Denoising Task-Based fMRI data:
a comparison between methods

Jur Kersten
4168763

Under supervision of

Wouter Schellekens,
Natalia Petridou &
Stefan van der Stigchel

October 2020

Abstract

Over the last decade BOLD-response related fMRI has seen impressive progress in resolution of the imaging technique. This progress however is limited by the degree to which fMRI can detect localised hemodynamic responses linked to activity and disconnect this from noise. Important noise factors that contribute to the BOLD response include respiration and cardiac cycle. In this research a comparison is made between different methods that pertain to the untanglement of signal from noise, i.e. denoising. Different methods are adopted using Principal Component Analysis (GLMDenoise), Independent Component Analysis (FSL Fix & GLM ICA), and a method using motion and physiological parameters (GLM mp-ph). This in an effort to retrieve signal of interest in a High Field (7T) fMRI visual experiment, comparing the effectiveness and implementational ease of each method. Unfortunately, no conclusive evidence could be reported for the methods efficiently and consistently denoising the data. These findings might be caused by methodological rather than neurovascular issues. Another possible cause are non-linearities in the data and suggest Machine Learning implementations might be useful in future research on fMRI and denoising.

Introduction

Functional magnetic resonance imaging (fMRI) is the most commonly used technique to investigate brain activity in humans. Its non-invasiveness means the technique is applicable to both the healthy population and patient groups, aiding the gathering of neural data and making data-driven methods and quantitative research possible for neural activity. The technique used in fMRI to measure neural activity is known as the Blood Oxygen Level-Dependent (BOLD) contrast (See Appendix for a more detailed introduction into fMRI & BOLD-contrast). This technique is sensitive to differences in blood oxygenation levels and depends on local metabolic demands to restore potentials and neurotransmitters after activity (Logothetis, 2008). The basis for BOLD-contrast is that changes in the ratio of paramagnetic deoxyhemoglobin to diamagnetic oxyhemoglobin in a localised area suggests neuronal activity preceded , (Logothetis & Wandell, 2004). A downside of BOLD fMRI measurements is that neuronal activity is not the only factor that can influence local concentrations of blood oxygenation. Identifying these other factors (i.e. noise) and separating them from the signal of interest is the aim of this research.

BOLD fMRI can detect the difference in blood-oxygenation due to the difference in magnetic properties of oxygen-rich and oxygen-poor blood dependent on the concentration of hemoglobin. This means that BOLD is sensitive to local concentrations of blood oxygenation, but also dependent on the total amount of deoxyhemoglobin (dHB) in the blood. Therefore the BOLD-signal is not only made up of neural activity but also of factors that alter cerebral blood flow and metabolism in the form of (physiological) noise and interference (Hillman, 2014; Hutton et al., 2011; Krüger & Glover, 2001; Triantafyllou et al., 2005). Three primary physiological noise-sources are (1) changes in cerebral blood flow due to variations in respiratory activity, (2) variations in the cardiac rate that may influence cerebral blood flow, volume, and oxygenation, and (3) processes that cause intrinsic fluctuations in blood flow and metabolism, including variations in blood pressure and resting-state neural activity. Furthermore, these physiological sources can alter the magnetic field (e.g. movement during respiratory activity (Liu, 2016).

Additionally, there is not a one-on-one relationship between fMRI data and a single neuron. In fMRI, the brain is divided into voxels (i.e. volumetric pixels, in this dataset the size of roughly one cubic millimetre) and given that a human brain contains 100 billion neurons (Herculano-Houzel, 2009), each cubic millimetre represents about 100.000 neurons. So, although the BOLD signal is intrinsically linked to the physiological and metabolic processes in the brain that modulate blood flow and neural activity is one of those processes, it is hard to detect and distinguish this in the data. To put it into perspective, a hemodynamic response function (HRF) is used to model the evoked response to a neural event. The HRF typically only accounts for a small percentage of signal change (2-5%) relative to the baseline BOLD signal (Bianciardi et al., 2009). Defining task-related BOLD changes is most often accomplished by fitting a General linear model (GLM) to the data, whereby the stimulus time course is convolved with a pre-specified canonical HRF (Friston, Jezzard, & Turner, 1994). For the reliability of this prediction, linearity of the BOLD signal to neuronal activity is critical (Siero et al., 2013) as a GLM cannot explain empirically observed non-linearities in the BOLD responses (Birn, Saad, & Bandettini, 2001).

Other non-neuronal contributions to the BOLD signal include; thermal noise inherent to the coils and electrical current in magnetic resonance imaging, instrumental drifts, artefacts due to hardware instabilities and signal changes due to head motion (Murphy, Birn, & Bandettini, 2013). All these different noise contributions have an effect on the measurement but can also have an effect on each other. This makes it

inherently difficult to filter noise from the signal of interest and represent the true activity in the brain. So, despite BOLD fMRI's promise and broad implementation within both scientific and medical fields extensive processing is needed in order to be able to interpret the data. Over the last decade, BOLD-response related fMRI has seen impressive progress in spatial resolution of the imaging technique, especially fMRI at high field strengths shows promise for research on specialized regions of the brain (Siero et al., 2013). Moreover, steps have been made to suppress the different kinds of machine noise: reducing thermal noise (Bartha et al., 2002), improving the effectiveness of multichannel coil arrays (Zwart et al., 2003), and monitoring drifts and spikes in the magnetic field in order to make the machine itself more stable.

However, noise still remains an issue, particularly physiological noise, with two primary components being the cardiac and respiratory cycle. During the cardiac cycle, blood volume in different places in the brain increases and decreases following the arterial pulse which results in dilation of vessels, which also influences cerebrospinal fluid (CSF). In the brain this forms an ebb and flow, with blood and CSF moving around continuously, which influences blood oxygen levels throughout the brain. Likewise, heart-rate variability and blood pressure changes alter the measurement of the BOLD-signal as well (Chang, Cunningham, & Glover, 2009). Additionally, the respiratory cycle causes signal fluctuations in various ways. Changes in respiratory volume have an effect on oxygen levels in the blood as well as on carbon-dioxide (CO2) levels, which acts as a vasodilator (Birn, Smith, Jones, & Bandettini, 2008). Along with this, the motion of the chest during breathing alters the very sensitive magnetic field, which can provide a significant component of physiological noise (Chang & Glover, 2009). One can imagine that it is hard to overcome these noise sources a priori: it is impossible to prevent a participant from breathing and having a heartbeat during scanning and thus the fluctuations in the rate of both pulse and breathing as well as the expansion and contraction of blood vessels. Therefore, these noise factors need to be dealt with after data acquisition. However, at this point the different noise factors are blended in with the signal and each other, which makes disentanglement of noise from signal of interest a challenging task.

A variety of noise filtering techniques have been proposed and implemented, with varying results. Multiple of these approaches are based on the extraction and estimation of signals that describe noise fluctuations, so called nuisance signals, and the process of nuisance regression. In most approaches, the final step in denoising is adding these estimated noise signals as a regressor in the design of a model that is fitted to the data. This GLM, which should describe the progression of noise over time can then be 'subtracted' from the data, which should leave you with denoised data (Caballero-Gaudes & Reynolds, 2017). This paper will describe and explain several of these noise filtering techniques, and implement them on data gathered by Nota, Piantoni & Petridou (2019) which focused on activation of the early visual area of the brain in an experiment that implemented different visual stimuli.

To start off, certain parameters can be derived from preprocessing steps such as estimated head movement. Space in an MRI machine during scanning is divided into voxels, and the BOLD-response is then measured within each of these voxels. When a participant moves, albeit slightly, this shift can cause thousands of neurons to shift from one voxel to another. Therefore, head movement correction is always a necessary preprocessing step in fMRI. Motion correction can be seen as a realignment of data which tries to make sure voxels stay within the same space. Not all movement related problems are fixed by motion correction as movement interacts with the field causing inhomogeneity of the magnetic field which can lead to both dropouts and distortions resulting in both negative and positive fluctuations of the signal.

Applying a linear regression on the BOLD signal using the movement parameters, or on parameters such as respiratory or cardiac rate is an easy and cost-efficient method for filtering interaction effects of movement and field inhomogeneity. These parameters are fitted to the data relatively easily and without much further processing, and can in some cases be an effective way of denoising the data (Kalthoff, Seehafer, Po, Wiedermann, & Hoehn, 2011). Chen, Jahanian & Glover (2017) however call this process into question, doubting whether a linear regression based on simple parameters can succeed in splitting noise from signal of interest as, in practice, non-linear variance occurs because spatial and spectral characteristics of motion artefacts differ among types and levels of head movements.

An often-used technique is Principal Component Analysis (PCA), which is a process with the goal of identifying patterns in data. From a dataset, PCA tries to detect correlations between variables and vectorises these variables. This is done in order to find vectors that maximise explained variance, and subsequently to project the variance onto a smaller dimensional space while preserving most of the information . The vectors that are formed first, correspond to the direction with the greatest variance in the data (Behzadi, Restom, Liau, & Liu, 2007; Muschelli et al., 2014). PCA tries to put the maximum possible information in the first component, then the maximum remaining information (i.e. orthogonal to the previous component) in the next component and so on. In other words, each component does not explain variance that has already been explained by other components. This method gives reduced dimensionality without losing much information by discarding the components with low information and considering the remaining components as your new variables. A method which utilises PCA is GLMDenoise, which is an automated technique made for denoising task-based fMRI data using minimal assumptions (Kay, Rokem, Winawer, Dougherty, & Wandell, 2013). In this technique the PCA is run on a voxel group that should contain no task-related voxels. Kay and colleagues report promising results on a variety of datasets. Even though using a non-task related voxel group is a common workaround for PCA, a limitation of this approach is that, in defining a pool of task-unrelated voxels, also the task-related voxels are analysed and selected (to not be part of this pool). Which is a form of circular analysis since the same data is used twice, with the risk of shaping the outcome before any analysis is done. Circular analysis is a bad practice since it can unjustifiably inflate statistical strength of any reported result and, in the worst case, lead to significant results being found in data that consists only of noise (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). Furthermore, there is also a risk of selecting false negative voxels, resulting in the nuisance regressors explaining task-related activity. This is a problem that is reported in different approaches for defining voxel-specific regressors using masks as well (Behzadi et al., 2007; Bianciardi et al., 2009; Soltysik, Thomasson, Rajan, & Biassou, 2015; Zwart, Gelderen, Fukunaga, & Duyn, 2008).

An alternative strategy for denoising fMRI data is Independent Component Analysis (ICA). ICA works from a similar base as PCA, finding components that maximise explained variance and removing correlations to compress data. Contrary to PCA, in ICA all components are equally important and its vectors are non-orthogonal. Where PCA compresses data, ICA separates data whereby each component can be seen as its own signal. A disadvantage of both of these methods is that they do not distinguish between voxels that are related to some task and those that are related to noise factors. Therefore, the explained variance of each component could contain a mixture of noise signals as well as signal of interest. Once the ICA is computed, the denoising process begins by distinguishing which components are linked to task-related neuronal activity and which components are related to noise. The noisy components can then be removed leaving a denoised dataset (Mckeown, 2003). In practice, using

ICA to identify noise components requires the manual selection of components that show noise and components that show task-related activity, which is very time consuming, difficult to reproduce and, additionally, requires expertise (Kelly et al., 2010). Hence, numerous procedures have been proposed in order to autoclassify the 'good' from the 'bad' components: clustering methods, naïve Bayes, logistic regression, support vector machines, decision trees, random forests or even a combination of several classifiers are among the ones named. One of the methods which uses a combination of classifiers is FSL FIX (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014). To capitalize on the advantages of different classifiers and compensate for their weaknesses, FSL FIX uses a k-nearest neighbour algorithm (k-NN), decision tree and a Support Vector Machine (SVM). A k-NN is reliable locally, but has difficulties to capture patterns in a full dataset, decision trees learn complex decision boundaries well and SVMs are capable of finding decision boundaries with the best chance of generalisation. FIX utilises an ensemble technique known as stacking (Wolpert, 1992), in which outputs of the individual classifiers become the input at higher levels. This stack of classifiers trains on several FLS Melodic ICAs, in which the components are manually classified as either noise or task-related. FSL FIX generates a number of spatial and temporal features for a set of noise-related ICA components. The ICAs are then split into different datasets using these features (full, feature-selected, temporal, spatial, temporal-feature-selected and spatial-feature-selected). Training is done on each dataset, and the higher-level classifier is trained based on the probabilistic output of the previously described stacked lower level classifiers (vector of probabilities whose range is from 0 i.e. perfect noise to 1 i.e. perfect signal). The final output is a single probability for each IC being noise/signal. For generalisation and to minimise the risk of over-fitting, FIX is tested using the leave-one-out (LOO) approach. After training, FSL FIX can be run on a Melodic ICA, which will generate a noise/signal probability for each IC and, using a threshold specified by the user remove the ICs deemed as noise from the data. Along with running FSL FIX on the dataset, a second method using an ICA can be constructed using manually selected ICs which are classified as noise. These components can then be used as parameters and fitted in a GLM and regressed out of the data similarly to the motion- and physiological parameters.

The aim of this research is denoising fMRI data of a visual experiment. This is done using the four different techniques described above using: *Motion and physiological parameters, hand-picked ICs, GLMDenoise* and *FSL Fix*. Analyses will be conducted not only on the merits of these techniques for the data, but also on their use and implementation efficiency, both computationally and in a user perspective. This, to answer the questions: Are the proposed methods effective in denoising data? And is it worth implementing one of the methods over the others?

Evidence supports the notion of (linear) relationships between respiration and hemodynamics to BOLD response (Chang & Glover, 2009; Siero et al., 2013). This suggests that denoising can be effective in cleaning up the data, but it is conditional to whether the proposed techniques are able to catch these relationships (and not task-related signal). Both GLMDenoise (Kay et al., 2013) and FSL Fix (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014) report promising results and are quite extensive techniques. Therefore a natural assumption would be that these methods outperform the other, 'simpler', methods. Generalizing techniques over different datasets and accurately recovering true task-evoked changes in BOLD signal has however proven to be difficult (Lindquist, Loh, Atlas, & Wager, 2009). So, whether these methods are effective on this particular dataset and whether their performance warrants further implementation merits investigation.

Methods

Scan protocol

The data used in this research was collected by Nota et al., (2019), using a 7T Philips Achieva scanner with a 32-channel high-density surface receive array (MRCoils BV, the Netherlands). The scanning sessions consisted of anatomical and functional scans. During the anatomical scans, structural images of partial brain coverage were obtained consisting of 50 slices covering the posterior occipital cortex (where the early visual area of the brain is located). This was recorded with a spatial resolution of 0.71 × 0.71 × 0.80 mm, field-of-view (FOV) (ap,fh,rl) = 40 × 160 × 40 mm, TR/TE = 7/3ms. Functional data was collected using Gradient Echo EPI with SENSE factor = 3.0, TR/TE = 850/27ms, flip angle = 60°, number of slices = 7, FOV = 7 × 128 × 128 mm covering the early visual areas, voxel size = 1.0 mm3 isotropic. Throughout all scans, the subjects' peripheral pulse and respiratory rate were recorded using a pulse oximeter and a respiratory belt.

Functional scans

During the session, a Hemodynamic Response Function (HRF) scan, four temporal pattern scans and two spatial pattern scans were completed. All scans contained different visual stimuli which were generated in Matlab (Mathworks, Natick, MA, USA) using PsychToolbox. The stimuli were windowed with a circular aperture and centred at fixation, with the remaining space in the window filled with neutral grey. In order to check and maintain subject attention, a fixation cross was positioned in the centre of each image, with the colour of this cross alternating between green and red at differing and randomised intervals. Reporting this change of colour was done by the subjects by pressing a button. The average performance across runs and subjects was 80% (Nota et al., 2019 for more specifics).

In the HRF-experiment only one condition was used, so trials differed somewhat in what the pattern looked like but not in duration, inter-stimulus interval (ISI), contrast, or intensity. The purpose of the HRF-condition was to investigate the hemodynamic response function per participant caused by a brief visual stimulus presented numerous times (32 repetitions) a (see figure 1.d). The temporal pattern experiment had 12 conditions which varied in duration (1:6) and ISI (7:12), both ascending from 17 ms to 533 ms, doubling at each condition. All conditions were randomised.

The spatial pattern experiment also had 12 conditions. However, these conditions varied in contrast, pattern type and sparsity. Five conditions varied in contrast ascending from low contrast to high contrast, three conditions show different patterns from the randomly oriented contrast pattern that can be seen in the other trials called grating, plaid and circular. The last four conditions have a descending amount of sparsity within the trials (see figure fig:stimuli).

Preprocessing

Certain preprocessing steps needed to be taken in order for the data to be ready for efficient denoising. The first step in preprocessing was motion correction, after which a correction for Echo-Planar Imaging (EPI) distortions was done. Both these steps were taken with the use of AFNI (Cox, 1996; Cox & Hyde, 1997). Also, a co-registration to the structural volume of the T1 scan was performed using ANTs (Avants et al., 2011). In order to interpolate and therefore smooth the data as little as possible all these preprocessing steps were combined, resulting in the use of only one interpolation step. For signal quality to be improved before denoising, slow scanner drifts were removed using a high pass filter with a cut-off at 0.01Hz. These slow scanner drifts are often thought to be caused by physiological noise or scanner-related noise (Bandettini, Jesmanowicz, Wong, & Hyde, 1993; Lee, Glover, & Meyer, 1995; A. M. Smith et al.,

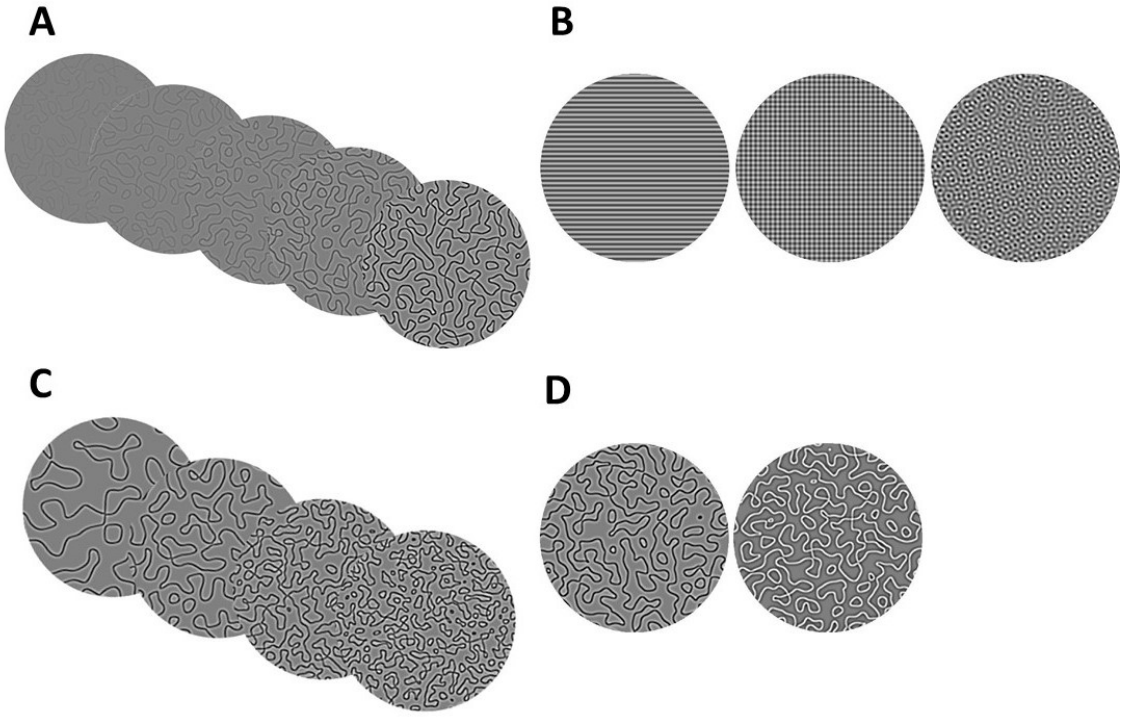

Figure 1: Examples of visual stimuli used in the experiment, with different stimuli pertaining to (A) contrast conditions, (B) grating, plaid and circular conditions and (C) sparsity conditions in the spatial pattern experiment and (D) an example stimuli of the HRF scan.

1999). With this, the data was also rescaled to percent signal change.

Afterwards, all denoising methods were performed and the results of each method were analysed and compared to the non-denoised dataset (baseline). Both the temporal and the spatial pattern scans had 12 conditions. These were each convolved using a canonical HRF, which depicts the typical BOLD response over time related to neuronal events which was done to accurately describe the form and, therefore, measure the impact that these conditions had on the data. The model timeseries for each of the conditions were fitted to the data using a linear regression in order to calculate regression coefficients (β-coefficients) and residuals for each condition and each voxel within the early visual cortex (V1, V2 & V3). On the basis of the regression coefficients, t-values were calculated for each voxel. In order to correct for the *multiple comparisons problem*, which is an issue found in fMRI due to the often large number of voxels that need to be analysed leading to a high chance of a type 1 error (Genovese, Lazar, & Nichols, 2002). False Discovery Rate (FDR) was used to correct for this problem and select a threshold for voxels to be included in the comparison at a more reasonable chance of false positives. In order to compare the results of different methods the same voxels needed to be used, for this reason the FDR-threshold was calculated on the basis of the HRF-scan (alpha = 0.05, dof = 399) with three different regions of interest within early visual cortex (V1, V2 &

V3). This resulted in a group of voxels showing significant activity, which varied per participant (356, 978 and 727 respectively).

Denoising

During the scans, six motion parameters were extracted by estimating rotation and translation along three axes: the 3 spatial dimensions. Even though during preprocessing *motion-correction* was already applied, making sure that each voxel stays the same, motion by itself causes fluctuations in the magnetic field during fMRI which, therefore, adds noise to the data. As mentioned before, cardiac and respiratory rates are also contributors to noise (Chang et al., 2009). Therefore, the recorded cardiac and respiratory rates were also added to the general linear model, resulting in a total of eight parameters for the first method of denoising. A general linear model was fitted for all voxels and each parameter. This resulted in beta-coefficients for each parameter in each voxel, which indicated the amplitude of the parameters' response. The beta-coefficients were then used to calculate the estimated contribution of each parameter to the signal and applied in a nuisance regression, which subtracts the estimated contribution of these noise parameters from the signal.

GLMdenoise is an automated denoising technique. More specifically, it requires both data and task designs in order to estimate the hemodynamic response function (HRF). The algorithm calculates beta weights for the designs and regressors that describe the baseline signal, selecting voxels that should be unrelated to the experiment (using cross validation as a check), after which a PCA is run on this task-unrelated group of voxels trying to prevent the task-signal from being included in the noise PCA. An added advantage of using GLMDenoise over other techniques would be the fact that GLMdenoise uses cross-validation on the number of components used as nuisance regressors, which should therefore automatically determine the appropriate number of elements to include. The downside to this is that multiple fMRI scan sessions/runs are needed. In the end, the algorithm comes up with principal components to include and uses these as nuisance regressors resulting in denoised data and what is left should better describe the 'true' neural activity in the data (Kay et al., 2013).

FMRIB Software Library (FSL) is collection of image analysis and statistical tools for MRI brain imaging data, including functional imaging (S. M. Smith et al., 2004). Using FSL Melodic with full registration settings utilising a structural image of the brain and no spatial smoothing, an Independent Component Analysis was performed. The generated components were then analysed, and the bad (i.e. noisy) components were selected to serve as nuisance regressors in a similar way compared to motion- and physiological parameters in the aforementioned GLM-nuisance regression approach. For most participants, the number of independent components that could be classified as noisy varied between 10-15 out of ~ 50 components. Example output of a threshold map of one component can be found in the supplementary material.

An alternative to hand-picking components from an ICA and using this in denoising is FSL FIX (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014) which, given an ICA as input, automatically scores components on a scale from bad to good (range of 0-1). In order to run FSL FIX, a training set was constructed, and a threshold parameter set at 15 which represents the cut-off for including components based on their given score. A value within the range of 5-20 is recommended by the makers, with lower values causing less false positives (i.e. less components falsely classified as noise) at the expense of more false negatives. After this, FSL FIX will clean the data based on the components with optional motion clean-up and high pass filtering, as well as a choice between unique variance clean-up or more aggressive full variance clean-up. The difference between these clean-up methods is that full variance clean-up does not account for vari-

ance within a component which might also be (partly) explained by a good component. In unique variance clean-up, an estimation of the contribution of both good and bad components is made in order to identify the noise specific variance. The chosen severity for clean-up determines to which extent signals, both noise and signal of interest, will be affected.

Analysis

To test whether signal quality improves and to determine the effect that the noise-clean-up methods have on the data different metrics will be used. The first is the temporal signal-to-noise ratio (tSNR), calculated for each method as well as the baseline. The tSNR is defined as follows:

$$tSNR = \frac{\overline{S}}{\sigma}$$

where $\overline{S}$ is the mean signal of the time series and σ is the standard deviation, both calculated for every voxel. The temporal and spatial patterns contained events that should theoretically lead to an increase in activity. For the temporal patterns these were the duration- and ISI-conditions and for the spatial patterns these were the contrast, pattern shape and sparsity conditions. For each of these four conditions, a regression coefficient can be calculated representing the strength of the task conditions to the signal. T-statistics were calculated on the task regression coefficients to test if the task conditions changed significantly relative to baseline, per denoising method. For these analyses, SPSS (version 26) will be used to run a linear mixed model (LMM), using the methods as fixed effects and including the participants as well as the scan sessions as random effects. Using an LMM, analyses can be done on the whole group whilst both inter-subject and inter-scan effects can be caught.

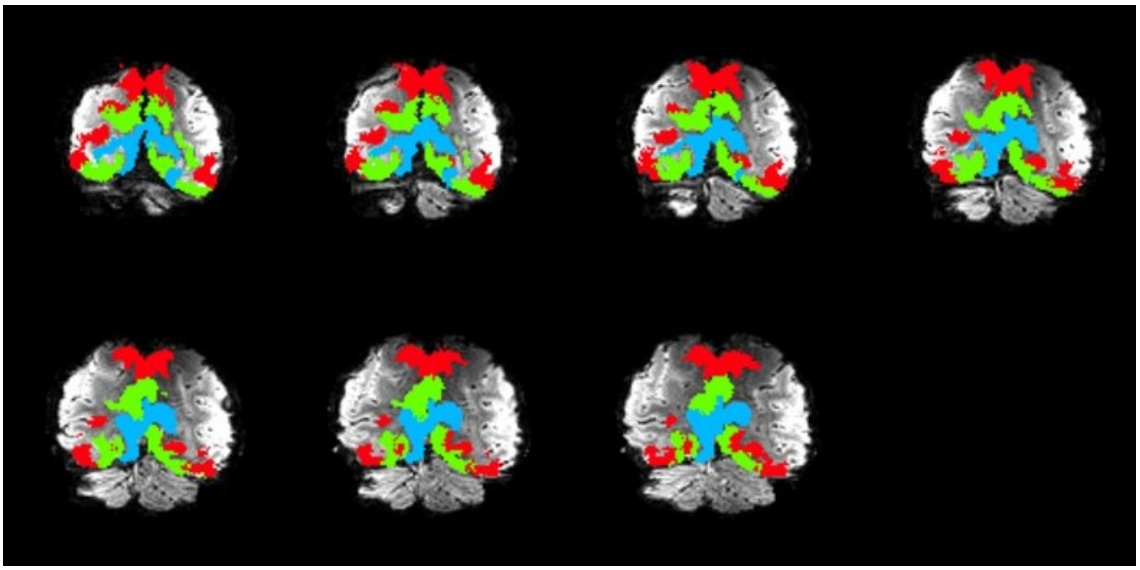

Figure 2: Images of the brain of one of the participants with the early visual areas V1 (blue), V2 (green) and V3 (red), posterior to anterior (from left to right) overlaid to show the structure of the three visual areas that are of interest in the analyses.

Results

To put the analyses into perspective, figure 2 shows a part of the brain of one of the participants with the regions of interest overlaid. It shows seven slices towards the back of the brain which contains the early visual cortex, which visualizes the region in which the analysed voxels are located. Figure 3 shows the beta-coefficients to the task and its progression through the conditions, visualizing differences between the methods and the baseline in the temporal scans. In figure 4 the mean beta-coefficients for the spatial pattern runs can be seen. Table 1 shows the outcomes of the linear mixed-effects model for the temporal pattern and spatial pattern runs respectively for each method compared to the baseline (i.e. undenoised preprocessed data). In these analyses, each method used the voxels that were selected using the HRF-scan, resulting in a group of 2061 voxels per scan (356, 978 and 727 from the participants respectively) where the temporal pattern experiment consisted of four scans and the spatial pattern experiment two scans.

GLM-mp-ph

The GLM using motion- and physiological parameters showed a significant decrease in tSNR compared to the baseline for the temporal pattern condition ($t = -9.46$, $p < .001$), whilst the SD per voxel also decreased significantly ($t = -9.039$, $p < .001$). A decrease in SD per voxel should result in an increase in tSNR unless this is counteracted by a decrease in signal as well, suggesting an increased negative BOLD signal whilst denoising with GLM-mp-ph. This effect however cannot be corroborated by task-related testing as neither β-slopes nor the mean t-values for the task showed significant change from the baseline. As for the spatial scans, an increased tSNR was found ($t = 14.470$, $p < .001$), which can be partly attributed to a decrease in SD per voxel ($t = -6.711$, $p < .001$) but implies an increase in BOLD-signal as well. However, for the spatial scans no trend of significant differences was found in task-signal with only the t-values for the grating-condition (cond. 6) significantly increasing ($t = 4.297$, $p < .001$). So, no substantial significant differences were found in task-signal compared to the baseline even though an increased (both positive and negative) BOLD signal and lowered SD per voxel were found, implying that GLM-mp-ph had an effect on the data but wasn't able to separate noise from signal of interest in both the temporal and the spatial scans.

GLMDenoise

Running GLMDenoise on both the temporal and spatial dataset and comparing these to the baseline revealed a significant decrease in SD per

Methods	tSNR		SD		β-slope (cond 1:6)		β-slope (cond 7:12)		Task t-values	
	t	Sig.	t	Sig.	t	Sig.	t	Sig.	t	Sig.
GLM - mp & ph	-9.460	<.001*	-9.039	<.001*	-1.127	.260	-.310	.757	-.335	.738
GLMDenoise	1.546	.122	-24.233	<.001*	-13.317	<.001*	-8.800	<.001*	-12.903	<.001*
GLM - ICA	-.046	.963	-7.446	<.001*	.533	.594	-.429	.668	-.899	.369
FSL Fix	.056	.955	7.279	<.001*	.233	.816	-.662	.508	-6.810	<.001*

Methods	tSNR		SD		β-slope (cond. 1:5)		β-slope (cond. 9:12)		Task t (cond. 6)		Task t (cond. 7)		Task t (cond. 8)	
	t	Sig.	t	Sig.	t	Sig.	t	Sig.	t	Sig.	t	Sig.	t	Sig.
GLM - mp & ph	14.470	<.001*	-6.711	<.001*	1.951	.051	.054	.957	4.297	<0.01*	-1.100	.271	1.791	.073
GLMDenoise	1.237	.216	-17.080	<.001*	-3.326	0.001*	3.759	<.001*	2.107	.035	-6.391	<.001*	-6.144	<.001*
GLM - ICA	.066	.948	-5.515	<.001*	-.371	.711	-1.023	.306	-3.648	<0.01*	5.444	<.001*	5.039	<.001*
FSL Fix	.040	.968	1.782	.075	-3.730	0.001*	4.840	<.001*	-1.835	.066	-6.111	<.001*	-4.671	<.001*

Table 1: Linear mixed model output tables showing the results of the fixed effect of each method for the temporal pattern (top) and spatial pattern (bottom). Each method was compared to the baseline (undenoised preprocessed data).

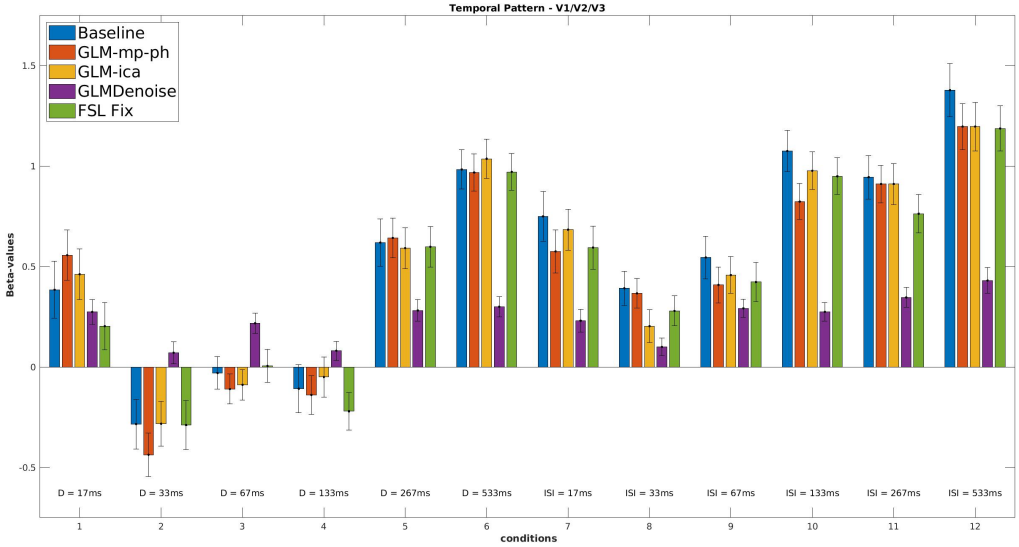

Figure 3: Depiction of the mean β values (y-axis) for every condition in the temporal pattern scan for all participants (x-axis) for every method. Error bars contain the standard error of the mean across voxels.

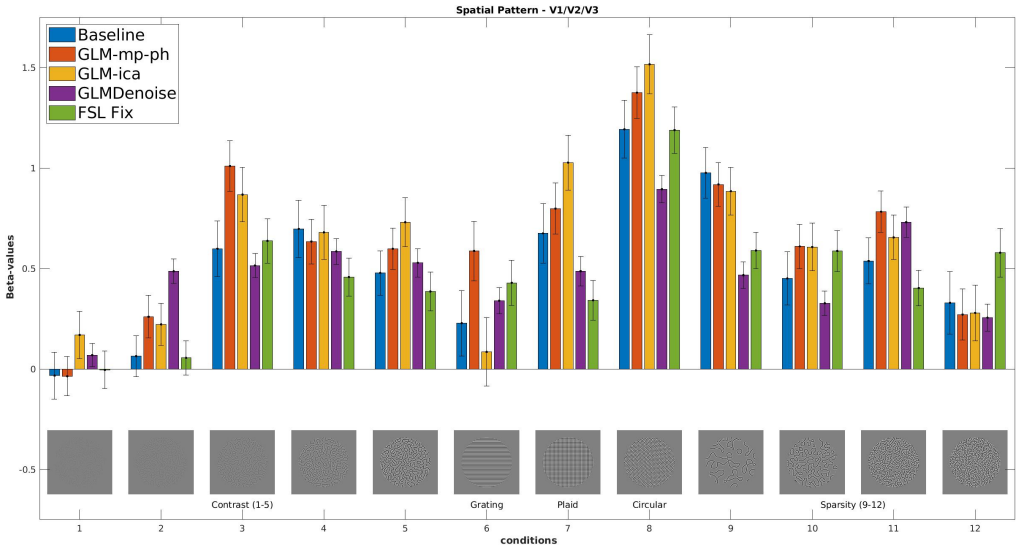

Figure 4: Depiction of the mean β values (y-axis) for every condition in the spatial pattern scan for all participants (x-axis) for every method. Error bars contain the standard error of the mean across voxels. The images depicted along the x-axis are example stimuli for each condition.

voxel (t = -24.233, p < .001) whilst no significant differences were found in tSNR (p = .122 and p = .216 respectively), which indicates a decrease in mean signal alongside the decrease in SD. This decrease in signal unfortunately also showed in the task-related values for the temporal scans, with a negative significant effect found in both β-slopes and the mean t-values for task denoting a loss of signal of interest and implying GLMDenoise filtered noise and signal of interest coincidentally. For the spatial scans, GLMDenoise had a negative significant effect on the β-slope of the contrast conditions (t = -3.326, p < .001) whilst having a positive significant effect on the β-slope of the sparsity conditions (t = 3.759, p < .001), which is contradictory.

GLM-ICA

Comparing the GLM with hand-picked independent components to the baseline dataset, the temporal scans showed no significant effect in terms of tSNR (t = -0.46, p = .963) whilst SD per voxel was reduced significantly (t = -7.446, p < .001). This is similar to the findings in GLMDenoise, which suggest a decrease in mean signal. No significant differences were found however in task-related values. The spatial pattern data showed a similar pattern for this method, with tSNR not substantially different from the baseline (t = -.066, p = 9.48) but a significant difference in SD (t = -5.515, p < .001). Significant effects were found in the task-related t-values for the grating (t = -3.648, p < .01), plaid (t = 5.444, p < .001) and circular conditions (t = 5.039, p < .001). This suggests a small improvement in task-signal for the pattern shape conditions whilst no task-signal differences were found in the contrast and sparsity conditions.

FSL Fix

The last comparison was made between the baseline and FSL Fix, which showed a significant positive effect on SD per voxel (t = 7.279, p < .001), meaning an increase in standard deviation per voxel after denoising whilst no significant improvement in tSNR was found. No significant effects were found for the β-slopes of both conditions, whilst a negative significant effect was found for the mean task t-values over all conditions (t = -6.810, p < .001). For the spatial scans, no significant effects were found in tSNR or SD for FSL Fix relative to the baseline, the task-related test however showed significance in both the β-slope of the contrast conditions (t = -3.730, p < .001) and sparsity conditions (t = 4.840, p < .001). Also, significant negative effects were found in task related t-values for the plaid condition (t = -6.111, p < .001) and circular condition (t = -4.671, p < .001). So, for the spatial dataset, only the sparsity conditions was positively impacted by FSL Fix denoising, whilst the other conditions are impacted negatively.

Discussion

This research provides an evaluation of methods for denoising fMRI data and as the results show, all methods perform inconsistently at best. The GLM with motion and physiological parameters incorporated had an effect on tSNR in both datasets. Furthermore, increased negative BOLD-signal was found in the temporal data and increased positive BOLD-signal in the spatial data. An increase in signal, albeit a negative BOLD-signal, could indicate that a method has merit. However, these changes were not found when testing for task-related signal as β-slopes and task-related t-values after denoising with the GLM-mp-ph did not significantly differ from the β-slopes and t-values found in the baseline. If this method was effective in its purpose of denoising, a change in task-related signal would have been observed in the comparison with the baseline. Therefore, it can be concluded that GLM-mp-ph had an effect on the data but wasn't able to separate noise from signal of interest.

GLMDenoise showed no differences to the

baseline when comparing tSNR, but did have a significant decrease in standard deviation per voxel. The method impacted the data in β-slopes as well, significantly reducing these. Figure 3 also shows the reduction in beta-coefficients in relation to the baseline. The decline of beta-coefficients does not necessarily mean unsuccessful denoising, as the coefficients might have been boosted by noise factors. However, GLMDenoise significantly decreased the task-related t-values as and the overall tendency of both BOLD-signal as well as signal of interest declining gives no indication of successful denoising. This might be due to the circular analysis which occurs during voxel selection for the PCA. When a clear distinction cannot be made between voxels that contain signal of interest and voxels that do not, the chance of this method filtering out signal of interest together with noise is high. GLMDenoise might have had this problem since it is congruent with results of the analysis i.e. both signal and signal of interest decreasing.

the GLM ICA was unable to significantly impact the data in terms of beta-slope or t-values. It seemed to have a very small effect on the data overall, suggesting that the ICs that were deemed to be noise and therefore used in the nuisance regression mostly contained noise that did not affect the early visual cortex, or at least not the voxels that were used in the analysis. A contributor to the lack of effect can be that hand-picking of components is a subjective process that is very dependent on a person's expertise and difficult to reproduce (Kelly et al., 2010). In this case, the researcher' expertise on ICAs and fMRI data might be called into question, being a student with only basic knowledge of the field. However, FSL Fix which uses a combination of classifiers to determine which components should be labelled noisy did not considerably outperform the GLM-ICA. FSL Fix showed similar results as the GLM-ICA method on the temporal data, except for the task-related t-values significantly declining compared to the baseline. In the spatial pattern data, FSL Fix did not improve the data in terms of tSNR or standard deviation per voxel, it did however alter the β-slopes significantly, with the contrast pattern increasing and the sparsity pattern decreasing contradicting each other. This inconsistency might be due to noise-boosted beta-coefficients in the sparsity conditions, however FSL Fix showed very low overall effect on denoising the data, and performed similarly as the 'low-tech' hand-picked IC method making the implementation of FSL Fix questionable when weighing the costs both computationally and time-wise in terms of creating a training-set. Compared to the time-consumption of FSL FIX, GLMDenoise is rather fast and needs only the data and task designs to work. A downside to GLMDenoise is that it can't work with high-pass filtered data, and therefore detrending needs to be done after denoising. Hence, GLMDenoise can be better described as a tool for pre-processing which can be quite easily implemented, which might be a useful tool for quick denoising in some but, given the results on this dataset, definitely not all datasets. All in all, no denoising method seemed to have a clear and positive effect in both the temporal and the spatial data and it seems that no method was able to make a clear distinction between noise and signal of interest which is problematic.

The downside of BOLD fMRI is that the response to a signal induced by neuronal activity is only small and therefore the total signal is usually very noisy (Bianciardi et al., 2009). However, the upside of this is that in theory a method which denoises the data even a little is quite effective since there is a lot to gain. In this research, however, none of the four different methods seemed to be very successful in their task. This raises some questions about the designs and execution of the experiments and whether the experimental set-up could be part of the reason why the data does not respond well to all denoising methods that were tried. For instance, the experiments were designed with conditions that should elicit a response that would gradually grow, this however

was not consistently found within and across subjects with some conditions even gradually declining or showing a negative overall BOLD-response. A shortcoming of the design might be that tasks had too many conditions. This meant there were only few repetitions per condition and small temporal spacing between each event. Within both the temporal and spatial designs, there were 12 conditions which each had three trials, which was all done within a timeframe of $\sim$ 182 seconds per scan (214 frames with a TR = 850ms). Evidence that suggests this could be the case is the substantially higher beta- and t-values in the HRF design which only contained one condition. Temporal spacing in the HRF design was similar to that of both the temporal and the spatial designs, but the HRF experiment contained many more repetitions. This is a trade-off between the amount of repetitions and the amount of conditions in which the temporal and the spatial designs inadvertently need more than one condition, yet less than 12 might be advisable.

This research was initially started with the intention of designing a machine learning algorithm for denoising data using motion and physiological parameters as a base, in order to explain empirically observed non-linearities in BOLD responses which cannot be caught using a GLM (Birn et al., 2001). Accordingly, non-linearity of the data is a possible reason why the proposed GLMs in this research have not shown sizeable or consistent results. However, the machine-learning approach did not yield interesting results with confounds being the amount of data and expertise within the field of fMRI. For denoising to be done successfully, it is important that a model is able to capture the complex and mingled noise factors which are often present in fMRI BOLD datasets. Capturing complex patterns in data is something that machine learning is particularly apt at and fuelled by increasing computer power and algorithmic advances, it has become a powerful tool for finding patterns in data (Biamonte et al., 2017). So machine learning inspired denoising tools might still offer the solution in the future.

Conclusion

This research has shown that implementing methods which have been validated on other datasets does not necessarily guarantee a good result. No conclusive evidence for any method outperforming the baseline or the other methods can be reported. Therefore, the methods are ineffective in denoising the current dataset. The basis for these results might however be methodological, having to do with data acquisition rather than denoising, and a critical look at the design of the experiment needs taken as no substantial claims about any denoising method can be done. An alternative contribution to the lack of denoising results could be non-linearities in the data which none of the proposed methods are able to catch. Answering the secondary research question, a denoising method which is easy to implement and might sometimes perform well but at least does not negatively impact the data might still be useful. With that in mind, only GLMDenoise might be used in terms of computational efficiency and a GLM using motion and physiological parameters might perform reasonably in some cases. Using ICA-methods is however both time consuming and did not yield promising results. Ultimately, denoising fMRI data has proven to be a difficult problem altogether which is in need of a method that can capture noise consistently.

References

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ants similarity metric performance in brain image registration. *NeuroImage*, *54*(3), 2033–2044. doi: 10.1016/j.neuroimage.2010.09.025

Bandettini, P. A., Jesmanowicz, A., Wong, E. C., & Hyde, J. S. (1993). Processing strategies for time-course data sets in functional mri of the human brain. *Magnetic Resonance in Medicine*, *30*(2), 161–173. doi: 10.1002/mrm.1910300204

Bartha, R., Michaeli, S., Merkle, H., Adriany, G., Andersen, P., Chen, W., ... Garwood, M. (2002). In vivo1h2ot?2 measurement in the human occipital lobe at 4t and 7t by carr-purcell mri: Detection of microscopic susceptibility contrast. *Magnetic Resonance in Medicine*, *47*(4), 742–750. doi: 10.1002/mrm.10112

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. doi: 10.1016/j.neuroimage.2007.04.042

Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., & Lloyd, S. (2017). Quantum machine learning. *Nature*, *549*(7671), 195–202. doi: 10.1038/nature23474

Bianciardi, M., Fukunaga, M., Gelderen, P. V., Horovitz, S. G., Zwart, J. A. D., 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. doi: 10.1016/j.mri.2009.02.004

Birn, R. M., Saad, Z. S., & Bandettini, P. A. (2001). Spatial heterogeneity of the nonlinear dynamics in the fmri bold response. *NeuroImage*, *14*(4), 817–826. doi: 10.1006/nimg.2001.0873

Birn, R. M., Smith, M. A., Jones, T. B., & Bandettini, P. A. (2008). The respiration response function: The temporal dynamics of fmri signal fluctuations related to changes in respiration. *NeuroImage*, *40*(2), 644–654. doi: 10.1016/j.neuroimage.2007.11.059

Caballero-Gaudes, C., & Reynolds, R. C. (2017). Methods for cleaning the bold fmri signal. *NeuroImage*, *154*, 128–149. doi: 10.1016/j.neuroimage.2016.12.018

Chang, C., Cunningham, J. P., & Glover, G. H. (2009). Influence of heart rate on the bold signal: The cardiac response function. *NeuroImage*, *44*(3), 857–869. doi: 10.1016/j.neuroimage.2008.09.029

Chang, C., & Glover, G. H. (2009). Effects of model-based physiological noise correction on default mode network anti-correlations and correlations. *NeuroImage*, *47*(4), 1448–1459. doi: 10.1016/j.neuroimage.2009.05.012

Cox, R. W. (1996). Afni: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173. doi: 10.1006/cbmr.1996.0014

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fmri data. *NMR in Biomedicine*, *10*(4-5), 171–178. doi: 10.1002/(sici)1099-1492(199706/08)10:4/5<171::aid-nbm453>3.0.co;2-l

Friston, K. J., Jezzard, P., & Turner, R. (1994). Analysis of functional mri time-series. *Human Brain Mapping*, *1*(2), 153–171. doi: 10.1002/hbm.460010207

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, *15*(4), 870–878. doi: 10.1006/nimg.2001.1037

Griffanti, L., Salimi-Khorshidi, G., Beckmann, C. F., Auerbach, E. J., Douaud, G., Sexton, C. E., ... et al. (2014). Ica-based artefact removal and accelerated fmri acquisition for improved resting state network imaging. *NeuroImage*, *95*, 232–247. doi: 10.1016/j.neuroimage.2014.03.034

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience*, *3*. doi: 10.3389/neuro.09.031.2009

Hillman, E. M. (2014, Aug). Coupling mechanism and significance of the bold signal: A status report. *Annual Review of Neuroscience*, *37*(1), 161–181. doi: 10.1146/annurev-neuro-071013-014111

Hutton, C., Josephs, O., Stadler, J., Featherstone, E., Reid, A., Speck, O., ... Weiskopf, N. (2011). The impact of physiological noise correction on fmri at 7 t. *NeuroImage*, *57*(1), 101–112. doi: 10.1016/j.neuroimage.2011.04.018

Kalthoff, D., Seehafer, J. U., Po, C., Wiedermann, D., & Hoehn, M. (2011). Functional connectivity in the rat at 11.7t: Impact of physiological noise in resting state fmri. *NeuroImage*, *54*(4), 2828–2839. doi: 10.1016/j.neuroimage.2010.10.053

Kay, K. N., Rokem, A., Winawer, J., Dougherty, R. F., & Wandell, B. A. (2013). Glmdenoise: a fast, automated technique for denoising task-based fmri data. *Frontiers in Neuroscience*, *7*. doi: 10.3389/fnins.2013.00247

Kelly, R. E., Alexopoulos, G. S., Wang, Z., Gunning, F. M., Murphy, C. F., Morimoto, S. S., ... et al. (2010). Visual inspection of independent components: Defining a procedure for artifact removal from fmri data. *Journal of Neuroscience Methods*, *189*(2), 233–245. doi: 10.1016/j.jneumeth.2010.03.028

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. doi: 10.1038/nn.2303

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

Lee, A. T., Glover, G. H., & Meyer, C. (1995). Discrimination of large venous vessels in time-course spiral blood-oxygen-level-dependent magnetic-resonance functional neuroimaging. *Magnetic Resonance in Medicine*, *33*(6), 745-754.

Lindquist, M. A., Loh, J. M., Atlas, L. Y., & Wager, T. D. (2009). Modeling the hemodynamic response function in fmri: Efficiency, bias and mis-modeling. *NeuroImage*, *45*(1). doi: 10.1016/j.neuroimage.2008.10.065

Liu, T. T. (2016). Noise contributions to the fmri signal: An overview. *NeuroImage*, *143*, 141–151. doi: 10.1016/j.neuroimage.2016.09.008

Logothetis, N. K. (2008). What we can do and what we cannot do with fmri. *Nature*, *453*(7197), 869–878. doi: 10.1038/nature06976

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the bold signal. *Annual Review of Physiology*, *66*(1), 735–769. doi: 10.1146/annurev.physiol.66.082602.092845

Mckeown, M. (2003). Independent component analysis of functional mri: what is signal and what is noise? *Current Opinion in Neurobiology*, *13*(5), 620–629. doi: 10.1016/j.conb.2003.09.012

Murphy, K., Birn, R. M., & Bandettini, P. A. (2013). Resting-state fmri confounds and cleanup. *NeuroImage*, *80*, 349–359. doi: 10.1016/j.neuroimage.2013.04.001

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

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. doi: 10.1016/j.neuroimage.2013.11.046

Siero, J. C. W., Ramsey, N. F., Hoogduin, H., Klomp, D. W. J., Luijten, P. R., & Petridou, N. (2013). Bold specificity and dynamics evaluated in humans at 7 t: Comparing gradient-echo and spin-echo hemodynamic responses. *PLoS ONE*, *8*(1). doi: 10.1371/journal.pone.0054560

Smith, A. M., Lewis, B. K., Ruttimann, U. E., Ye, F. Q., Sinnwell, T. M., Yang, Y., ... Frank, J. A. (1999). Investigation of low frequency drift in fmri signal. *NeuroImage*, *9*(5), 526–533. doi: 10.1006/nimg.1999.0435

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., ... et al. (2004). Advances in functional and structural mr image analysis and implementation as fsl. *NeuroImage*, *23*. doi: 10.1016/j.neuroimage.2004.07.051

Soltysik, D. A., Thomasson, D., Rajan, S., & Biassou, N. (2015). Improving the use of principal component analysis to reduce physiological noise and motion artifacts to increase the sensitivity of task-based fmri. *Journal of Neuroscience Methods*, *241*, 18–29. doi: 10.1016/j.jneumeth.2014.11.015

Triantafyllou, C., Hoge, R., Krueger, G., Wiggins, C., Potthast, A., Wiggins, G., & Wald, L. (2005). Comparison of physiological noise at 1.5 t, 3 t and 7 t and optimization of fmri acquisition parameters. *NeuroImage*, *26*(1), 243–250. doi: 10.1016/j.neuroimage.2005.01.007

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241–259. doi: 10.1016/s0893-6080(05)80023-1

Zwart, J. A. D., Gelderen, P. V., Fukunaga, M., & Duyn, J. H. (2008). Reducing correlated noise in fmri data. *Magnetic Resonance in Medicine*, *59*(4), 939–945. doi: 10.1002/mrm.21507

Zwart, J. A. D., Ledden, P. J., Gelderen, P. V., Bodurka, J., Chu, R., & Duyn, J. H. (2003). Signal-to-noise ratio and parallel imaging performance of a 16-channel receive-only brain coil array at 3.0 tesla. *Magnetic Resonance in Medicine*, *51*(1), 22–26. doi: 10.1002/mrm.10678

Appendix

Basics of MRI: T1- and T2-weighted imaging

Magnetic resonance imaging (MRI) is a technique which uses the magnetic properties of protons within molecules in order to make detailed pictures of the anatomy and the physiological processes of the body. These protons normally have a random orientation, but a powerful magnetic field can align them. While aligned, applying a radiofrequency (RF) pulse will excite the protons which causes them to change angle of orientation and makes them spin in 'phase' with each other (i.e. resonance). After the RF pulse, the protons start to get back to their original state within the magnetic field by (1) *realignment* to the magnetic field and (2) *dephasing* of spinning protons (i.e. loss of resonance). Realignment and dephasing to the original state can be detected and result in two types of signal: *T1 signal* relating to the time it takes for protons to realign with the magnetic field and *T2 signal* which is calculated using the speed of dephasing. Both realignment and dephasing vary in speed depending on which tissue type the protons are located in. Within fat tissue protons realign quickly resulting in a high T1 signal whilst protons in water (or fluid) have a slow dephasing speed and give a high T2 signal. Using RF pulse sequences which exploit these phenomenon T1- and T2-weighted images can be formed which highlight only fat tissue (T1) or fat tissue and fluids (T2). These images can show different textures in the brain and distinguished e.g. Cerebrospinal fluid (CSF) which will show up dark on T1-weighted images and white on the T2-weighted images since it is a fluid and contains no fat. Figure 5 shows the differences between a T1-weighed and a T2-weighted scan, where CSF can be distinguished in the middle of the brain as it is dark on the T1 image but very bright on the T2 image.

fMRI: T2 imaging and BOLD-contrast*

In order to be able to generate functional MRI images, T2* images need to be created which highlight the presence of blood products. Figure 6 shows the difference between T2 and T2* images, in which the T2* image of a part of the brain clearly highlights a spot which is dark and messy on the T2 image. This is a spot on the brain which contains a lot of blood called a hemangioma i.e. a collection of small blood vessels forming a lump. Blood-Oxygen Dependent (BOLD) contrast can be used on these T2* images, showing localised oxygen-richness in the brain over time. BOLD-contrast measures changes in the ratio of paramagnetic deoxyhemoglobin to diamagnetic oxyhemoglobin, taking advantage of the difference in magnetic properties between oxygen-rich and oxygen-poor blood. As the proportion of deoxyhemoglobin decreases, the MR signal increases. Several reasons exist for the decrease of proportion deoxyhemoglobin (and thus a BOLD signal increase), one of which is increased blood flow due to the heightened energy requirements of active tissue: more oxygen-rich blood can be found in places that need the oxygen to produce energy after (or during still ongoing) activation of neurons in that area. The signal that can be contributed to this mechanism is called signal of interest, which forms the basis for BOLD fMRI giving insight into brain activity. However, only a small part of up to ten percent of BOLD signal fluctuations can be attributed to increase in energy requirements, which means there is a lot of 'unwanted' signal i.e. noise stemming from physiological factors. An example of a physiological factor that has effect on the BOLD signal is heart rate variability which both causes fluctuations in the oxygen-richness of blood in the brain as well as changes in the sensitive magnetic field due to the slight movement of the body as the heart pounds. Substantial processing

and analysis is needed to correct for different unwanted but inevitable influences to the inherently noise-prone BOLD-signal.

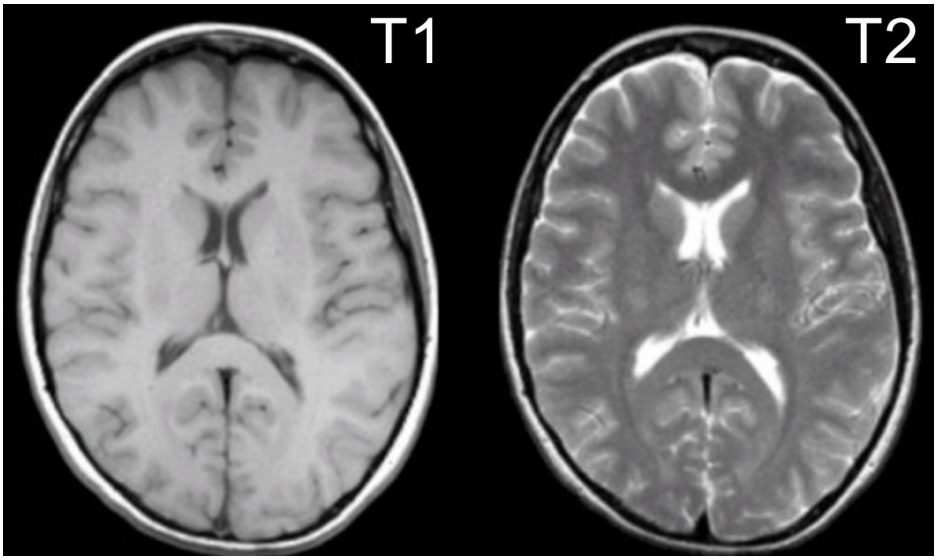

Figure 5: T1- and a T2-weighted image of a part of the brain, showing the contrast between the different imaging techniques.

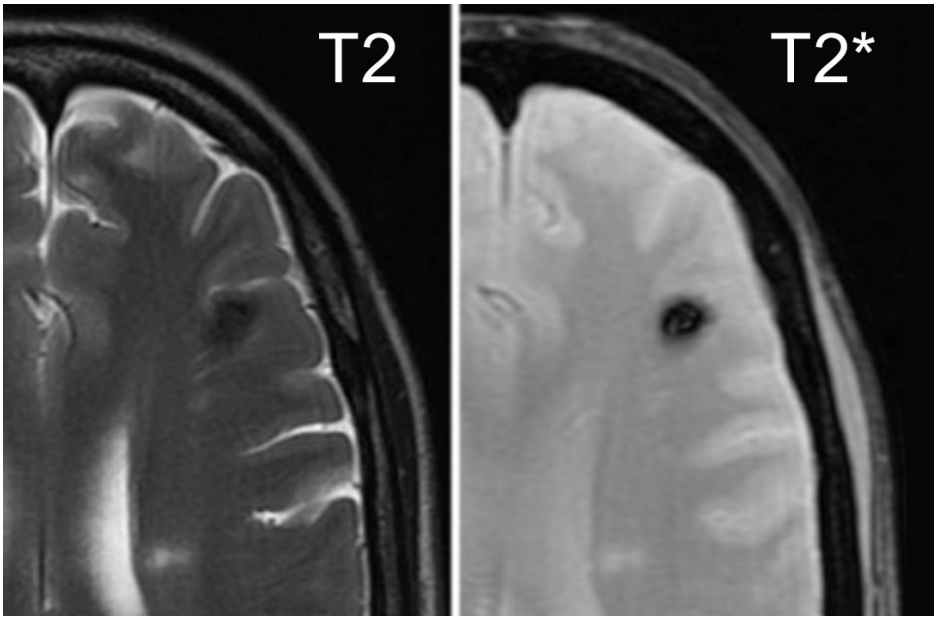

Figure 6: T2- and a T2*-weighted images one part of the brain, showing a hemangioma.

Supplementary Material

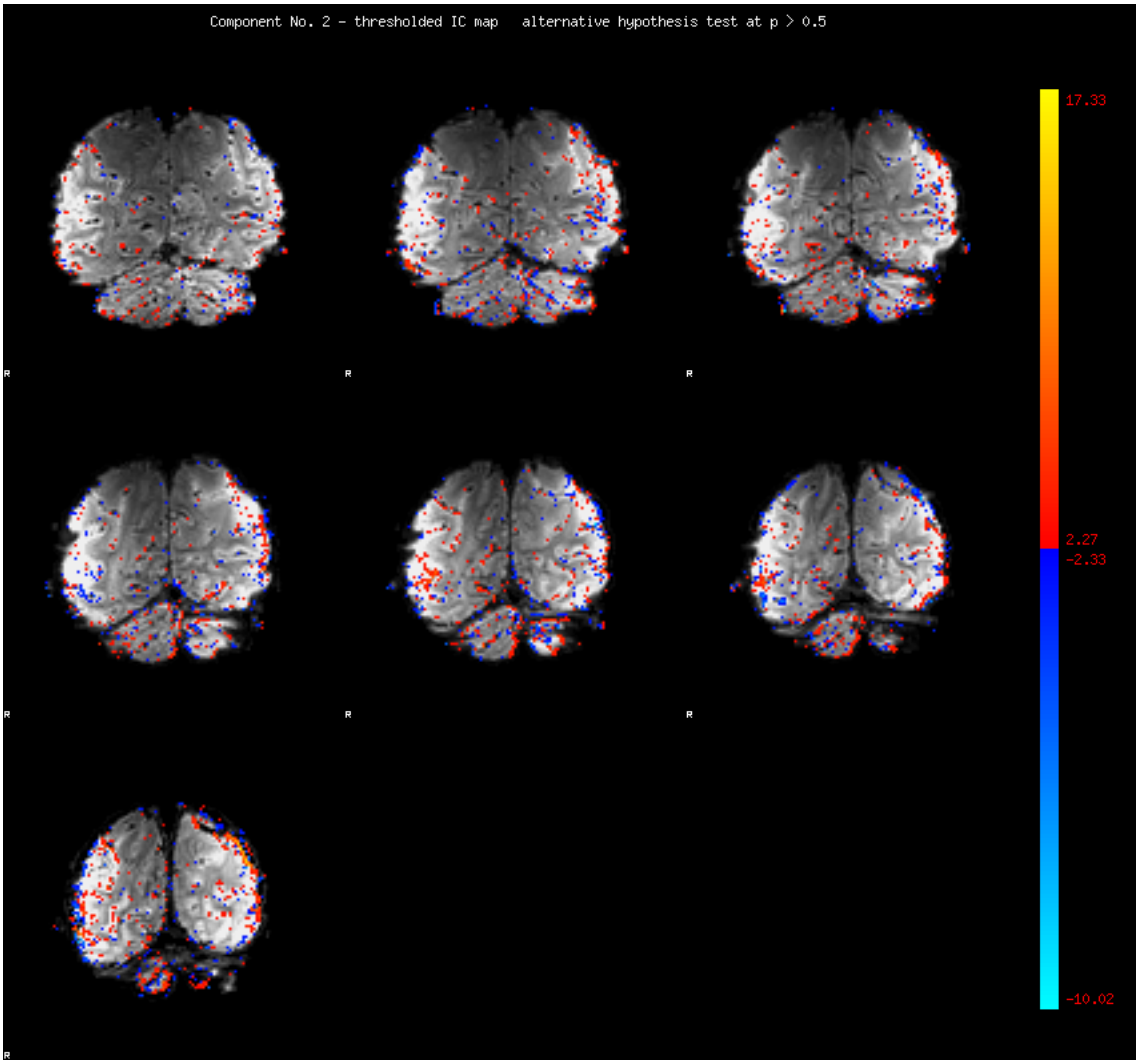

Figure 7: *Participant 11, Component Nr.2 of 51.* An example of what an independent component looks like after a threshold function has been applied, showing which voxels correlate highly (positively or negatively) with each other. Presented anterior to posterior (7 slices).