ICA-based Denoising Strategies in Breath-Hold Induced Cerebrovascular Reactivity Mapping with Multi Echo BOLD fMRI.

Stefano Moia[a,b,*], Maite Termenon[a], Eneko Uruñuela[a], Rachael C. Stickland[c], Molly G. Bright[c,d], and César Caballero-Gaudes[a]

a. Basque Center on Cognition, Brain and Language, Donostia, Spain
b. University of the Basque Country (UPV/EHU), Donostia, Spain
c. Physical Therapy and Human Movement Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL
d. Biomedical Engineering, McCormick School of Engineering, Northwestern University, Evanston, IL

* Corresponding author: s.moia@bcbl.eu

Abstract

Performing a BOLD functional MRI (fMRI) acquisition during breath-hold (BH) tasks is a non-invasive, robust method to estimate cerebrovascular reactivity (CVR). However, movement and breathing-related artefacts caused by the BH can substantially hinder CVR estimates due to their high temporal collinearity with the effect of interest, and attention has to be paid when choosing which analysis model should be applied to the data. In this study, we evaluate the performance of multiple analysis strategies based on lagged general linear models applied on multi-echo BOLD fMRI data, acquired in ten subjects performing a BH task during ten sessions, to obtain subject-specific CVR and haemodynamic lag estimates. The evaluated approaches range from conventional regression models including drifts and motion timecourses as nuisance regressors applied on single-echo or optimally-combined data, to more complex models including regressors obtained from multi-echo independent component analysis with different grades of orthogonalization in order to preserve the effect of interest, i.e. the CVR. We compare these models in terms of their ability to make signal intensity changes independent from motion, as well as the reliability as measured by voxelwise intraclass correlation coefficients of both CVR and lag maps over time. Our results reveal that a conservative independent component analysis model applied on the optimally-combined multi-echo fMRI signal offers the largest reduction of motion-related effects in the signal, while yielding reliable CVR amplitude and lag estimates, although a conventional regression model applied on the optimally-combined data results in similar estimates. This work demonstrate the usefulness of multi-echo based fMRI acquisitions and independent component analysis denoising for precision mapping of CVR in single subjects based on BH paradigms, fostering its potential as a clinically-viable neuroimaging tool for individual patients. It also proves that the way in which

data-driven regressors should be incorporated in the analysis model is not straight-forward due to their complex interaction with the BH-induced BOLD response.

Keywords

Cerebrovascular reactivity, breath-hold, multi-echo fMRI, independent component analysis, denoising, precision functional mapping

1 Introduction

Cerebrovascular reactivity (CVR) is a physiological response of the cerebral vessels to vasodilatory or vasoconstrictive stimuli. Mapping of the CVR response provides an important indicator of cerebrovascular health. In recent years, functional magnetic resonance imaging (fMRI), either based on the blood oxygenation level-dependent (BOLD) contrast, arterial spin labelling, or a mixture of both, has demonstrated its effectiveness as a method to assess CVR. As a result, its use is spreading into clinical practice, where its potential as a diagnostic measure is being ascertained in different diseases, spanning from vascular diseases (Hartkamp, Bokkers, van Osch, de Borst, & Hendrikse, 2017; Markus & Cullinane, 2001; Webster et al., 1995; Ziyeh et al., 2005), to stroke and aphasia (Krainik, Hund-Georgiadis, Zysset, & Von Cramon, 2005; Van Oers et al., 2018), brain tumors (Fierstra et al., 2018; Zacà, Jovicich, Nadar, Voyvodic, & Pillai, 2014), neurodegenerative diseases (Camargo et al., 2015; Glodzik, Randall, Rusinek, & de Leon, 2013; Marshall et al., 2014), hypertension (Iadecola & Davisson, 2008; Leoni et al., 2011; Tchistiakova, Anderson, Greenwood, & Macintosh, 2014), lifestyle habits (Friedman et al., 2008; Gonzales et al., 2014), sleep apnea (Buterbaugh et al., 2015; Prilipko, Huynh, Thomason, Kushida, & Guilleminault, 2014), and traumatic brain injury or concussions (Churchill, Hutchison, Graham, & Schweizer, 2020; Markus & Cullinane, 2001).

CVR measurements are obtained by evoking a vasodilatory response during imaging. This is typically done by injecting intravenous acetazolamide, or by exposing the subject to gas challenges with computerised dynamic deployment of CO_2 and O_2. However, acetazolamide is an invasive technique not indicated for vulnerable subjects (e.g. elderly or children), while gas challenges require dedicated setups and can also cause discomfort in some subjects, which might induce anxiety and thus potentially bias CVR measurement (Urback, MacIntosh, & Goldstein, 2017). A valid alternative is the use of voluntary respiratory challenges, such as paced deep breathing or breath-hold (BH) tasks (Bright, Bulte, Jezzard, & Duyn, 2009; Kastrup, Li, Takahashi, Glover, & Moseley, 1998). In fact, it has been shown that CO_2 changes in the blood due to breathing tasks elicit a CVR response that is equivalent to that of inhaled CO_2 (Kastrup, Krüger, Neumann-

Haefelin, & Moseley, 2001; Tancredi & Hoge, 2013). A BH task can be successfully implemented in young children and elderly subjects (Handwerker, Gazzaley, Inglis, & D'Esposito, 2007; Thomason, Burrows, Gabrieli, & Glover, 2005), and it is a robust measurement even if subjects are not able to hold their breath for as long as instructed (Bright & Murphy, 2013a). Moreover, BH-induced CVR has been shown to be a reliable measurement across different sessions of MRI, both in the short (same day) and long term (Peng, Yang, Chen, & Shih, 2019), in terms of spatial reliability (i.e. comparing variability of voxels across multiple sessions in one subject) and general reliability (i.e. average CVR value across sessions and within subjects) (Lipp, Murphy, Caseras, & Wise, 2015; Magon et al., 2009). Both short and long term reliability of BH-induced CVR were found to be comparable to that of other non-invasive means of estimating CVR, such as resting state fMRI (P. Liu et al., 2017), inhaled gas challenges (Dengel et al., 2017; Evanoff et al., 2020; Leung, Kim, & Kassner, 2016), Fourier modelling of a BH task (Pinto, Jorge, Sousa, Vilela, & Figueiredo, 2016), and a paced deep breathing task (Sousa, Vilela, & Figueiredo, 2014).

However, BOLD fMRI data exhibit signal variation arising from different sources, most of which corresponds to hardware-related artefacts and drifts, head motion, confounding physiological fluctuations, and other sources of noise (Bianciardi et al., 2009; Jorge, Figueiredo, van der Zwaag, & Marques, 2013). It is important that the signal variance associated with these confounding signals is accounted for and minimized during preprocessing or data analyses (Caballero-Gaudes & Reynolds, 2017; T. T. Liu, 2016). The artefacts induced by voluntary and involuntary movement are a particularly problematic source of noise for task-based fMRI experiments, mainly in block designs where large head movement leads to bias in estimates of the task-induced activity (Johnstone et al., 2006) and in particular experimental paradigms, such as in overt speech production where the articulation of words makes head movement considerable (Barch et al., 1999; Soltysik & Hyde, 2006; Xu et al., 2014). This concern with task-induced movement artefacts extends to respiration tasks: the experimental design is similar to that of block designs, but the amount of motion associated with paced breathing, deep breaths, or "recovery" breaths following a BH task can be very prominent and concur with the pattern of the task. Moreover, respiration can perturb the B0 field due to the change of air in the lungs (Raj, Anderson, & Gore, 2001) and introduce aliasing artefacts or pseudo-movement effects in the signal (Gratton et al., 2020; Pais-Roldán, Biswal, Scheffler, & Yu, 2018; Power, Lynch, et al., 2019).

There are different ways to account for motion effects on task-based fMRI data analysis. For instance, such effects can be reduced during acquisition by implementing an event-related task paradigm (Birn, Bandettini, Cox, & Shaker, 1999; Birn, Cox, & Bandettini, 2004). However, in a BH task the periods of apnoea are typically between 10 and 20 seconds in duration to achieve a robust and reproducible vasodilatory response (Bright & Murphy, 2013a; Magon et al., 2009), and

are not readily adapted to a brief event-related design. The most straight-forward approach is then to include the realignment translation and rotation parameters in the analysis model (Friston, Williams, Howard, Frackowiak, & Turner, 1996). Treating such timecourses, as well as their derivatives and non-linear expansions, as regressors of non-interest in the regression model can account for part of the motion-related variance of the signal, thus improving the estimation of the task effects. Another widely adopted method to remove motion-related effects, as well as noise in general, is to decompose the fMRI data using Principal Component Analysis or Independent Component Analysis (ICA) in order to identify and remove components that are mostly related to motion or other sources of noise (Behzadi, Restom, Liau, & Liu, 2007; Griffanti et al., 2014; Muschelli et al., 2014; R. H. R. Pruim, Mennes, Buitelaar, & Beckmann, 2015; R. H. R. Pruim, Mennes, Rooij, et al., 2015; Salimi-Khorshidi, Smith, & Nichols, 2011).

Alternatively, noise in fMRI can be reduced by using multi-echo (ME) acquisitions that sample the data at multiple successive echo times (TE). A weighted combination of the multiple echoes based on each voxel's $T_2^\star$ value (Posse et al., 1999) or temporal signal-to-noise ratio (Poser, Versluis, Hoogduin, & Norris, 2006) can smear out random noise and enhance the sensitivity to the BOLD contrast. In fact, compared with single-echo data, this optimal combination can improve the mapping of neuronal activity at 3 Tesla (Fernandez, Leuchs, Sämann, Czisch, & Spoormaker, 2017) and 7T (Puckett et al., 2018), with results comparable to other preprocessing techniques requiring extra data such as RETROICOR (Atwi et al., 2018). Optimal combination of multiple echo volumes can also improve BH-induced CVR mapping sensitivity, specificity, repeatability and reliability (Cohen & Wang, 2019).

Furthermore, assuming monoexponential decay, the voxelwise fMRI signal (S) can be expressed in signal percentage change as:

$$\frac{S-\bar{S}}{\bar{S}}=\Delta\rho-TE\cdot\Delta R_2^\star+n\ , \tag{1}$$

where $\Delta\rho$ represents non-BOLD related changes in net magnetisation, $\Delta R_2^\star$ represents BOLD-related susceptibility changes, and *n* denotes the random noise (Kundu, Inati, Evans, Luh, & Bandettini, 2012). As the BOLD-related signal can be expressed as a function of the TE, whereas noise-related non-BOLD changes in the net magnetization are independent of TE, the information available in multiple echoes can be leveraged for the purpose of denoising. For example, in a dual-echo acquisition where the first TE is sufficiently short, the first echo signal mainly captures changes in $\Delta\rho$ rather than in $\Delta R_2^\star$. It is then possible to remove artefactual effects, through voxelwise regression, from the second echo signal acquired at a longer TE with appropriate BOLD contrast (Bright & Murphy, 2013b). Collecting more echoes opens up the possibility of applying

ICA and classifying independent components into BOLD-related (i.e. describing ΔR_2^* fluctuations with a linear TE-dependency) or noise (i.e. independent of TE, related to non-BOLD fluctuations in the net magnetization $\Delta\rho$), an approach known as multi-echo independent component analysis (ME-ICA) (Kundu et al., 2013, 2012, 2017). Compared to single-echo data denoising, ME-ICA can improve the mapping of task-induced activation (DuPre, Luh, & Spreng, 2016; Gonzalez-Castillo et al., 2016; Lombardo et al., 2016), for example in challenging paradigms with slow-varying stimuli (Evans, Kundu, Horovitz, & Bandettini, 2015) or language mapping and laterality (Amemiya, Yamashita, Takao, & Abe, 2019). It also outperforms single-echo ICA-based denoising of resting-state fMRI data (Dipasquale et al., 2017), which is particularly beneficial more efficient and reliable functional connectivity mapping in individual subjects (Lynch, Power, Dubin, Gunning, & Liston, 2020) and in brain regions where traditional single-echo acquisitions offer reduced signal-to-noise ratio, such as the basal forebrain (Markello, Spreng, Luh, Anderson, & De Rosa, 2018). Furthermore, ME-ICA also enhances the deconvolution of neuronal-related signal changes (Caballero-gaudes, Moia, Panwar, Bandettini, & Gonzalez-castillo, 2019).

However, up to now, the operation of ME-ICA has not been evaluated thoroughly in experimental paradigms with unavoidable task-correlated artefacts. Under such scenarios, one question that remains open is how to obtain the right trade-off between removing as much noise as possible and saving as much signal of interest as possible (Bright & Murphy, 2015; Griffanti et al., 2014). In this study, we acquire ME-fMRI data during a BH task in a precision functional mapping experiment (Gordon et al., 2017) and assess the efficiency of different nuisance regression models to remove artefacts that are highly correlated with the effect of interest, i.e. the CVR response. In particular, we compare traditional nuisance regression approaches, applied to single- or multi-echo data, and three different ME-ICA denoising approaches ranging from aggressive to conservative. For each denoising strategy, we assess the correlation of the cleaned signal with measures of motion, and evaluate the amplitude and lag of the CVR signal response in terms of their physiological interpretability and inter-session reliability.

2 Material and methods

2.1 Participants

Ten healthy subjects with no record of psychiatric or neurological disorders (5F, age range 24-40 y at the start of the study) underwent ten MRI sessions in a 3T Siemens PrismaFit scanner with a 64-channel head coil. Each session took place one week apart, on the same day of the week and at the same time of the day.

All participants had to meet several further requirements, i.e. being non-smokers and refrain from smoking for the whole duration of the experiment, and not suffering from respiratory or cardiac

health issues. They were also instructed to refrain from consuming caffeinated drinks for two hours before the session. Informed consent was obtained before each session, and the study was approved by the local ethics committee.

2.2 Data acquisition and MRI session

Within the MRI session, subjects performed a BH task while T2*-weighted ME-fMRI data was acquired with the simultaneous multislice (a.k.a. multiband, MB) gradient-echo planar imaging sequence provided by the Center for Magnetic Resonance Research (CMRR, Minnesota) (Moeller et al., 2010; Setsompop et al., 2012) with the following parameters: 340 scans, TR = 1.5 s, TEs = 10.6/28.69/46.78/64.87/82.96 ms, flip angle = 70°, MB acceleration factor = 4, GRAPPA = 2 with Gradient-echo reference scan, 52 slices with interleaved acquisition, Partial-Fourier = 6/8, FoV = 211x211 mm^2, voxel size = 2.4x2.4x3 mm^3, Phase Encoding = AP, bandwidth=2470 Hz/px, LeakBlock kernel reconstruction (Cauley, Polimeni, Bhat, Wald, & Setsompop, 2014) and SENSE coil combination (Sotiropoulos et al., 2013). Single-band reference (SBRef) images were also acquired for each TE. The BH task was preceded by 64 minutes of ME-fMRI scanning, consisting of three task-based and four 10-minute resting state acquisitions, which are not part of the current study. The BH task always followed a resting state run.. A pair of Spin Echo echo planar images (EPI) with opposite phase-encoding (AP or PA) directions and identical volume layout (TR = 2920 ms, TE = 28.6 ms, flip angle = 70°) were also acquired before each functional run in order to estimate field distortions. A T1-weighted MP2RAGE image (Marques et al., 2009) (TR = 5 s, TE = 2.98 ms, TI1 = 700 ms, TI2 = 2.5 s,, flip angle 1 = 4°, flip angle 2 = 5°, GRAPPA = 3, 176 slices, FoV read = 256 mm, voxel size = 1x1x1 mm^3, TA = 662 s) and a T2-weighted Turbo Spin Echo image (Hennig, Nauerth, & Friedburg, 1986) (TR = 3.39 s, TE = 389 ms, GRAPPA = 2, 176 slices, FoV read = 256 mm, voxel size = 1x1x1 mm^3, TA = 300 s) were also collected at the end and at the beginning of each MRI session, respectively.

During the fMRI acquisition runs exhaled CO_2 and O_2 levels were monitored and recorded using a nasal cannula (Intersurgical) with an ADInstruments ML206 gas analyser unit and transferred to a BIOPAC MP150 physiological monitoring system where scan triggers were simultaneously recorded. Photoplethismography and respiration effort data were also measured via the BIOPAC system, but these physiological signals were not used in the current study. All signals were sampled at 10 kHz.

2.3 Breath-hold task

Following Bright and Murphy (2013a), the BH paradigm consisted of eight repetitions of a BH trial composed of four paced breathing cycles of 6 s each, an apnoea (BH) of 20 s, an exhalation of 3 s, and 11 s of “recovery” breathing (unpaced) (i.e. total trial duration of 58 s) (Figure 1). Subjects

were instructed prior to scanning about the importance of the exhalations preceding and following the apnoea. Without these exhalations providing CO_2 measurements, the change in systemic CO_2 levels achieved by each BH cannot be robustly estimated; as a result, CVR (%BOLD/mmHg CO_2 change) cannot be estimated quantitatively. Participants were instructed textually throughout the task through a mirror screen located in the head coil.

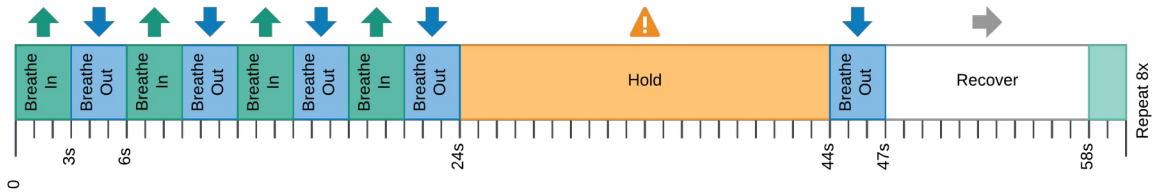

Figure 1: Schematic of Breath-Hold trial. Apnoea was preceded and followed by exhalations.

2.4 MRI data preprocessing

The DICOM files of the MRI data were transformed into nifti files with dcm2nii (Li, Morgan, Ashburner, Smith, & Rorden, 2016) and formatted into Brain Imaging Data Structure (Gorgolewski et al., 2016) with heudiconv (Halchenko et al., 2019).

MRI data were minimally preprocessed with custom scripts based mainly in FSL (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012), AFNI (Cox, 1996), and ANTs (Tustison et al., 2014). In brief, the T2-weighted image was skull-stripped and co-registered to the MP2RAGE image along with the brain mask. The latter was applied to the MP2RAGE image, that then was segmented into gray matter (GM), white matter and cerebrospinal fluid tissues (Avants, Tustison, Wu, Cook, & Gee, 2011). The MP2RAGE image was normalised to an asymmetric version of the MNI152 6th generation template at 1 mm resolution (Grabner et al., 2006), while the T2-weighted volume was co-registered to the skull-stripped single-band reference image (SBRef) of the first echo. The first 10 volumes of the functional data were discarded to allow the signal to achieve a steady state of magnetisation. Image realignment to the SBRef was computed on the first echo, and the estimated rigid-body spatial transformation was then applied to all other echoes (Jenkinson, Bannister, Brady, & Smith, 2002; Jenkinson & Smith, 2001). A brain mask obtained from the SBRef volume was applied to all the echoes. The different echo timeseries were optimally combined (OC) voxelwise by weighting each timeseries contribution by its T_2^* value (Posse et al., 1999). Next, ME-ICA decomposition was performed with tedana (DuPre et al., 2019) using the minimum description length criterion for estimation of the number of components (Harris, 1978; Li et al., 2016). The independent components (ICs) were then manually classified by SM and CCG into two categories (rejected or accepted components) based on temporal, spatial, spectral and TE-

dependence features of each component (Griffanti et al., 2017) (see Supplementary materials). Finally, a distortion field correction was performed on the OC volume using Topup (Andersson, Skare, & Ashburner, 2003), and the BOLD timeseries was converted in signal percentage change. For comparison, the dataset acquired at the second echo time (TE_2 = 28.6 ms) was used as a surrogate for standard single-echo (SE) acquisitions. This volume followed the same preprocessing steps as the OC volume, except for the optimal combination and the ICA decomposition.

2.5 CO_2 trace processing and CVR estimation

The files exported from the AcqKnowledge software were transformed and formatted into BIDS with phys2bids (The phys2bids developers et al., 2019). Then, the CVR estimation followed the same steps as those reported in Moia, Stickland, and colleagues (2020).

The CO_2 timecourse was processed using custom scripts in Python 3.6.7. Briefly, the CO_2 timecourse was downsampled to 40 Hz to reduce computational costs. The end-tidal peaks were automatically and manually individuated, then the amplitude envelope was demeaned and convolved with a canonical HRF to obtain the $P_{ET}CO_2$*hrf* trace. In order to account for measurement delay, the $P_{ET}CO_2$*hrf* trace was shifted to maximise the cross-correlation with the average timecourse of an eroded version of the GM mask (bulk shift) (Yezhuvath, Lewis-Amezcua, Varghese, Xiao, & Lu, 2009). This step was performed on both OC and the SE data (see Supplementary figure 1).

A lagged general linear model (L-GLM) approach was adopted in this study for CVR estimation (Moia, Stickland, et al., 2020) in order to model temporal offsets between the $P_{ET}CO_2$ recording and the CVR response across voxels that occur due to measurement and physiological delays (Donahue et al., 2016; Geranmayeh, Wise, Leech, & Murphy, 2015; Murphy, Harris, & Wise, 2011; Sousa et al., 2014; Tong, Bergethon, & Frederick, 2011). Sixty shifted versions of the $P_{ET}CO_2$*hrf* trace were created, ranging between ±9 s from the bulk shift, with a shift increment of 0.3 s (fine shift). This temporal range was based on previous literature, which rarely reports haemodynamic lags over ±8 s in healthy individuals (Bright et al., 2009; Donahue et al., 2016; Sousa et al., 2014). For each shift, a GLM was defined with a design matrix comprised of the shifted $P_{ET}CO_2$*hrf* timecourse as the regressor of interest, and different combinations of nuisance regressors (see below) in order to examine their efficiency in modelling artefactual signals of the voxel timeseries that might degrade CVR estimates. The simultaneous fitting of the nuisance regressors and the regressor of interest (i.e. the shifted $P_{ET}CO_2$*hrf* trace) is preferable, rather than denoising via nuisance regression prior to the analysis (Jo et al., 2013; Lindquist, Geuter, Wager, & Caffo, 2019; Moia, Stickland, et al., 2020).

Five different modelling strategies were evaluated, varying which nuisance regressors were included in the design matrix or how they were derived from ME-ICA:

1. A L-GLM model on the SE data where the design matrix includes the motion parameters and their temporal derivatives (denoted as *Mot*), Legendre polynomials of up to the fourth order (denoted as *Poly*), together with the $P_{ET}CO_2$*hrf* trace (SE-MPR):

$$Y_{SE} = P_{ET} CO_2 hrf + Mot + Poly + n \tag{2}$$

2. The same model applied on the OC data (OC-MPR):

$$Y_{OC} = P_{ET} CO_2 hrf + Mot + Poly + n \tag{3}$$

3. An *aggressive* model applied on the OC data (ME-AGG) in which the design matrix also includes the timecourses of the ME-ICA rejected components (denoted as IC_{rej}) added to the design matrix of the L-GLM, orthogonalised with respect to the motion parameters, their temporal derivatives, and Legendre polynomials of up to the fourth order. This orthogonalisation step was performed to maintain a low Variance Inflation Factor in this model, and thus not bias the CVR estimation, without altering the relative variance explained by the original nuisance regressors and the regressor of interest (Mumford, Poline, & Poldrack, 2015):

$$Y_{OC} = P_{ET} CO_2 hrf + Mot + Poly + [IC_{rej} \perp (Mot, Poly)] + n \tag{4}$$

4. A *moderate* model applied on the OC data (ME-MOD) in which the timecourses of the ME-ICA rejected components are also orthogonalised with respect to the $P_{ET}CO_2$*hrf* trace (i.e. the regressor of interest describing the CVR response):

$$Y_{OC} = P_{ET} CO_2 hrf + Mot + Poly + [IC_{rej} \perp (P_{ET} CO_2 hrf, Mot, Poly)] + n \tag{5}$$

5. A *conservative* model applied on the OC data (ME-CON) in which the timeseries of the ME-ICA rejected components are orthogonalised with respect to the $P_{ET}CO_2$*hrf* trace and the ME-ICA accepted components (denoted as IC_{acc}):

$$Y_{OC} = P_{ET} CO_2 hrf + Mot + Poly + [IC_{rej} \perp (P_{ET} CO_2 hrf, IC_{acc}, Mot, Poly)] + n \tag{6}$$

In the models above, Y_{SE} and Y_{OC} are the SE and OC voxel timeseries respectively and n denotes the random noise.

For each modelling strategy and each of the sixty shifted $P_{ET}CO_2$*hrf* traces, the corresponding L-GLM was fitted via orthogonal least squares using AFNI. Then, for each voxel, the beta coefficient (i.e. weight) of the best fine-shifted $P_{ET}CO_2$*hrf* trace, corresponding to the L-GLM model with maximum coefficient of determination (R^2), was selected. Finally, the beta coefficients expressed in BOLD signal percentage change over Volts ($BOLD_{SPC}$/V) were rescaled to be expressed in BOLD

percentage over millimetres of mercury (%BOLD/mmHg) as indicated by the gas analyser manufacturer[1].

In this way, a lag-optimised CVR map and a t-value map were obtained, together with the associated lag map representing the voxelwise delay from the bulk shift, for each analysis pipeline. To account for sixty comparisons computed in the L-GLM approach (one per lagged regressor), the CVR and lag maps were thresholded at p<0.05 adjusted with the Šidák correction (Bright, Tench, & Murphy, 2017; Šidák, 1967), and the voxels that were not statistically significant were excluded. The maps were further thresholded on the basis of the lag: those voxels in which the optimal lag was at or adjacent to the boundary (i.e. $|lag| \geq 8.7\,s$) were considered not truly optimised and not readily physiologically plausible in healthy subjects and therefore masked in all maps (Moia, Stickland, et al., 2020).

2.6 Evaluation of motion removal

For each type of L-GLM analysis, 4-D volumes representing the modelled noise variance were reconstructed by multiplying the optimised beta coefficient maps of the regressors of non-interest by their timeseries using 3dSynthesize in AFNI. Then, they were subtracted from the OC or the SE data to obtain five different denoised datasets. DVARS, the root of the spatial mean square of the first derivative of the signal (Smyser et al., 2010), was computed on each denoised dataset as:

$$DVARS_t = \sqrt{\langle [I_t(x) - I_{t-1}(x)]^2 \rangle}\,, \tag{7}$$

where $I_t(x)$ is the image intensity of voxel x and at time t and $\langle \cdots \rangle$ indicates the spatial average over the whole brain. These DVARS timeseries were compared with the Framewise Displacement (FD) time courses (Power, Barnes, Snyder, Schlaggar, & Petersen, 2012), computed using the realignment parameters estimated during preprocessing using the fsl_motion_outliers tool as:

$$FD_t = |\Delta d_x| + |\Delta d_y| + |\Delta d_z| + |\Delta \alpha| + |\Delta \beta| + |\Delta \gamma|\,, \tag{8}$$

where t denotes the time, d_x , d_y , d_z are the translational displacements along the three axes, α , β , γ are the rotational displacements of pitch, yaw, and roll, and $\Delta d_x = d_{x,t-1} - d_{x,t}$ (and similarly for the other parameters). DVARS was also computed on the SE volume before preprocessing (SE-PRE) to serve as a reference, as its relationship with FD should be at its maximum prior to the effects of motion being removed.

1 https://www.adinstruments.com/support/knowledge-base/it-possible-measure-expired-gasses-partial-pressure-mmhg-rather-percentage. We used the formula $CO_2[mmHg] = (P_{atm} - P_{vap})[mmHg] \cdot 10 \cdot CO_2[V]/100[V]$, where $CO_2[V]$ is the original CO_2 timeseries, P_{atm} is the atmospheric pressure in the laboratory at the moment of acquisition, and P_{vap} is the water vapour pressure associated with expired air. The values of $P_{atm} = 759$ and $P_{vap} = 47$ were used for all the sessions.

In order to visualise the CVR responses to a BH trial, the average timeseries within GM was extracted from each denoised dataset from each model SE-PRE, SE-MPR, OC-MPR, ME-AGG, ME-MOD, ME-CON, as well as from SE-PRE. These timeseries were transformed to BOLD percentage signal change, then the response to individual BH trials from each session were extracted using the timing of the third paced breathing cycle as a reference onset, and averaged together for each subject. The DVARS and FD timeseries followed the same process, except that the FD timeseries were not expressed in percentage.

Finally, we compared how many BH trials were necessary in each pipeline to achieve a robust estimation of the BH response. The Manhattan distance from a pool of a gradually increasing number of trials to the average BOLD response over all BH trials (across the ten sessions, 80 trials in total) was also computed for each analysis model and subject.

2.7 Reliability and variability analysis

In order to assess the reliability of each analysis model, the thresholded CVR, t-value and lag maps of each session were normalised with a nearest neighbour interpolation to the MNI152 template (Grabner et al., 2006). Then, the intraclass correlation coefficient (ICC) was computed voxelwise on the CVR and lag maps using a regularized multilevel mixed effect model in 3dICC, in order to take into account the standard error of CVR and lag for each session in the ICC estimation (Chen et al., 2018). ICC assesses the reliability of a metric by comparing the intersubject, intrasubject, and total variability of that metric:

$$ICC(2,1)=\rho_2=\frac{\sigma_{subj}^2}{\sigma_{subj}^2+\sigma_{sess}^2+\sigma_n^2} \tag{9}$$

which is equivalent to:

$$\hat{\rho}_2=\frac{MS_{subj}-MS_n}{\frac{k}{n}(MS_{sess}-MS_n)+MS_{subj}+(k-1)MS_n} \tag{10}$$

where MS_{subj}, MS_{sess}, and MS_n are the mean squares of the effects of subjects, sessions, and residuals respectively, k is the number of sessions, and n the number of subjects (Chen et al., 2018; Mcgraw & Wong, 1996; Shrout & Fleiss, 1979). ICC(2,1) was chosen since both subjects and sessions were considered random effects. High ICC scores indicate high reliability, where the intrasubject variability is lower than the intersubject variability. Note that, since 3dICC uses the t-statistic map associated with the estimation of the CVR, CVR and lag maps used in this computation were thresholded only on the basis of the lag and not on the basis of the t-statistic.

2.8 Methods and data availability

In order to guarantee the replication of methods and results, all the code has been prepared to be run in a Singularity container based on a Ubuntu 18.04 Neurodebian OS. The container is publicly available at https://git.bcbl.eu/smoia/euskalibur_container, the methods pipeline is available at https://github.com/smoia/EuskalIBUR_dataproc, while all of the MRI images, physiological data, and manual classification used in this study are available in OpenNeuro (EuskalIBUR dataset).

3 Results

Three subjects were excluded due to poor performance of the BH task, mainly due to inadequate execution of the exhalations preceding and following the apnoea which prevented accurate determination of the $P_{ET}CO_2$*hrf* traces (see Supplementary Materials). Hence, seven subjects were used for subsequent analyses (4F, age 25-40y).

3.1 Motion removal across pipelines

Figure 2a illustrates the relationship between FD and DVARS in the raw data (SE-PRE) and after removing the reconstructed noise of each analysis model from the SE or OC volume for a representative subject; each point represents a timepoint and each line represents the linear regression between both timeseries in one session. The corresponding figures for the remaining subjects are available as Supplementary Material (supplementary figure 2). Figure 2b shows the same plot considering all the subjects and sessions. The optimal combination (OC-MPR) of ME data reduces DVARS compared to single-echo (SE-MPR), although a similar relationship is observed between DVARS and FD in both approaches. This relationship is mitigated in the moderate (ME-MOD) and conservative (ME-CON) denoising approaches, which show similar modulatory effects on it. Note that this similarity is common, but not the same for all the subjects; for instance, ME-MOD showed higher impact than ME-CON for two subjects (subject 003 and 007), while the opposite pattern was observed in two other subjects (subject 004 and 009) (see Supplementary figure 2). The aggressive strategy (ME-AGG) is the most successful in reducing motion-related effects described by FD on DVARS.

Figure 3a plots the average percentage DVARS (left column) and average GM percentage BOLD response (central column) of all the BH trials across all of the sessions of a representative subject. The FD trace features a clear peak right after the end of the apnoea (highlighted in grey), likely associated with large head movement arefacts caused by the recovery breaths following the apnoea period. The percentage DVARS curves of the SE-PRE, SE-MPR and OC-MPR denoised timeseries reflect this peak in FD, which is absent in the ME-ICA based denoising timeseries, indicating a strong influence of movement on the signal intensity changes. All DVARS curves present a peak at a later time (between timepoints 25 and 30) that, as DVARS is akin to the first derivative of the

BOLD signal changes, may agree with the return to the baseline seen in the BOLD response. The percentage BOLD signal change curves feature a delayed peak compared to the FD trace, reflecting a delayed CVR response compared to instantaneous head movements associated with respiration. However, they also feature a modulation in the BOLD signal change in correspondence with the peak in the FD trace, with the exception of ME-MOD and ME-AGG. The flattened DVARS and BOLD responses seen for ME-AGG indicate that the inclusion of the ME-ICA rejected components substantially removes part of the true CVR response, compared with the OC-MPR time courses. The average percentage DVARS and percentage BOLD response of the other subjects can be found in the Supplementary Materials (Supplementary figure 3).

Figure 3b plots the Manhattan distance between the average of *N* trials and the average of all 80 BH trials as *N* increases from 1 to 80. ME-AGG tends to be more similar to the total average compared to all the other timeseries. For most of the subjects, SE-MPR, OC-MPR and ME-MOD have a similar behaviour and need more trials than SE-PRE, ME-CON and ME-AGG to converge to the total average. Note that the convergence to the analysis-specific ‘ground truth’ BH response is not monotonic and fluctuates across trials of the same session and across sessions, indicating that the convergence does not depend only on the number of BH trials, but also on their quality and possible physiological variability in the CVR response across trials and sessions.

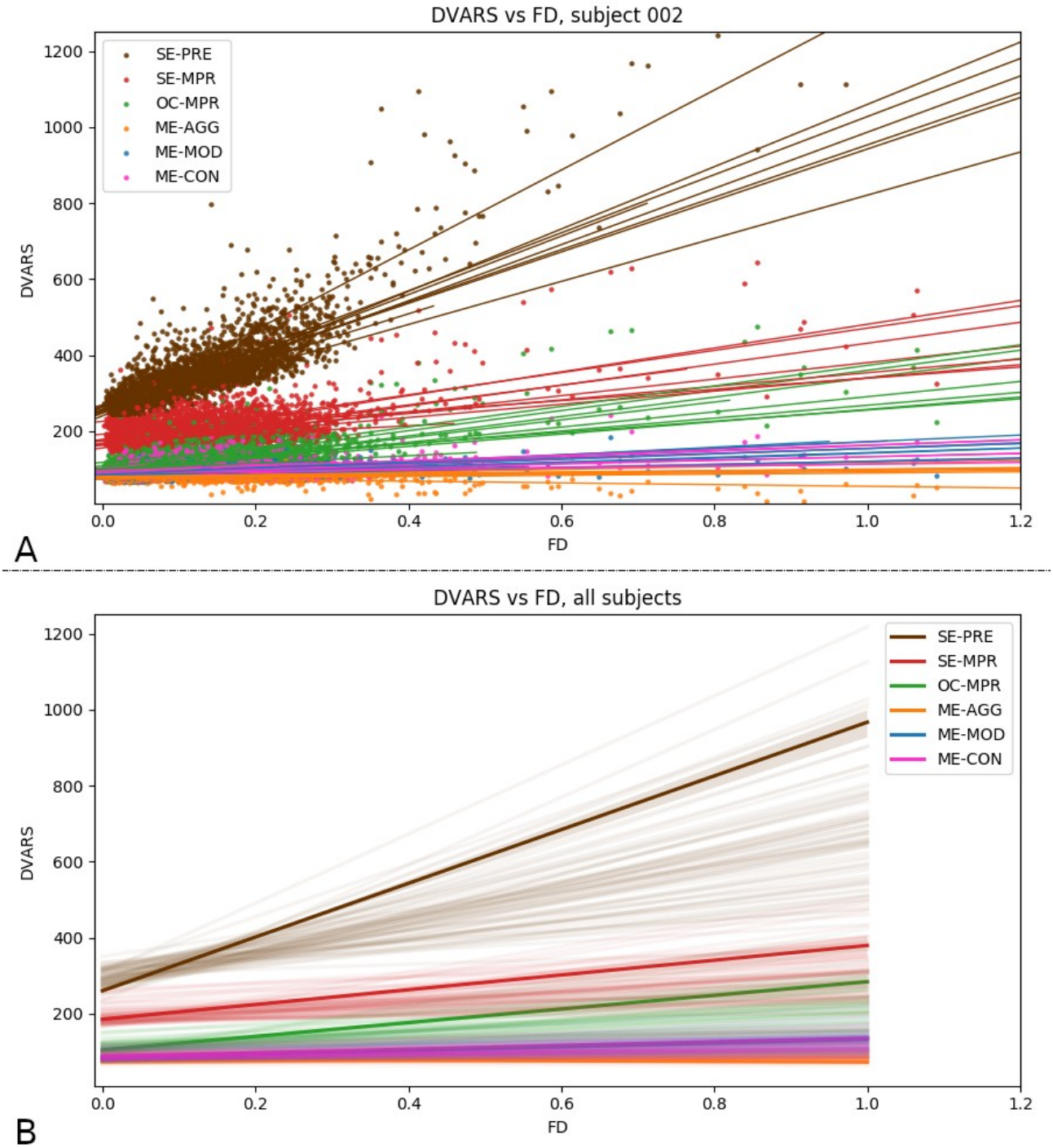

Figure 2: (A) Relation between the DVARS of the denoised data following different analysis pipelines and FD for a representative subject. Each point represents a timepoint, each line the linear regression between both timeseries in a session. In general, OC-MPR shows lower DVARS than SE-MPR, but similar modulation of the DVARS-FD relationship. All the ICA denoising solutions performs better in reducing motion-related effects described by FD on DVARS. Between the ICA solutions, ME-AGG performs the best in reducing this relationship, while ME-MOD and ME-CON seem to be equivalent. (B) DVARS vs. FD for all the subjects. Each transparent line represents a session, the solid line represents the estimation across subjects and sessions. Similar patterns to the representative subject are shown. SE-PRE: raw data; SE-MPR: single-echo; OC-MPR: optimally combined; ME-AGG: aggressive; ME-MOD: moderate; ME-CON: conservative. The relation between DVARS and FD of the other subjects can be found in the Supplementary Material (Supplementary figure 2).

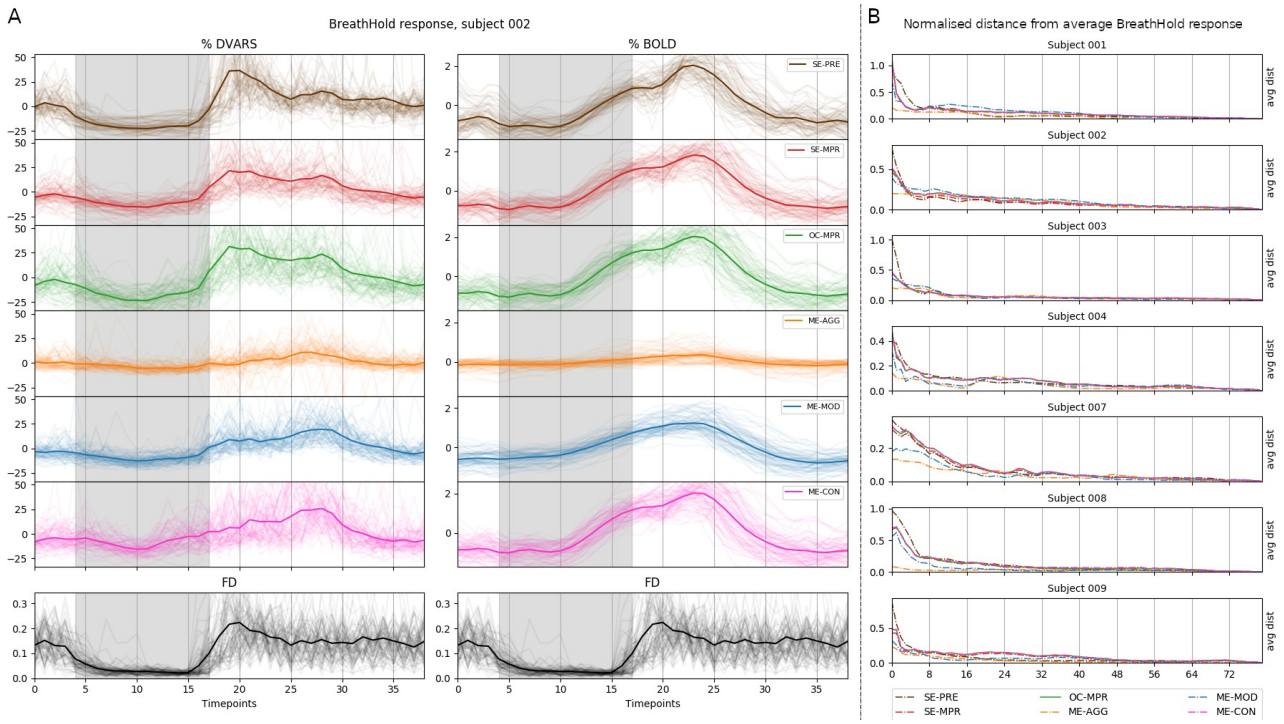

Figure 3: (A) Average GM %DVARS and %BOLD response of all BH trials across ten sessions for the same representative subject. The apnoea period is highlighted in grey. Each transparent line is a trial, the solid line is the average across all the trials. (B) Manhattan distance between the average of N trials and the average of all 80 BH trials as N increases from 1 to 80 for each subject. Each vertical line divides the number of trials in each session. SE-PRE: raw data; SE-MPR: single-echo; OC-MPR: optimally combined; ME-AGG: aggressive; ME-MOD: moderate; ME-CON: conservative. The average %DVARS and %BOLD response of the other subjects can be found in the Supplementary Material (Supplementary figure 2)

3.2 CVR maps and reliability of the pipelines

Figure 4 and 5 show CVR and lag maps respectively, for all analysis strategies and all sessions of a representative subject (subject 002). The CVR maps were masked to exclude the voxels that were not statistically significant or whose lag is at the boundary of the explored range and might not be truly optimised or physiologically plausible. SE-MPR features more spatial variation and speckled noise in CVR and lag estimates of voxels within the same brain region compared to ME approaches like OC-MPR or ME-CON. In general, the ME-AGG and ME-MOD approaches do not yield CVR maps with as much clear distinction between brain tissues or delineation of the cortical folding and subcortical structures (e.g. see putamen and caudate nucleus) as obtained with the OC-MPR and ME-CON models. Among the ICA-based approaches, the adoption of an aggressive (ME-AGG) or moderate (ME-MOD) modelling strategy results in lag maps without anatomically defined patterns, as well as a higher rate of voxels with a lag estimation that is not within physiologically plausible range, and in CVR maps with lower responses and fewer significant voxels. ME-AGG also produces CVR maps with a higher percentage of negative values than any other analysis model, and a reduced CVR response in voxels near the posterior part of the superior sagittal and transverse

sinuses. The CVR and lag maps of other subjects are available in the *S*upplementary Material (Supplementary figure 4 and 5).

In order to assess the reliability of each model, we computed voxelwise ICC(2,1) maps for both CVR and haemodynamic lag. Figure 6 depicts the ICC(2,1) maps for all analysis strategies for both CVR and lag maps, as well as their distributions. High ICC scores indicate that the intra-subject variability is lower than the inter-subject variability, hence the estimations of CVR or haemodynamic lag can be considered consistent across sessions. Conversely, low ICC scores indicate that the inter-subject variability is low compared to the intra-subject variability, hence the estimations of CVR and haemodynamic lag cannot be considered consistent across sessions. Following the classification given by Cicchetti (2001), an ICC score lower than 0.4 is considered poor, lower than 0.6 fair, lower than 0.75 good, and equal or higher than 0.75 excellent.

In terms of whole brain CVR reliability, the ME-CON demonstrated excellent reliability (spatial average across the whole brain of 0.86 ± 0.16) as well as the highest among all methods tested, closely followed by the OC-MPR (excellent, 0.85 ± 0.16), SE-MPR (excellent, 0.81 ± 0.19), and ME-MOD (excellent, 0.79 ± 0.19), while ME-AGG had a fair reliability (0.46 ± 0.22). If only voxels in GM are considered, the ICC of all approaches increases slightly (0.88 ± 0.14, 0.87 ± 0.15, 0.85 ± 0.17, 0.82 ± 0.17, and 0.49 ± 0.22 for ME-CON, OC-MPR, SE-MPR, ME-MOD, and ME-AGG respectively). It can be observed that ME-AGG exhibits a considerable number of voxels with poor reliability (ICC below 0.4). These voxels are mostly located in white matter, and they also exhibit lower ICC values in the other analyses. In terms of whole-brain lag reliability, OC-MPR performed the best (good reliability, 0.67 ± 0.21), closely followed by ME-CON (good reliability, 0.66 ± 0.21). SE-MPR, ME-MOD, and ME-AGG demonstrated fair lag reliability (0.6 ± 0.22 and 0.42 ± 0.19, 0.41 ± 0.20, respectively). Considering only GM voxels, the reliability of all the approaches increases minimally (0.68 ± 0.21, 0.67 ± 0.21, 0.61 ± 0.21, 0.43 ± 0.19, 0.42 ± 0.20, for OC-MPR, ME-CON, SE-MPR, ME-MOD, and ME-AGG respectively). The reliability of CVR lag estimates was lower than that of CVR amplitude estimates, even though certain cortical regions, such as the visual and motor cortices, also show excellent ICC values with OC-MPR and ME-CON. Interestingly, it can be observed that ME-MOD offers excellent ICC values for the CVR response amplitude in grey matter voxels, whereas they are poor for the lag estimates.

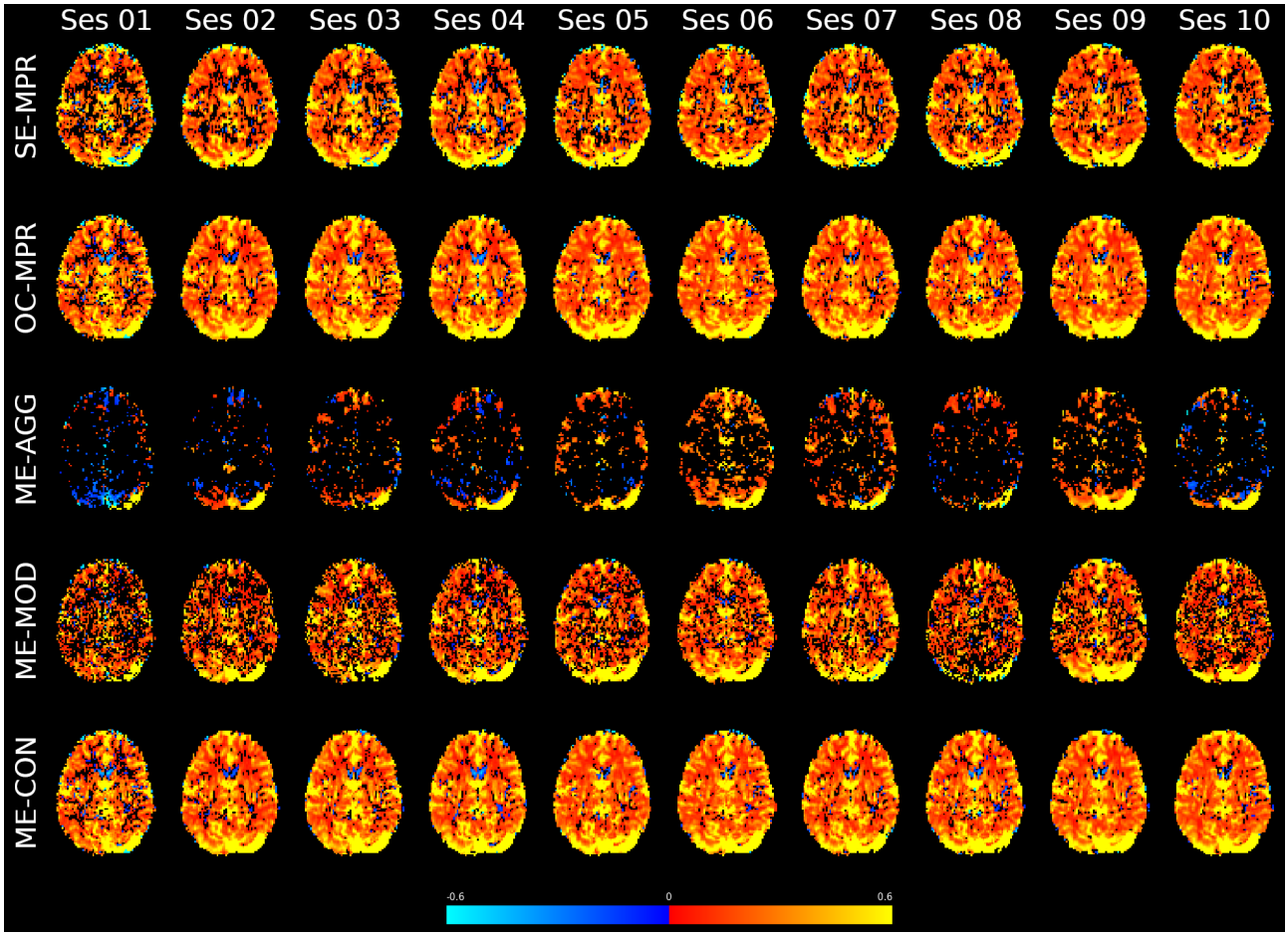

Figure 4: CVR map obtained with the different lagged-GLM analysis for all the sessions of a representative subject (subject 002). Note the low CVR response in ME-AGG, depicting numerous voxels with a negative values, as well as the increased amount of masked voxels in SE-MPR, ME-AGG and ME-MOD. SE-MPR: single-echo; OC-MPR: optimally combined; ME-AGG: aggressive; ME-MOD: moderate; ME-CON: conservative. The CVR maps of other subjects are available in the supplementary materials.

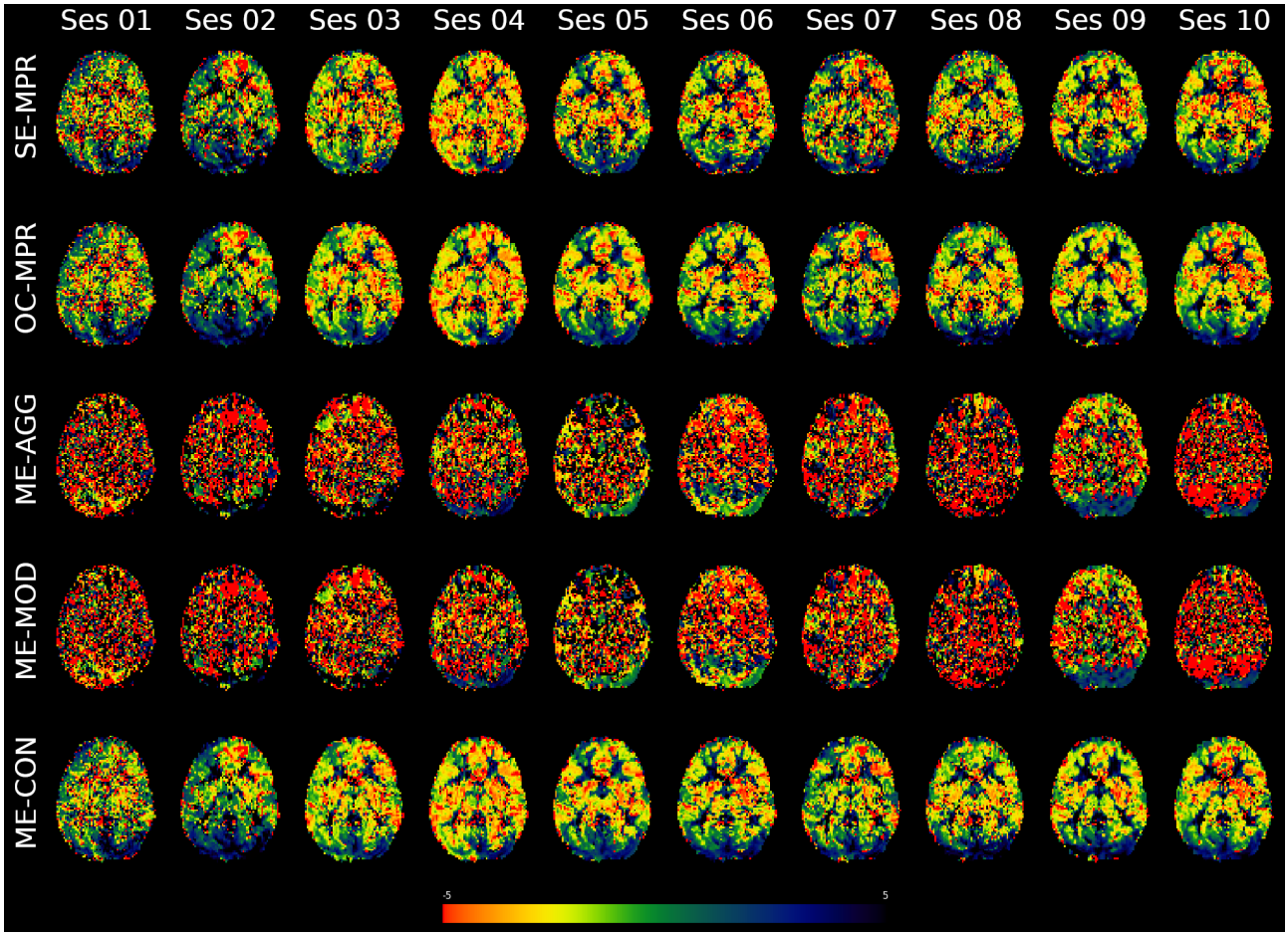

Figure 5: CVR lag map obtained with the different lagged-GLM analysis, for all the sessions of a representative subject (same as figure 4). These lag maps represent the delay between the best shifted version of the $P_{ET}CO_2$*hrf trace and the bulk shift (i.e. the best match between average grey matter signal and* $P_{ET}CO_2$*hrf trace). The scale from -5 to +5 represents earlier to later hemodynamic responses. Note the lack of anatomically informative patterns in ME-MOD and ME-AGG. SE-MPR: single-echo; OC-MPR: optimally combined; ME-AGG: aggressive; ME-MOD: moderate; ME-CON: conservative. The lag maps of other subjects are available in the supplementary materials.*

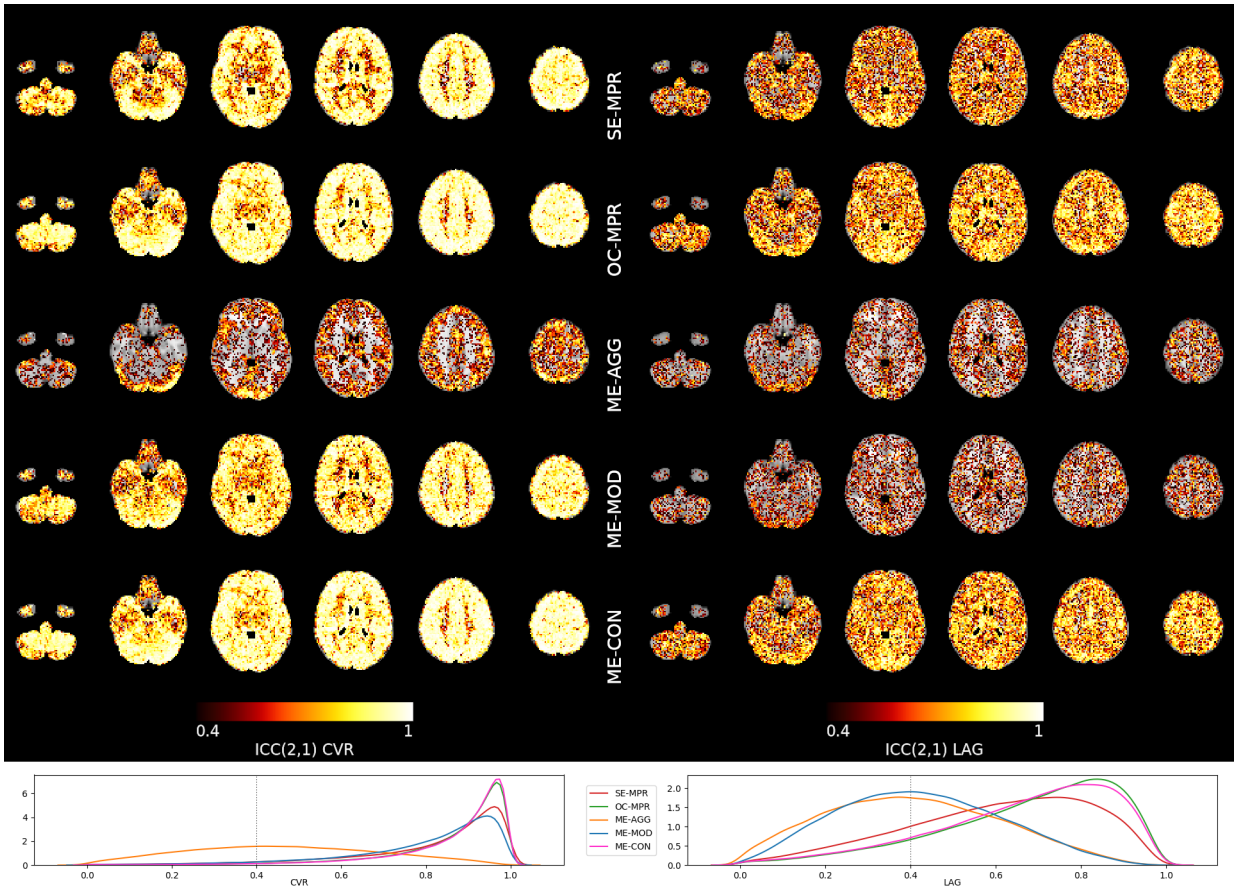

Figure 6: ICC(2,1) maps of CVR (left) and haemodynamic lag (right) for each analysis pipeline. The maps are thresholded at 0.4 since scores lower than it indicate poor reliability . A high ICC score indicates that the inter-subject variability is higher than the intra-session variability, while a low ICC score suggest that the variability across sessions is the same as the one across subjects. The bottom rows depict the whole brain distribution of ICC scores across voxels. Note how OC-MPR and ME-CON have generally higher ICC scores than the other approaches, and are very similar to each other, while ME-AGG has the lowest ICC scores for both CVR and lag maps. SE-MPR: single-echo; OC-MPR: optimally combined; ME-AGG: aggressive; ME-MOD: moderate; ME-CON: conservative. The distribution of ICC scores across GM voxels only is available in the Supplementary Material (Supplementary figure 6).

4 Discussion

In this study, we compared five different analysis strategies based on a lagged-GLM model (Moia, Stickland, et al., 2020) to simultaneously remove motion-related effects and non-BOLD artefacts in the BOLD fMRI signal while estimating CVR and haemodynamic lag in order to identify the best modelling approach for BH paradigms in which prominent task-correlated artefacts coexist with the effect of interest.

Among all possible modelling strategies, the five presented here were included in our analysis for different reasons. The optimal combination of ME fMRI data, with subsequent motion and Legendre polynomial regression (MPR), was expected to remove more noise and improve reliability of the CVR estimation compared to MPR on single-echo data, which is the standard approach for BH CVR estimation (Cohen & Wang, 2019). ICA-based approaches are known to

outperform traditional MPR in typical denoising fMRI data, possibly because they can identify and separate artefactual sources in the data in a data-driven and non-linear manner (Griffanti et al., 2014; R. H. R. R. Pruim et al., 2015; Salimi-Khorshidi et al., 2014). We did not apply ICA to single-echo data because it has already been demonstrated that ICA-based denoising applied to OC data outperforms ICA denoising applied to single-echo data (Dipasquale et al., 2017) and the ICs estimated from OC data might not have matched ICs obtained from single-echo data, making such comparison less straight forward than the one based on MPR.

Spatial ICA decomposition is applied to fMRI data more often than temporal ICA decomposition, as the latter requires many more samples in time than normally available. Having many sessions for each subject, temporal ICA could have been leveraged in this study. In fact, temporal ICA could be more appropriate than spatial ICA to estimate a proper decomposition of timeseries sources (Smith et al., 2012), improving the modelling of temporal noise (Glasser et al., 2018), and potentially leading to better disentanglement of noise from CVR effects. However, we decided to apply spatial ICA in order to maintain the independence of each session, both to simulate a more common denoising approach to fMRI data, and to be able to capture session-specific noise contributions that could have been missed otherwise. Further studies could compare temporal and spatial ICA denoising for CVR mapping when many temporal samples have been collected in the same session, for instance reducing the TR by acquiring fewer echoes. Here, our decision to acquire five echoes, instead of conventional multi-echo protocols with three or four echoes, was made to facilitate and improve the classification of the ICs based on their TE-dependence.

The choice of comparing different levels of orthogonalisation of only the ICA-based nuisance regressors compared to regressors of interest might seem in contrast with previous literature, that suggests that orthogonalisation of collinear confounding factors could lead to misinterpreted results (Mumford et al., 2015). Our results clearly demonstrated that using the original (e.g., non orthogonalised) rejected ICs as nuisance regressors in the analysis (ME-AGG) removes the CVR effect of interest (see Figures 3, 4, 5 and 6). To decide which regressors should be orthogonalised, and with respect to what, we considered the different origin of the nuisance regressors. While Legendre polynomials and motion parameters can be considered adequate models of noise sources in the data, intrinsic data-driven regressors may well contain variance related to the effect of interest, especially as spatial ICA was adopted and because of the high collinearity between the $P_{ET}CO_2$*hrf*, motion, physiological adaptations to vascular dilation (e.g. cerebrospinal fluid flows), or changes in the magnetisation related to breathing (Raj et al., 2001). In these scenarios, it becomes more important to understand how to properly implement ICA denoising in order to preserve the effect of interest. For these reasons, three different ICA-based approaches were selected, from an

aggressive strategy to a conservative approach, to assess if they preserved the BOLD effects related to the CVR response happening at different lags.

As hypothesized, all of the ME-based solutions outperformed the SE-MPR model in their ability to account for the effect of motion, summarized in terms of FD, on the fMRI signal intensity changes, described in terms of DVARS (see Figure 2). Furthermore, all of the ICA-based strategies outperformed traditional MPR, and within ICA-based strategies, the aggressive one (ME-AGG) showed the best performance. However, observing the average DVARS and BOLD responses (Figure 3) and the CVR and lag maps (Figures 4 and 5) it becomes evident how aggressive and moderate approaches result in lower estimates of CVR responses, even compared to the SE-MPR approach. Similarly, these two approaches result in the estimated haemodynamic lag hitting the boundaries of a physiologically plausible lag range in healthy adults. The substantial reduction in the CVR estimates in the aggressive approach (Figures 3 and 4) occurs because the effect of interest can also be explained as a linear combination of the timecourses of rejected ICs related to motion, vascular effects or large susceptibility changes due to chest expansions and contractions while performing the BH task (Caballero-Gaudes & Reynolds, 2017; Griffanti et al., 2017). As for the moderate approach, the lower estimates of CVR could be due to the fact that orthogonalising data-driven nuisance regressors with respect to the $P_{ET}CO_2$*hrf* trace per sé is not sufficient to save all the variance associated to real CVR. The $P_{ET}CO_2$ trace can only be estimated during exhalations, hence it is unable to capture local dynamic signal changes that are captured by ICs timeseries. Furthermore, CVR has a sigmoidal non-linear relation with the $P_{ET}CO_2$*hrf* trace (Bhogal et al., 2014), and the local BH-induced BOLD response has a complex shape, in terms of response amplitude and temporal delays, due to multiple physiological factors (Magon et al., 2009) that must be accounted for in order to improve its estimation. Our results demonstrate that these local complexities might be adequately captured by the linear combination of the accepted ICs timecourses, and not removing this variance from the rejected ICs when they are included as nuisance regressors in the model is detrimental (as observed with ME-MOD and ME-AGG approaches). In other words, only a conservative approach (ME-CON) that preserves the BOLD variance associated with local CVR responses performs well, while also reducing motion-related effects more than with the conventional MPR models.

To further explore the benefit of different modelling strategies, we assessed the reliability of the resulting CVR and haemodynamic lag maps over the course of two and a half months (ten sessions) using ICC(2,1). To our knowledge, this was the first time that CVR reliability was tested over the course of ten sessions in individual subjects, and the first time that intersession haemodynamic lag reliability was tested. The ME-CON and OC-MPR strategies featured the greatest reliability for

CVR and lag estimation, while the ME-AGG and ME-MOD approaches produced lower reliability values than even the simple SE-MPR model.

The lag maps are computed as the temporal offset related to the bulk shift, which is obtained by aligning the average GM BOLD response with the $P_{ET}CO_2$*hrf* trace. If the bulk shift computation is misestimated this would create a systematic bias in the estimated lag maps, potentially reducing the apparent intersession reliability. While the CVR reliability should not be affected by this issue, due to the use of a lagged-GLM approach that can overcome bulk shift misestimation (see session 4 of subject 007 in Supplementary figure 4 and 5), the true lag map reliability might be higher than reported here.

Regarding CVR reliability, the whole-brain average reliability of SE-MPR was comparable to long-term reliability (days or weeks apart) found in previous studies of CVR induced by BH (Peng et al., 2019), by paced deep breathing (Sousa et al., 2014), or by gas challenges (Leung et al., 2016), and higher than that reported in other studies on BH induced CVR estimated with a non-lagged optimized $P_{ET}CO_2$*hrf* trace (Lipp et al., 2015) or with Fourier modelling (Pinto et al., 2016), and by gas challenges (Dengel et al., 2017; Evanoff et al., 2020). Consequently, the reliability of CVR estimates obtained with the optimal combination dataset and conservative ME-ICA modelling approaches were found higher than those previously reported in the literature. However, all strategies produced a reliability that was lower than the short-term (within-session) reliability reported in BH induced CVR (Peng et al., 2019), resting state based CVR (P. Liu et al., 2017), and gas challenge induced CVR (Leung et al., 2016), although lower intersession reliability in gas challenges has also been reported (Dengel et al., 2017; Evanoff et al., 2020). Note that the reliability observed in this study seems to be globally higher and spatially less variable than that reported in previous studies (Lipp et al., 2015; Sousa et al., 2014). However, discrepancies in the reliability measurements might be related to the different methods used to compute the CVR maps and the ICC score itself.

Using ICC to test reliability has the drawback that higher scores might be related to the presence of residual task-correlated motion effects that artificially stabilise the CVR estimation and reduce intrasubject variability compared to intersubject variability. In fact, recent studies have shown that individuals have particular movement patterns during fMRI sessions that may be a stable characteristic of a person (Bolton et al., 2020) related to stable physical characteristics, such as body mass index (Ekhtiari, Kuplicki, Yeh, & Paulus, 2019) and could even be a heritable characteristic (Couvy-Duchesne et al., 2014; Hodgson et al., 2017). If subjects have similar motion patterns across the 10 repeated sessions, fMRI responses might appear more similar than they truly are, and the ICC might be inflated by such effects. Moreover, higher spatial reliability does not necessarily mean higher accuracy: a denoising strategy might be systematically misestimating CVR or

haemodynamic lag. The fact that both optimal combination with traditional nuisance regressio and the conservative ME-ICA denoising approaches resulted in similar CVR and lag spatial patterns and exhibited higher reliability than the single-echo model, while at the same time reduced the apparent effect of motion on the data variance, suggests that such drawbacks are mitigated in our data. However, further studies could compare different BH analysis strategies with a CVR estimation based on an independent computerised gas delivery protocol.

Another possibility would be to assess CVR in resting state fMRI, either measuring resting fluctuations in exhaled CO_2 levels (Golestani, Wei, & Chen, 2016; Lipp et al., 2015), or by using a band of the power spectrum of the global signal as a regressor of interest (P. Liu et al., 2017, 2020). Such method might be more robust to motion collinearity, as the amount of movement in each breath is less pronounced and not consistently time-locked to the paradigm cues. At the same time, the lower amplitude of intrinsic CO_2 fluctuations relative to BH CO_2 change might also make this approach more susceptible to general motion effects and other sources of variance (e.g. neural or artefactual) unrelated to CO_2. Moreover, previous work has shown that the optimal temporal shift between BOLD and $P_{ET}CO_2$ is hard to reliably identify in resting state data alone, in contrast to BH datasets where the temporal shift can be reliably identified (Bright et al., 2017; Stickland, Ayyagari, Zvolanek, & Bright, 2020). Resting state fMRI methods may therefore be inappropriate to use with the lagged-GLM approach that we have applied here. Either way, the analyses presented in this study can be easily implemented in other CVR assessment pipelines to mitigate the dependence of the response on motion. Beyond BH-based CVR studies, similar conclusions might be applicable to other experimental paradigms that present high collinearity between the expected task induced activity and artefactual sources, such as in overt speech production with long trial durations (Birn et al., 1999, 2004; Gracco, Tremblay, & Pike, 2005), that aim to use (ME-) ICA-based nuisance regressors as part of the model.

Note that MPR and ICA denoising are not the only viable options to reduce motion effects on fMRI and BH-induced CVR in particular: advanced setups can be used to reduce motion during the acquisition itself. For instance, subject specific moulded head casts can be used to reduce head motion (Power, Silver, et al., 2019). Mounting an MRI compatible camera or tracker in the scanner enables prospective motion correction techniques (Faraji-Dana, Tam, Chen, & Graham, 2016; Maziero, Rondinoni, Marins, Stenger, & Ernst, 2020; Parkes, Fulcher, Yücel, & Fornito, 2018; Schulz et al., 2014; Zaitsev, Akin, LeVan, & Knowles, 2017) or concurrent field monitoring enables the dynamic correction of field distortions dynamically (Vannesjo et al., 2015; Wilm et al., 2015) in order to effectively reduce effects of motion and magnetic field susceptibility changes. However, such advanced setups are not always available.

A limitation of this study is that the results are influenced by the manual classification of the ICA components performed by two of the authors. Despite being based on the automatic classification made by *tedana*, we adopted a manual approach because often multiple ICs clearly exhibiting CVR-related timeseries and spatial maps were misclassified as noise. This manual classification was made with a cautious approach: if an IC seemed to be temporally and spatially related to the CVR response, it was accepted. Manual classification is still considered the gold standard for the classification of ICA components when performed by experts, despite the introduction of automatic classification algorithms (Griffanti et al., 2017), calling for further improvements in the automatic classification of (ME-)ICA components for BH tasks.

5 Conclusion

Breath Holding (BH) is a non-invasive, robust way to estimate cerebrovascular reactivity (CVR). However, due to the task-correlated movement introduced by the BH task, attention has to be paid when choosing an appropriate modelling strategy to remove movement-related effects while preserving the effect of interest ($P_{ET}CO_2$). We compared different multi-echo (ME) independent component analysis (ICA) based denoising strategies to the standard data acquisition and analysis procedure, i.e. single-echo motion parameters regression. We found that a conservative ICA-based approach best removes motion-related effects while obtaining reliable CVR and lag responses, although a simple optimal combination of ME data with motion parameters regression provides similar improvements in reliability compared to single-echo data acquisition.

6 Acknowledgements

CRediT: Stefano Moia: Conceptualisation, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing (OD), Writing (RE), Visualisation, Funding acquisition; Maite Termenon: Methodology, Supervision, Writing (RE); Eneko Uruñuela: Investigation, Writing (RE); Rachael C. Stickland: Methodology, Writing (RE); Molly G. Bright: Methodology, Supervision, Resources, Writing (RE); César Caballero-Gaudes: Conceptualisation, Methodology, Investigation, Resources, Writing (RE), Supervision, Project Administration, Funding acquisition.

The authors would like to thank Gang Chen for his help in the application of 3dICC, and Vicente Ferrer for collaborating in data acquisition.

This research was supported by the European Union's Horizon 2020 research and innovation program (Marie Skłodowska-Curie grant agreement No. 713673), a fellowship from La Caixa Foundation (ID 100010434, fellowship code LCF/BQ/IN17/11620063), the Spanish Ministry of Economy and Competitiveness (Ramon y Cajal Fellowship, RYC-2017- 21845), the Spanish State Research Agency (BCBL "Severo Ochoa" excellence accreditation, SEV- 2015-490), the Basque

Government (BERC 2018-2021 and PIBA_2019_104), the Spanish Ministry of Science, Innovation and Universities (MICINN; PID2019-105520GB-100 and FJCI-2017-31814), and the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award number K12HD073945.

7 Bibliography

Amemiya, S., Yamashita, H., Takao, H., & Abe, O. (2019). Integrated multi-echo denoising strategy improves identification of inherent language laterality. *Magnetic Resonance in Medicine*, *81*(5), 3262–3271. https://doi.org/10.1002/mrm.27620

Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, *20*(2), 870–888. https://doi.org/10.1016/S1053-8119(03)00336-7

Atwi, S., Metcalfe, A. W. S., Robertson, A. D., Rezmovitz, J., Anderson, N. D., & MacIntosh, B. J. (2018). Attention-related brain activation is altered in older adults with white matter hyperintensities using multi-echo fMRI. *Frontiers in Neuroscience*, *12*(OCT), 1–14. https://doi.org/10.3389/fnins.2018.00748

Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A., & Gee, J. C. (2011). An open source multivariate framework for N-tissue segmentation with evaluation on public data. *Neuroinformatics*, *9*(4), 381–400. https://doi.org/10.1007/s12021-011-9109-y

Barch, D. M. D. M., Sabb, F. W. F. W., Carter, C. S. C. S., Braver, T. S. T. S., Noll, D. C. D. C., & Cohen, J. D. J. D. (1999). Overt verbal responding during fMRI scanning: empirical investigations of problems and potential solutions. *NeuroImage*, *10*(6), 642–657.

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), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Bhogal, A. A., Siero, J. C. W., Fisher, J. A., Froeling, M., Luijten, P., Philippens, M., & Hoogduin, H. (2014). Investigating the non-linearity of the BOLD cerebrovascular reactivity response to targeted hypo/hypercapnia at 7T. *NeuroImage*, *98*, 296–305. https://doi.org/10.1016/j.neuroimage.2014.05.006

Bianciardi, M., Fukunaga, M., van Gelderen, P., Horovitz, S. G., de Zwart, J. A., Shmueli, K., & Duyn, J. H. (2009). Sources of functional magnetic resonance imaging signal fluctuations in the human brain at rest: a 7 T study. *Magnetic Resonance Imaging*, *27*(8), 1019–1029. https://doi.org/10.1016/j.mri.2009.02.004

Birn, R. M., Bandettini, P. A., Cox, R. W., & Shaker, R. (1999). Event-related fMRI of tasks involving brief motion. *Human Brain Mapping*, *7*(2), 106–114. https://doi.org/10.1002/(SICI)1097-0193(1999)7:2<106::AID-HBM4>3.0.CO;2-O

Birn, R. M., Cox, R. W., & Bandettini, P. A. (2004). Experimental designs and processing strategies for fMRI studies involving overt verbal responses. *NeuroImage*, *23*(3), 1046–1058. https://doi.org/10.1016/j.neuroimage.2004.07.039

Bolton, T. A. W., Kebets, V., Glerean, E., Zöller, D., Li, J., Yeo, B. T. T., … Van De Ville, D. (2020). Agito ergo sum: Correlates of spatio-temporal motion characteristics during fMRI. *NeuroImage*, *209*(June 2019). https://doi.org/10.1016/j.neuroimage.2019.116433

Bright, M. G., Bulte, D. P., Jezzard, P., & Duyn, J. H. (2009). Characterization of regional heterogeneity in cerebrovascular reactivity dynamics using novel hypocapnia task and BOLD fMRI. *NeuroImage*, *48*(1), 166–175. https://doi.org/10.1016/j.neuroimage.2009.05.026

Bright, M. G., & Murphy, K. (2013a). Reliable quantification of BOLD fMRI cerebrovascular reactivity despite poor breath-hold performance. *NeuroImage*, *83*, 559–568. https://doi.org/10.1016/j.neuroimage.2013.07.007

Bright, M. G., & Murphy, K. (2013b). Removing motion and physiological artifacts from intrinsic BOLD fluctuations using short echo data. *NeuroImage*, *64*(1), 526–537. https://doi.org/10.1016/j.neuroimage.2012.09.043

Bright, M. G., & Murphy, K. (2015). Is fMRI “noise” really noise? Resting state nuisance regressors remove variance with network structure. *NeuroImage*, *114*, 158–169. https://doi.org/10.1016/j.neuroimage.2015.03.070

Bright, M. G., Tench, C. R., & Murphy, K. (2017). Potential pitfalls when denoising resting state fMRI data using nuisance regression. *NeuroImage*, *154*(December 2016), 159–168. https://doi.org/10.1016/j.neuroimage.2016.12.027

Buterbaugh, J., Wynstra, C., Provencio, N., Combs, D., Gilbert, M., & Parthasarathy, S. (2015). Cerebrovascular Reactivity in Young Subjects with Sleep Apnea. *Sleep*, *38*(2), 241–250. https://doi.org/10.5665/sleep.4406

Caballero-gaudes, C., Moia, S., Panwar, P., Bandettini, P. A., & Gonzalez-castillo, J. (2019). A deconvolution algorithm for multi-echo functional MRI: Multi-echo Sparse Paradigm Free Mapping. *NeuroImage*, *202*, 1–26. https://doi.org/10.1016/j.neuroimage.2019.116081

Caballero-Gaudes, C., & Reynolds, R. C. (2017). Methods for cleaning the BOLD fMRI signal. *NeuroImage*, *154*(December 2016), 128–149. https://doi.org/10.1016/j.neuroimage.2016.12.018

Camargo, C. H. F., Martins, E. A., Lange, M. C., Hoffmann, H. A., Luciano, J. J., Young Blood, M. R., … Miyoshi, E. (2015). Abnormal Cerebrovascular Reactivity in Patients with Parkinson's Disease. *Parkinson's Disease*, *2015*. https://doi.org/10.1155/2015/523041

Cauley, S. F., Polimeni, J. R., Bhat, H., Wald, L. L., & Setsompop, K. (2014). Interslice leakage artifact reduction technique for simultaneous multislice acquisitions. *Magnetic Resonance in Medicine*, *72*(1), 93–102. https://doi.org/10.1002/mrm.24898

Chen, G., Taylor, P. A., Haller, S. P., Kircanski, K., Stoddard, J., Pine, D. S., … Cox, R. W. (2018). Intraclass correlation: Improved modeling approaches and applications for neuroimaging. *Human Brain Mapping*, *39*(3), 1187–1206. https://doi.org/10.1002/hbm.23909

Churchill, N. W., Hutchison, M. G., Graham, S. J., & Schweizer, T. A. (2020). Cerebrovascular reactivity after sport concussion: From acute injury to 1 year after medical clearance. *Frontiers in Neurology*, *11*(July), 1–11. https://doi.org/10.3389/fneur.2020.00558

Cicchetti, D. V. (2001). The precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology*, *23*(5), 695–700. https://doi.org/10.1076/jcen.23.5.695.1249

Cohen, A. D., & Wang, Y. (2019). Improving the Assessment of Breath-Holding Induced Cerebral Vascular Reactivity Using a Multiband Multi-echo ASL/BOLD Sequence. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-41199-w

Couvy-Duchesne, B., Blokland, G. A. M., Hickie, I. B., Thompson, P. M., Martin, N. G., de Zubicaray, G. I., … Wright, M. J. (2014). Heritability of head motion during resting state functional MRI in 462 healthy twins. *NeuroImage*, *102*(P2), 424–434. https://doi.org/10.1016/j.neuroimage.2014.08.010

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(29), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Dengel, D. R., Evanoff, N. G., Marlatt, K. L., Geijer, J. R., Mueller, B. A., & Lim, K. O. (2017). Reproducibility of blood oxygen level-dependent signal changes with end-tidal carbon dioxide alterations. *Clinical Physiology and Functional Imaging*, *37*(6), 794–798. https://doi.org/10.1111/cpf.12358

Dipasquale, O., Sethi, A., Marcella Laganà, M., Baglio, F., Baselli, G., Kundu, P., … Cercignani, M. (2017). Comparing resting state fMRI de-noising approaches using multi- and single-echo acquisitions. *Clinical Imaging Sciences Centre, Brighton and Sussex Medical School*, (3). https://doi.org/10.1371/journal.pone.0173289

Donahue, M. J., Strother, M. K., Lindsey, K. P., Hocke, L. M., Tong, Y., & Frederick, B. D. B. (2016). Time delay processing of hypercapnic fMRI allows quantitative parameterization of cerebrovascular reactivity and blood flow delays. *Journal of Cerebral Blood Flow and Metabolism*, *36*(10), 1767–1779. https://doi.org/10.1177/0271678X15608643

DuPre, E., Luh, W. M., & Spreng, R. N. (2016). Multi-echo fMRI replication sample of autobiographical memory, prospection and theory of mind reasoning tasks. *Scientific Data*, *3*(October), 1–9. https://doi.org/10.1038/sdata.2016.116

DuPre, E., Salo, T., Markello, R., Kundu, P., Whitaker, K., & Handwerker, D. (2019, February 6). *ME-ICA/tedana: 0.0.6*. https://doi.org/10.5281/ZENODO.2558498

Ekhtiari, H., Kuplicki, R., Yeh, H. wen, & Paulus, M. P. (2019). Physical characteristics not psychological state or trait characteristics predict motion during resting state fMRI. *Scientific Reports*, *9*(1), 1–8. https://doi.org/10.1038/s41598-018-36699-0

Evanoff, N. G., Mueller, B. A., Marlatt, K. L., Geijer, J. R., Lim, K. O., & Dengel, D. R. (2020). Reproducibility of a ramping protocol to measure cerebral vascular reactivity using functional magnetic resonance imaging. *Clinical Physiology and Functional Imaging*, *40*(3), 183–189. https://doi.org/10.1111/cpf.12621

Evans, J. W., Kundu, P., Horovitz, S. G., & Bandettini, P. A. (2015). Separating slow BOLD from non-BOLD baseline drifts using multi-echo fMRI. *NeuroImage*, *105*, 189–197. https://doi.org/10.1016/j.neuroimage.2014.10.051

Faraji-Dana, Z., Tam, F., Chen, J. J., & Graham, S. J. (2016). A robust method for suppressing motion-induced coil sensitivity variations during prospective

correction of head motion in fMRI. *Magnetic Resonance Imaging*, *34*(8), 1206–1219. https://doi.org/10.1016/j.mri.2016.06.005

Fernandez, B., Leuchs, L., Sämann, P. G., Czisch, M., & Spoormaker, V. I. (2017). Multi-echo EPI of human fear conditioning reveals improved BOLD detection in ventromedial prefrontal cortex. *NeuroImage*, *156*(May), 65–77. https://doi.org/10.1016/j.neuroimage.2017.05.005

Fierstra, J., van Niftrik, C., Piccirelli, M., Bozinov, O., Pangalu, A., Krayenbühl, N., … Regli, L. (2018). Diffuse gliomas exhibit whole brain impaired cerebrovascular reactivity. *Magnetic Resonance Imaging*, *45*(September 2017), 78–83. https://doi.org/10.1016/j.mri.2017.09.017

Friedman, L., Turner, J. A., Stern, H., Mathalon, D. H., Trondsen, L. C., & Potkin, S. G. (2008). Chronic smoking and the BOLD response to a visual activation task and a breath hold task in patients with schizophrenia and healthy controls. *NeuroImage*, *40*(3), 1181–1194. https://doi.org/10.1016/j.neuroimage.2007.12.040

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996). Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine*, *35*(3), 346–355. https://doi.org/10.1002/mrm.1910350312

Geranmayeh, F., Wise, R. J. S., Leech, R., & Murphy, K. (2015). Measuring vascular reactivity with breath-holds after stroke: A method to aid interpretation of group-level BOLD signal changes in longitudinal fMRI studies. *Human Brain Mapping*, *36*(5), 1755–1771. https://doi.org/10.1002/hbm.22735

Glasser, M. F., Coalson, T. S., Bijsterbosch, J. D., Harrison, S. J., Harms, M. P., Anticevic, A., … Smith, S. M. (2018). Using temporal ICA to selectively remove global noise while preserving global signal in functional MRI data. *NeuroImage*, *181*(December 2017), 692–717. https://doi.org/10.1016/j.neuroimage.2018.04.076

Glodzik, L., Randall, C., Rusinek, H., & de Leon, M. J. (2013). Cerebrovascular Reactivity to Carbon Dioxide in Alzheimer's Disease. *Journal of Alzheimer's Disease*, *35*(3), 427–440. https://doi.org/10.3233/JAD-122011

Golestani, A. M., Wei, L. L., & Chen, J. J. (2016). Quantitative mapping of cerebrovascular reactivity using resting-state BOLD fMRI: Validation in healthy adults. *NeuroImage*, *138*, 147–163. https://doi.org/10.1016/j.neuroimage.2016.05.025

Gonzales, M. M., Tarumi, T., Mumford, J. A., Ellis, R. C., Hungate, J. R., Pyron, M., … Haley, A. P. (2014). Greater BOLD response to working memory in endurance-trained adults revealed by breath-hold calibration. *Human Brain Mapping*, *35*(7), 2898–2910. https://doi.org/10.1002/hbm.22372

Gonzalez-Castillo, J., Panwar, P., Buchanan, L. C., Caballero-Gaudes, C., Handwerker, D. A., Jangraw, D. C., … Bandettini, P. A. (2016). Evaluation of multi-echo ICA denoising for task based fMRI studies: Block designs, rapid event-related designs, and cardiac-gated fMRI. *NeuroImage*, *141*, 452–468. https://doi.org/10.1016/j.neuroimage.2016.07.049

Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., … Dosenbach, N. U. F. (2017). Precision Functional Mapping of Individual Human Brains. *Neuron*, *95*(4), 791-807.e7. https://doi.org/10.1016/j.neuron.2017.07.011

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*. https://doi.org/10.1007/978-1-4020-6754-9_1720

Grabner, G., Janke, A. L., Budge, M. M., Smith, D., Pruessner, J., & Collins, D. L. (2006). Symmetric Atlasing and Model Based Segmentation: An Application to the Hippocampus in Older Adults. In R. Larsen, M. Nielsen, & J. Sporring (Eds.), *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2006: 9th International Conference, Copenhagen, Denmark, October 1-6, 2006. Proceedings, Part II* (pp. 58–66). https://doi.org/10.1007/11866763_8

Gracco, V. L., Tremblay, P., & Pike, B. (2005). Imaging speech production using fMRI. *NeuroImage*, *26*(1), 294–301. https://doi.org/10.1016/j.neuroimage.2005.01.033

Gratton, C., Dworetsky, A., Coalson, R. S., Adeyemo, B., Laumann, T. O., Wig, G. S., … Campbell, M. C. (2020). Removal of high frequency contamination from motion estimates in single-band fMRI saves data without biasing functional connectivity. *NeuroImage*, *217*(April), 116866. https://doi.org/10.1016/j.neuroimage.2020.116866

Griffanti, L., Douaud, G., Bijsterbosch, J., Evangelisti, S., Alfaro-Almagro, F., Glasser, M. F., … Smith, S. M. (2017). Hand classification of fMRI ICA noise components. *NeuroImage*, *154*(June 2016), 188–205. https://doi.org/10.1016/j.neuroimage.2016.12.036

Griffanti, L., Salimi-Khorshidi, G., Beckmann, C. F., Auerbach, E. J., Douaud, G., Sexton, C. E., … Smith, S. M. (2014). ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*, *95*, 232–247. https://doi.org/10.1016/j.neuroimage.2014.03.034

Halchenko, Y., Goncalves, M., di Oleggio Castello, M. V., Ghosh, S., Hanke, M., Dae, … Kahn, A. (2019, December). *nipy/heudiconv v0.6.0*. https://doi.org/10.5281/zenodo.3579455

Handwerker, D. A., Gazzaley, A., Inglis, B. A., & D'Esposito, M. (2007). Reducing vascular variability of fMRI data across aging populations using a breathholding task. *Human Brain Mapping*, *28*(9), 846–859. https://doi.org/10.1002/hbm.20307

Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, *66*(1), 51–83. https://doi.org/10.1109/PROC.1978.10837

Hartkamp, N. S., Bokkers, R. P. H., van Osch, M. J. P., de Borst, G. J., & Hendrikse, J. (2017). Cerebrovascular reactivity in the caudate nucleus, lentiform nucleus and thalamus in patients with carotid artery disease. *Journal of Neuroradiology*, *44*(2), 143–150. https://doi.org/10.1016/j.neurad.2016.07.003

Hennig, J., Nauerth, A., & Friedburg, H. (1986). RARE imaging: A fast imaging method for clinical MR. *Magnetic Resonance in Medicine*, *3*(6), 823–833. https://doi.org/10.1002/mrm.1910030602

Hodgson, K., Poldrack, R. A., Curran, J. E., Knowles, E. E., Mathias, S., Göring, H. H. H., … Glahn, D. C. (2017). Shared Genetic Factors Influence Head Motion During MRI and Body Mass Index. *Cerebral Cortex (New York, N.Y. : 1991)*, *27*(12), 5539–5546. https://doi.org/10.1093/cercor/bhw321

Iadecola, C., & Davisson, R. L. (2008). Hypertension and Cerebrovascular Dysfunction. *Cell Metabolism*, *7*(6), 476–484. https://doi.org/10.1016/j.cmet.2008.03.010

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

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6

Jo, H. J., Gotts, S. J., Reynolds, R. C., Bandettini, P. A., Martin, A., Cox, R. W., & Saad, Z. S. (2013). *Effective Preprocessing Procedures Virtually Eliminate Distance-Dependent Motion Artifacts in Resting State FMRI*. *2013*.

Johnstone, T., Ores Walsh, K. S., Greischar, L. L., Alexander, A. L., Fox, A. S., Davidson, R. J., & Oakes, T. R. (2006). Motion correction and the use of motion covariates in multiple-subject fMRI analysis. *Human Brain Mapping*, *27*(10), 779–788. https://doi.org/10.1002/hbm.20219

Jorge, J., Figueiredo, P., van der Zwaag, W., & Marques, J. P. (2013). Signal fluctuations in fMRI data acquired with 2D-EPI and 3D-EPI at 7 Tesla. *Magnetic Resonance Imaging*, *31*(2), 212–220. https://doi.org/10.1016/j.mri.2012.07.001

Kastrup, A., Krüger, G., Neumann-Haefelin, T., & Moseley, M. E. (2001). Assessment of cerebrovascular reactivity with functional magnetic resonance imaging: Comparison of CO2 and breath holding. *Magnetic Resonance Imaging*, *19*(1), 13–20. https://doi.org/10.1016/S0730-725X(01)00227-2

Kastrup, A., Li, T., Takahashi, A., Glover, G. H., & Moseley, M. E. (1998). Functional Magnetic Resonance Imaging of Regional Cerebral Blood Oxygenation Changes During Breath Holding. *Stroke*, *29*(12), 2641–2645. https://doi.org/doi: 10.1161/01.STR.29.12.2641

Krainik, A., Hund-Georgiadis, M., Zysset, S., & Von Cramon, D. Y. (2005). Regional impairment of cerebrovascular reactivity and BOLD signal in adults after stroke. *Stroke*, *36*(6), 1146–1152. https://doi.org/10.1161/01.STR.0000166178.40973.a7

Kundu, P., Brenowitz, N. D., Voon, V., Worbe, Y., Vertes, P. E., Inati, S. J., … Bullmore, E. T. (2013). Integrated strategy for improving functional connectivity mapping using multiecho fMRI. *Proceedings of the National Academy of Sciences*, *110*(40), 16187–16192. https://doi.org/10.1073/pnas.1301725110

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*, 1759–1770. https://doi.org/10.1016/j.neuroimage.2011.12.028

Kundu, P., Voon, V., Balchandani, P., Lombardo, M. V., Poser, B. A., & Bandettini, P. A. (2017). Multi-echo fMRI: A review of applications in fMRI denoising and

analysis of BOLD signals. *NeuroImage*, *154*(March), 59–80. https://doi.org/10.1016/j.neuroimage.2017.03.033

Leoni, R. F., Paiva, F. F., Henning, E. C., Nascimento, G. C., Tannús, A., De Araujo, D. B., & Silva, A. C. (2011). Magnetic resonance imaging quantification of regional cerebral blood flow and cerebrovascular reactivity to carbon dioxide in normotensive and hypertensive rats. *NeuroImage*, *58*(1), 75–81. https://doi.org/10.1016/j.neuroimage.2011.06.030

Leung, J., Kim, J. A., & Kassner, A. (2016). Reproducibility of cerebrovascular reactivity measures in children using BOLD MRI. *Journal of Magnetic Resonance Imaging*, *43*(5), 1191–1195. https://doi.org/10.1002/jmri.25063

Li, X., Morgan, P. S., Ashburner, J., Smith, J., & Rorden, C. (2016). The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of Neuroscience Methods*, *264*, 47–56. https://doi.org/10.1016/j.jneumeth.2016.03.001

Lindquist, M. A., Geuter, S., Wager, T. D., & Caffo, B. S. (2019). Modular preprocessing pipelines can reintroduce artifacts into fMRI data. *Human Brain Mapping*, *40*(8), 2358–2376. https://doi.org/10.1002/hbm.24528

Lipp, I., Murphy, K., Caseras, X., & Wise, R. G. (2015). Agreement and repeatability of vascular reactivity estimates based on a breath-hold task and a resting state scan. *NeuroImage*, *113*, 387–396. https://doi.org/10.1016/j.neuroimage.2015.03.004

Liu, P., Li, Y., Pinho, M., Park, D. C., Welch, B. G., & Lu, H. (2017). Cerebrovascular reactivity mapping without gas challenges. *NeuroImage*, *146*(November 2016), 320–326. https://doi.org/10.1016/j.neuroimage.2016.11.054

Liu, P., Xu, C., Lin, Z., Sur, S., Li, Y., Yasar, S., … Lu, H. (2020). Cerebrovascular reactivity mapping using intermittent breath modulation. *NeuroImage*, *215*(March), 116787. https://doi.org/10.1016/j.neuroimage.2020.116787

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

Lombardo, M. V., Auyeung, B., Holt, R. J., Waldman, J., Ruigrok, A. N. V., Mooney, N., … Kundu, P. (2016). Improving effect size estimation and statistical power with multi-echo fMRI and its impact on understanding the neural systems supporting mentalizing. *NeuroImage*, *142*, 55–66. https://doi.org/10.1016/j.neuroimage.2016.07.022

Lynch, C. J., Power, J. D., Dubin, M., Gunning, F., & Liston, C. (2020). Rapid Precision Functional Mapping of Individuals using Multi-Echo fMRI. *Accepted in the 2020 Organisation of Human Brain Mapping (OHBM) Annual Meeting, Momtreal, Canad*.

Magon, S., Basso, G., Farace, P., Ricciardi, G. K., Beltramello, A., & Sbarbati, A. (2009). Reproducibility of BOLD signal change induced by breath holding. *NeuroImage*, *45*, 702–712. https://doi.org/10.1016/j.neuroimage.2008.12.059

Markello, R. D., Spreng, R. N., Luh, W. M., Anderson, A. K., & De Rosa, E. (2018). Segregation of the human basal forebrain using resting state functional MRI. *NeuroImage*, *173*(October 2017), 287–297. https://doi.org/10.1016/j.neuroimage.2018.02.042

Markus, H., & Cullinane, M. (2001). Severely impaired cerebrovascular reactivity predicts stroke and TIA risk in patients with carotid artery stenosis and occlusion. *Brain*, *124*(3), 457–467. https://doi.org/10.1093/brain/124.3.457

Marques, J. P., Kober, T., Krueger, G., Van Der Zwaag, W., Van De Moortele, P.-F., & Gruetter, R. (2009). MP2RAGE, a self bias-field corrected sequence for improved segmentation and T 1 -mapping at high field. *NeuroImage*, *49*, 1271–1281. https://doi.org/10.1016/j.neuroimage.2009.10.002

Marshall, O., Lu, H., Brisset, J. C., Xu, F., Liu, P., Herbert, J., … Ge, Y. (2014). Impaired cerebrovascular reactivity in multiple sclerosis. *JAMA Neurology*, *71*(10), 1275–1281. https://doi.org/10.1001/jamaneurol.2014.1668

Maziero, D., Rondinoni, C., Marins, T., Stenger, V. A., & Ernst, T. (2020). Prospective motion correction of fMRI: Improving the quality of resting state data affected by large head motion. *NeuroImage*, *212*(August 2019), 116594. https://doi.org/10.1016/j.neuroimage.2020.116594

Mcgraw, K. O., & Wong, S. P. (1996). Forming Inferences About Some Intraclass Correlation Coefficients. *Psychological Methods*, *1*(1), 30–46. Retrieved from papers2://publication/uuid/D94ABE24-4E43-4996-83B0-32BE18139462

Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E., Strupp, J., Harel, N., & Uğurbil, K. (2010). 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. *Magnetic Resonance in Medicine*, *63*(5), 1144–1153. https://doi.org/10.1002/mrm.22361

Moia, S., Stickland, R. C., Ayyagari, A., Termenon, M., Caballero-gaudes, C., & Bright, M. G. (2020). Voxelwise optimization of hemodynamic lags to improve

regional CVR estimates in breath-hold fMRI. *Accepted in Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Montreal, Canada;*

Mumford, J. A., Poline, J. B., & Poldrack, R. A. (2015). Orthogonalization of regressors in fMRI models. *PLoS ONE*, *10*(4), 1–11. https://doi.org/10.1371/journal.pone.0126255

Murphy, K., Harris, A. D., & Wise, R. G. (2011). Robustly measuring vascular reactivity differences with breath-hold: Normalising stimulus-evoked and resting state BOLD fMRI data. *NeuroImage*, *54*(1), 369–379. https://doi.org/10.1016/j.neuroimage.2010.07.059

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

Pais-Roldán, P., Biswal, B., Scheffler, K., & Yu, X. (2018). Identifying respiration-related aliasing artifacts in the rodent resting-state fMRI. *Frontiers in Neuroscience*, *12*(NOV), 1–14. https://doi.org/10.3389/fnins.2018.00788

Parkes, L., Fulcher, B., Yücel, M., & Fornito, A. (2018). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *NeuroImage*, *171*(July 2017), 415–436. https://doi.org/10.1016/j.neuroimage.2017.12.073

Peng, S.-L., Yang, H.-C., Chen, C.-M., & Shih, C.-T. (2019). Short- and long-term reproducibility of BOLD signal change induced by breath-holding at 1.5 and 3 T. *NMR in Biomedicine*. https://doi.org/10.1002/nbm.4195

Pinto, J., Jorge, J., Sousa, I., Vilela, P., & Figueiredo, P. (2016). Fourier modeling of the BOLD response to a breath-hold task: Optimization and reproducibility. *NeuroImage*, *135*, 223–231. https://doi.org/10.1016/j.neuroimage.2016.02.037

Poser, B. A., Versluis, M. J., Hoogduin, J. M., & Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine*, *55*(6), 1227–1235. https://doi.org/10.1002/mrm.20900

Posse, S., Wiese, S., Gembris, D., Mathiak, K., Kessler, C., Grosse-Ruyken, M. L., … Kiselev, V. G. (1999). Enhancement of BOLD-contrast sensitivity by single-shot multi-echo functional MR imaging. *Magnetic Resonance in Medicine*, *42*(1), 87–

97. https://doi.org/10.1002/(SICI)1522-2594(199907)42:1<87::AID-MRM13>3.0.CO;2-O

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, *59*(3), 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Power, J. D., Lynch, C. J., Silver, B. M., Dubin, M. J., Martin, A., & Jones, R. M. (2019). Distinctions among real and apparent respiratory motions in human fMRI data. *NeuroImage*, *201*(April), 116041. https://doi.org/10.1016/j.neuroimage.2019.116041

Power, J. D., Silver, B. M., Silverman, M. R., Ajodan, E. L., Bos, D. J., & Jones, R. M. (2019). Customized head molds reduce motion during resting state fMRI scans. *NeuroImage*, *189*(October 2018), 141–149. https://doi.org/10.1016/j.neuroimage.2019.01.016

Prilipko, O., Huynh, N., Thomason, M. E., Kushida, C. A., & Guilleminault, C. (2014). An fMRI study of cerebrovascular reactivity and perfusion in obstructive sleep apnea patients before and after CPAP treatment. *Sleep Medicine*, *15*(8), 892–898. https://doi.org/10.1016/j.sleep.2014.04.004

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., Rooij, D. Van, Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. Supplementary Material. *NeuroImage*.

Pruim, R. H. R. R., Mennes, M., Rooij, D. Van, Llera, A., Buitelaar, J. K., Beckmann, C. F., … 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

Puckett, A. M., Bollmann, S., Poser, B. A., Palmer, J., Barth, M., & Cunnington, R. (2018). Using multi-echo simultaneous multi-slice (SMS) EPI to improve functional MRI of the subcortical nuclei of the basal ganglia at ultra-high field (7T). *NeuroImage*, *172*(December 2017), 886–895. https://doi.org/10.1016/j.neuroimage.2017.12.005

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

Salimi-Khorshidi, G., Smith, S. M., & Nichols, T. E. (2011). Adjusting the effect of nonstationarity in cluster-based and TFCE inference. *NeuroImage*, *54*(3), 2006–2019. https://doi.org/10.1016/j.neuroimage.2010.09.088

Schulz, J., Siegert, T., Bazin, P. L., Maclaren, J., Herbst, M., Zaitsev, M., & Turner, R. (2014). Prospective slice-by-slice motion correction reduces false positive activations in fMRI with task-correlated motion. *NeuroImage*, *84*, 124–132. https://doi.org/10.1016/j.neuroimage.2013.08.006

Setsompop, K., Gagoski, B. A., Polimeni, J. R., Witzel, T., Wedeen, V. J., & Wald, L. L. (2012). Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic Resonance in Medicine*, *67*(5), 1210–1224. https://doi.org/10.1002/mrm.23097

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Šidák, Z. (1967). Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association*, *62*(318), 626–633.

Smith, S. M., Miller, K. L., Moeller, S., Xu, J., Auerbach, E. J., Woolrich, M. W., … Ugurbil, K. (2012). Temporally-independent functional modes of spontaneous brain activity. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(8), 3131–3136. https://doi.org/10.1073/pnas.1121329109

Smyser, C. D., Inder, T. E., Shimony, J. S., Hill, J. E., Degnan, A. J., Snyder, A. Z., & Neil, J. J. (2010). Longitudinal analysis of neural network development in preterm infants. *Cerebral Cortex*, *20*(12), 2852–2862. https://doi.org/10.1093/cercor/bhq035

Soltysik, D. A., & Hyde, J. S. (2006). Strategies for block-design fMRI experiments during task-related motion of structures of the oral cavity. *NeuroImage*, *29*(4), 1260–1271. https://doi.org/10.1016/j.neuroimage.2005.08.063

Sotiropoulos, S. N., Moeller, S., Jbabdi, S., Xu, J., Andersson, J. L., Auerbach, E. J., … Lenglet, C. (2013). Effects of image reconstruction on fiber orientation mapping from multichannel diffusion MRI: Reducing the noise floor using SENSE. *Magnetic Resonance in Medicine*, *70*(6), 1682–1689. https://doi.org/10.1002/mrm.24623

Sousa, I., Vilela, P., & Figueiredo, P. (2014). Reproducibility of hypocapnic cerebrovascular reactivity measurements using BOLD fMRI in combination with a paced deep breathing task. *NeuroImage*, *98*, 31–41. https://doi.org/10.1016/j.neuroimage.2014.04.049

Stickland, R., Ayyagari, A., Zvolanek, K., & Bright, M. G. (2020). *Short breathing tasks at the start of a resting state fMRI scan: feasible measures of cerebrovascular reactivity and hemodynamic lag (ISMRM2020: Oral Power Pitch, #1112)*. https://doi.org/10.13140/RG.2.2.23894.88644

Tancredi, F. B., & Hoge, R. D. (2013). Comparison of cerebral vascular reactivity measures obtained using breath-holding and CO 2 inhalation. *Journal of Cerebral Blood Flow and Metabolism*, *33*(7), 1066–1074. https://doi.org/10.1038/jcbfm.2013.48

Tchistiakova, E., Anderson, N. D., Greenwood, C. E., & Macintosh, B. J. (2014). Combined effects of type 2 diabetes and hypertension associated with cortical thinning and impaired cerebrovascular reactivity relative to hypertension alone in older adults. *NeuroImage: Clinical*, *5*, 36–41. https://doi.org/10.1016/j.nicl.2014.05.020

The phys2bids developers, Alcalá, D., Ayyagari, A., Bright, M., Ferrer, V., Gaudes, C. C., … Zvolanek, K. (2019, December). *physiopy/phys2bids: BIDS formatting of physiological recordings*. https://doi.org/10.5281/zenodo.3586045

Thomason, M. E., Burrows, B. E., Gabrieli, J. D. E., & Glover, G. H. (2005). Breath holding reveals differences in fMRI BOLD signal in children and adults. *NeuroImage*, *25*(3), 824–837. https://doi.org/10.1016/j.neuroimage.2004.12.026

Tong, Y., Bergethon, P. R., & Frederick, B. de B. (2011). An improved method for mapping cerebrovascular reserve using concurrent fMRI and near-infrared spectroscopy with Regressor Interpolation at Progressive Time Delays (RIPTiDe). *NeuroImage*, *56*(4), 2047–2057. https://doi.org/10.1016/j.neuroimage.2011.03.071

Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., … Avants, B. B. (2014). Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *NeuroImage*, *99*, 166–179. https://doi.org/10.1016/j.neuroimage.2014.05.044

Urback, A. L., MacIntosh, B. J., & Goldstein, B. I. (2017). Cerebrovascular reactivity measured by functional magnetic resonance imaging during breath-hold challenge: A systematic review. *Neuroscience and Biobehavioral Reviews*, *79*(April), 27–47. https://doi.org/10.1016/j.neubiorev.2017.05.003

Van Oers, C. A. M. M., Van Der Worp, H. B., Kappelle, L. J., Raemaekers, M. A. H., Otte, W. M., & Dijkhuizen, R. M. (2018). Etiology of language network changes during recovery of aphasia after stroke. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-19302-4

Vannesjo, S. J., Wilm, B. J., Duerst, Y., Gross, S., Brunner, D. O., Dietrich, B. E., … Pruessmann, K. P. (2015). Retrospective correction of physiological field fluctuations in high-field brain MRI using concurrent field monitoring. *Magnetic Resonance in Medicine*, *73*(5), 1833–1843. https://doi.org/10.1002/mrm.25303

Webster, M. W., Makaroun, M. S., Steed, D. L., Smith, H. A., Johnson, D. W., & Yonas, H. (1995). Compromised cerebral blood flow reactivity is a predictor of stroke in patients with symptomatic carotid artery occlusive disease. *Journal of Vascular Surgery*, *21*(2), 338–345. https://doi.org/10.1016/S0741-5214(95)70274-1

Wilm, B. J., Nagy, Z., Barmet, C., Vannesjo, S. J., Kasper, L., Haeberlin, M., … Pruessmann, K. P. (2015). Diffusion MRI with concurrent magnetic field monitoring. *Magnetic Resonance in Medicine*, *74*(4), 925–933. https://doi.org/10.1002/mrm.25827

Xu, Y., Tong, Y., Liu, S., Chow, H. M., AbdulSabur, N. Y., Mattay, G. S., & Braun, A. R. (2014). Denoising the speaking brain: Toward a robust technique for correcting artifact-contaminated fMRI data under severe motion. *NeuroImage*, *103*, 33–47. https://doi.org/10.1016/j.neuroimage.2014.09.013

Yezhuvath, U. S., Lewis-Amezcua, K., Varghese, R., Xiao, G., & Lu, H. (2009). On the assessment of cerebrovascular reactivity using hypercapnia BOLD MRI. *NMR in Biomedicine*, *22*(7), 779–786. https://doi.org/10.1002/nbm.1392

Zacà, D., Jovicich, J., Nadar, S. R., Voyvodic, J. T., & Pillai, J. J. (2014). Cerebrovascular reactivity mapping in patients with low grade gliomas undergoing presurgical sensorimotor mapping with BOLD fMRI. *Journal of*

Magnetic Resonance Imaging, *40*(2), 383–390. https://doi.org/10.1002/jmri.24406

Zaitsev, M., Akin, B., LeVan, P., & Knowles, B. R. (2017). Prospective motion correction in functional MRI. *NeuroImage*, *154*(November 2016), 33–42. https://doi.org/10.1016/j.neuroimage.2016.11.014

Ziyeh, S., Rick, J., Reinhard, M., Hetzel, A., Mader, I., & Speck, O. (2005). Blood oxygen level-dependent MRI of cerebral CO2 reactivity in severe carotid stenosis and occlusion. *Stroke*, *36*(4), 751–756. https://doi.org/10.1161/01.STR.0000157593.03470.3d