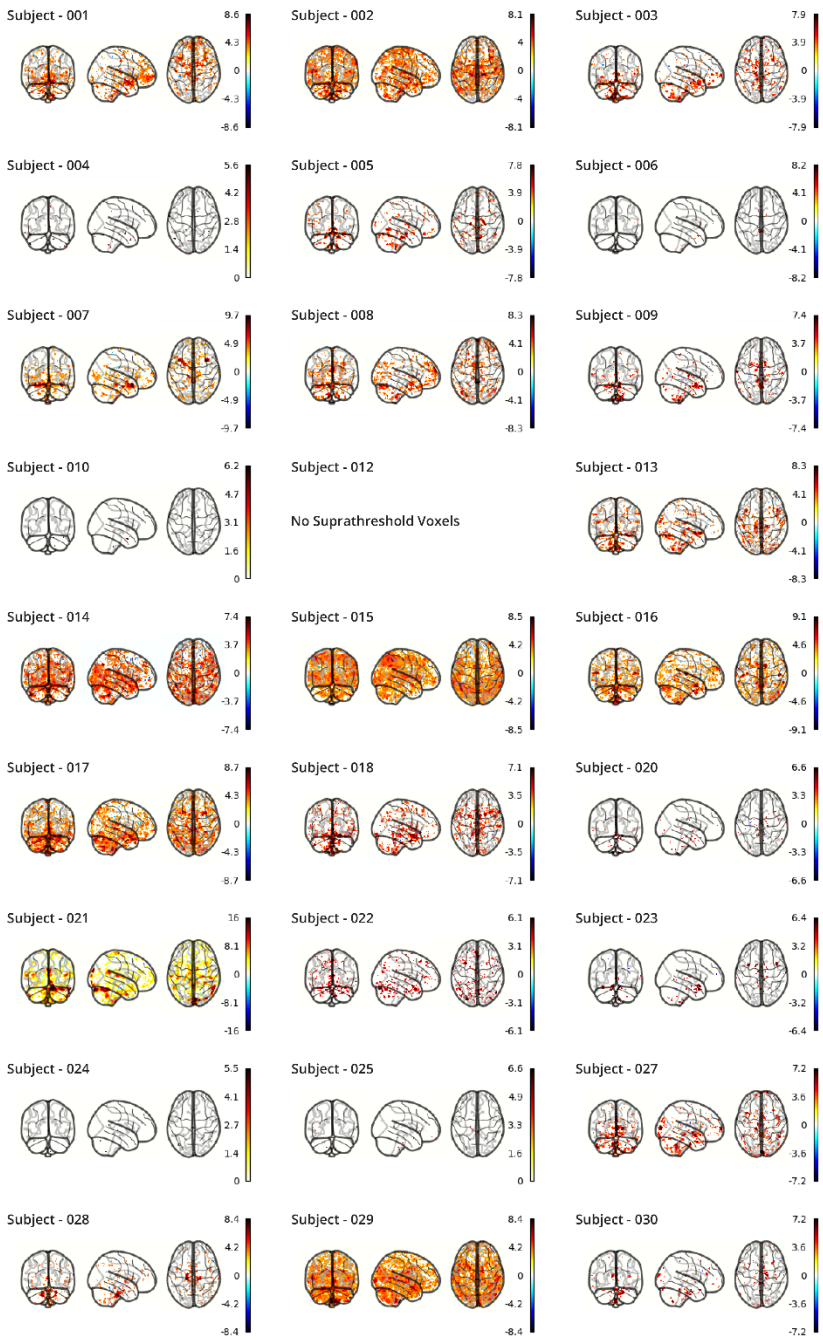

Figure 4. Subject-wise spatial representation of individually explained variance uniquely attributed to RETROICOR with AROMA in the model (thresholded at p_{FDR} <.05).

Figure 5

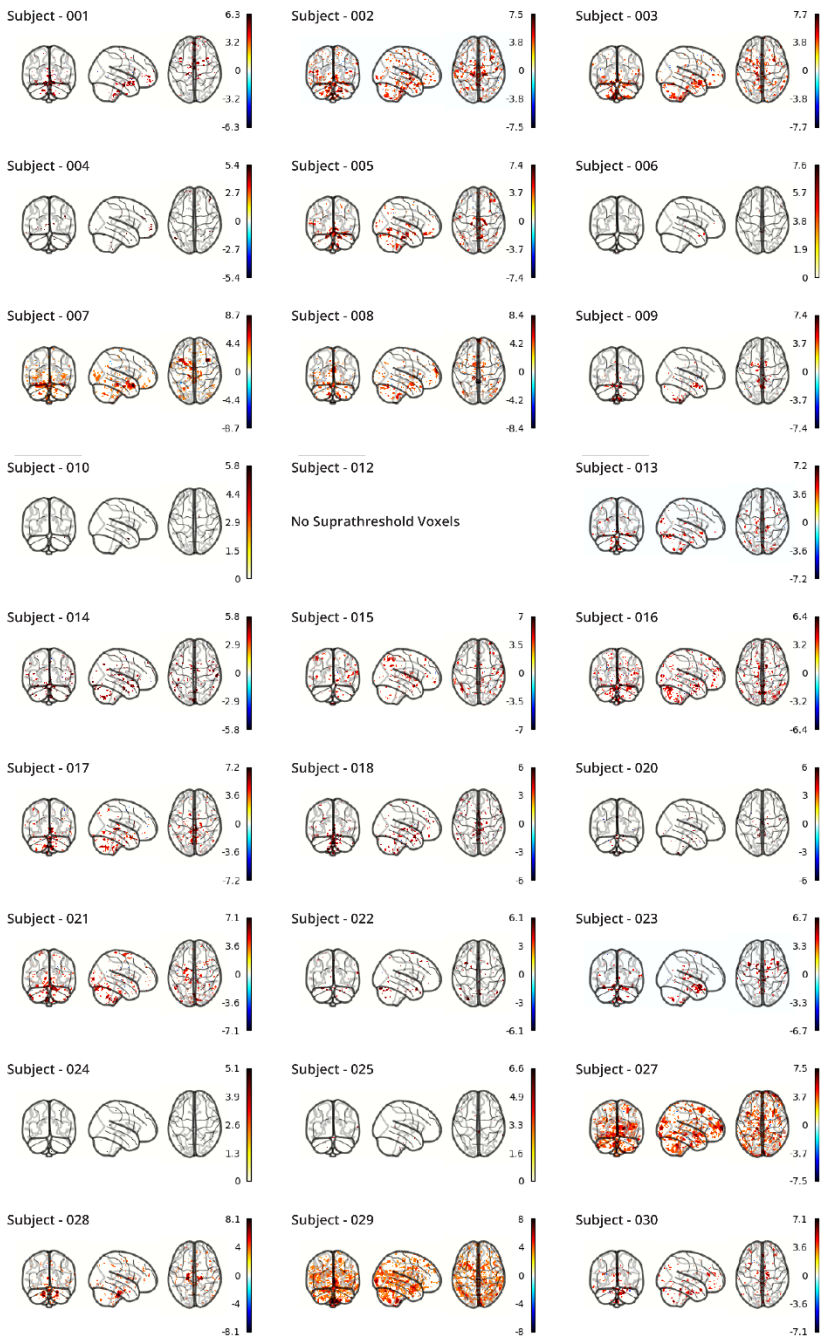

Figure 5. Subject-wise spatial representation of individually explained variance uniquely attributed to RETROICOR with AROMA and aCompCor in the model (thresholded at p_{FDR} <.05).

Figure 6

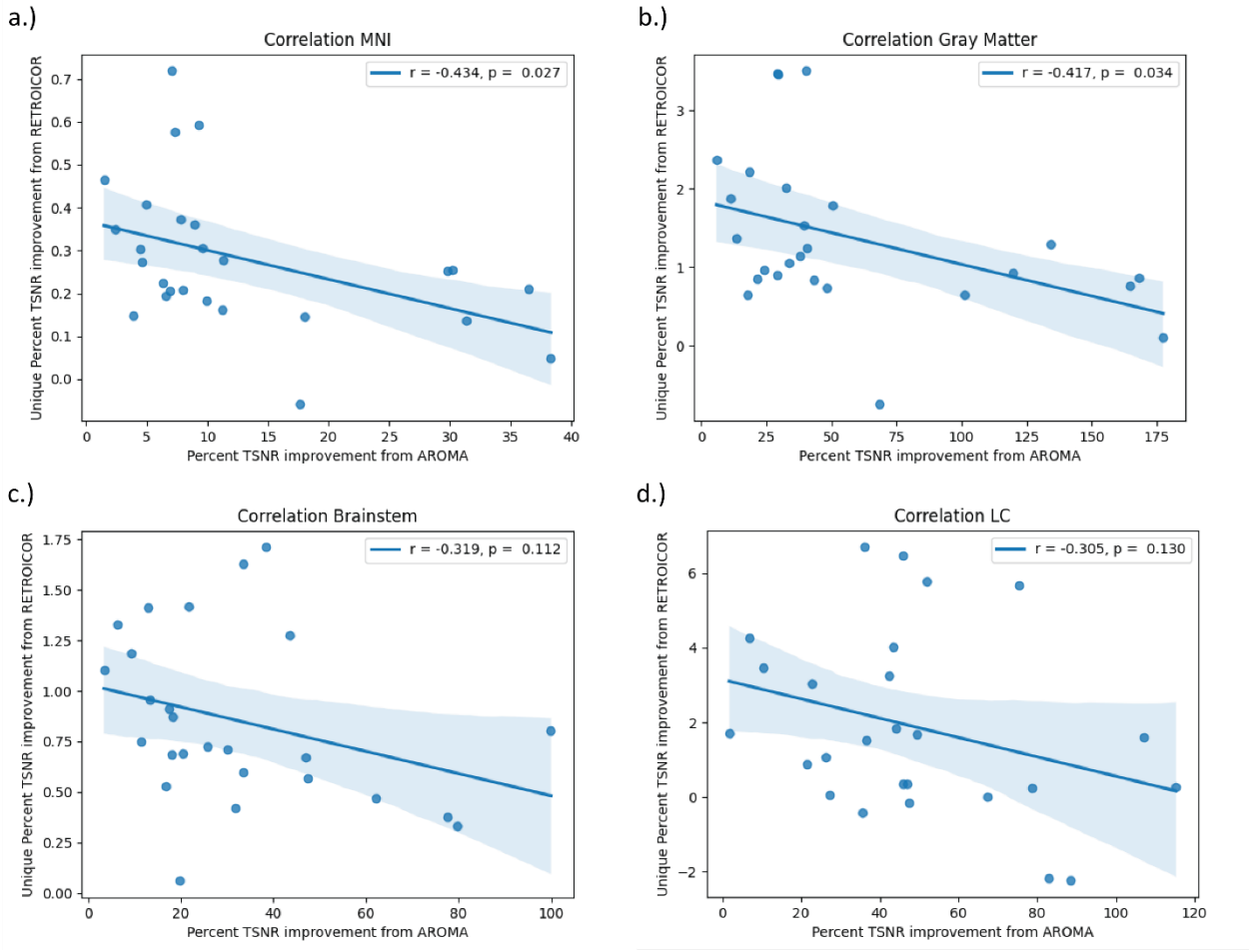

Figure 6. Correlation between unique improvement of RETROICOR and improvement of ICA-AROMA on tSNR for each of the four tested ROIs. Error band represents 95% confidence intervals.

Figure 7

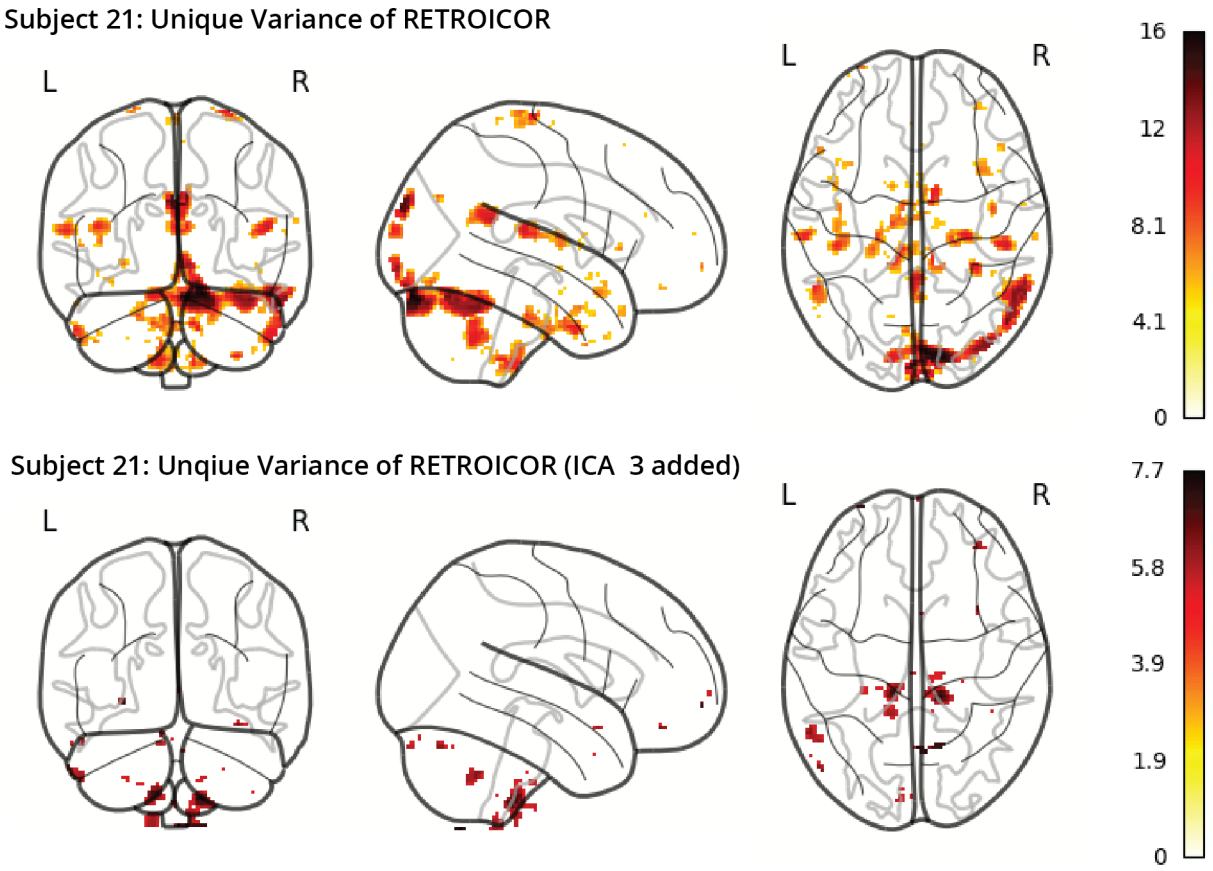

Figure 7. Unique explained variance of RETROICOR after adding a misclassified ICA component to the array of AROMA regressors for subject 021.