

## FIACH: A biophysical model for automatic retrospective noise control in fMRI



Tim M. Tierney<sup>a,\*</sup>, Louise J. Weiss-Croft<sup>c</sup>, Maria Centeno<sup>a</sup>, Elhum A. Shamshiri<sup>a</sup>, Suejen Perani<sup>a,b</sup>, Torsten Baldeweg<sup>c</sup>, Christopher A. Clark<sup>a</sup>, David W. Carmichael<sup>a</sup>

<sup>a</sup> Developmental Imaging and Biophysics, UCL Institute of Child Health, University College London, UK

<sup>b</sup> Department of Basic and Clinical Neuroscience, KCL Institute of Psychiatry, Psychology & Neuroscience, London, UK

<sup>c</sup> Cognitive Neuroscience and Neuropsychiatry, UCL Institute of Child Health, University College London, London, UK

### ARTICLE INFO

#### Article history:

Received 17 April 2015

Accepted 14 September 2015

Available online 28 September 2015

Editor: Steven Smith

#### Keywords:

Automatic  
Retrospective  
Noise  
fMRI  
Children

### ABSTRACT

Different noise sources in fMRI acquisition can lead to spurious false positives and reduced sensitivity. We have developed a biophysically-based model (named FIACH: Functional Image Artefact Correction Heuristic) which extends current retrospective noise control methods in fMRI. FIACH can be applied to both General Linear Model (GLM) and resting state functional connectivity MRI (rs-fcMRI) studies. FIACH is a two-step procedure involving the identification and correction of non-physiological large amplitude temporal signal changes and spatial regions of high temporal instability. We have demonstrated its efficacy in a sample of 42 healthy children while performing language tasks that include overt speech with known activations. We demonstrate large improvements in sensitivity when FIACH is compared with current methods of retrospective correction. FIACH reduces the confounding effects of noise and increases the study's power by explaining significant variance that is not contained within the commonly used motion parameters. The method is particularly useful in detecting activations in inferior temporal regions which have proven problematic for fMRI. We have shown greater reproducibility and robustness of fMRI responses using FIACH in the context of task induced motion. In a clinical setting this will translate to increasing the reliability and sensitivity of fMRI used for the identification of language lateralisation and eloquent cortex. FIACH can benefit studies of cognitive development in young children, patient populations and older adults.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

### Introduction

#### The Problem of noise in fMRI

The question of how to separate signal from noise in fMRI is one of crucial importance. This is primarily due to the fact fMRI signals are weak in magnitude (a few percent) and are therefore easily concealed by increased noise levels. Whether the noise is due to subject motion, physiological processes or scanner hardware it all can lead to spurious results (Murphy et al., 2013). Specifically with regards to subject motion error variance is increased resulting in reduced sensitivity (Friston et al., 1996) and false positives (see Fig. 1). In rs-fcMRI has been demonstrated that noise due to head motion can result in a spurious increase in local connectivity (Van Dijk et al., 2011) and that many of these issues are not remedied fully by regression of the realignment parameters (RPs) from the signal (Power et al., 2011).

#### Motion

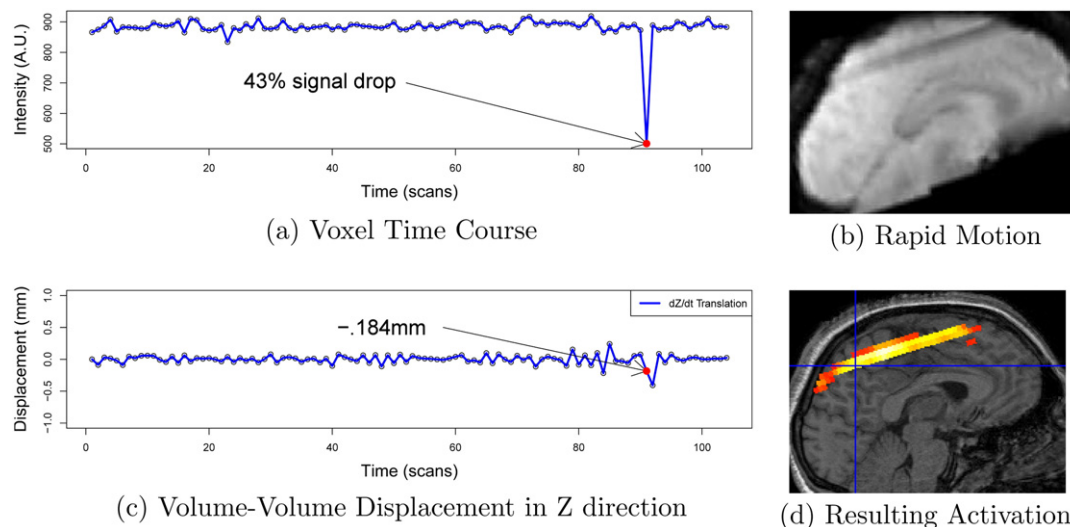
To account for the effects of motion induced noise a variety of retrospective methods have been proposed. Many motion-focused correction methods rely on modelling some transformation of the parameters attained from volumetric spatial realignment or from some measure of signal intensity across the whole brain such as DVARS (Derivative of VARIation) or mean global signal (Friston et al., 1996; Smyser et al., 2010; Desjardins et al., 2001; Lemieux et al., 2007; Wilke, 2012).

The use of methods incorporating global signal have been debated in the literature ever since it was demonstrated that regressing out global signal can result in spurious negative correlations (Murphy et al., 2009). Furthermore, if an artefact's structure is spatially specific (as is the case in Fig. 1b) it will only be partially reflected in the global signal. This will result in less efficient control of the artefact when it comes to the regression stage of the analysis.

This viewpoint is supported by Beall and Lowe (2014) who demonstrate that global signal is a poor reflection of motion related signal changes at each voxel. They conclude that it is not justifiable as a

\* Corresponding author at: 30 Guilford Street, UCL Institute of Child Health, University College London, London, UK.

E-mail address: [tim.tierney.12@ucl.ac.uk](mailto:tim.tierney.12@ucl.ac.uk) (T.M. Tierney).



**Fig. 1.** The effect of through-plane motion. Through-plane motion produces a spurious observation in the time series (a). This artefact presents spatially with a slice-specific profile (b). The effect on the time series is not reflected in the through plane derivative of the realignment parameter despite the fact that the effect of the motion is predominantly in this direction. The red sphere shows the time point where the motion occurred (c). The resulting statistical map shares the same spatial location as the artefact (d) due to motion-task correlation. The contrast is an F-contrast across the canonical haemodynamic response and its derivatives thresholded at FWE ( $p < .05$ ).

method of motion control. This is in contrast to Power et al. (2014) who demonstrate that global signal reduces motion related signal across voxels.

#### Localised large amplitude signal changes

While correction methods that are based on the volumetric RPs are useful (Lund et al., 2005) they can have limitations when used as subject exclusion criteria or metrics of data quality. This is because they do not model motion but, net displacement relative to a reference sampled every TR. This is particularly problematic in paediatric or patient populations where head motion may be faster than the volumetric sampling rate (TR). We present an illustrative example of this point where the volumetric RP is a poor reflection of subject motion in a 2D gradient echo sequence with axial acquisition (Fig. 1).

By not fully modelling the noise effects in the data, a false positive has been obtained in Fig. 1d. This false positive was obtained even after using scan nulling (1 mm framewise displacement threshold, similar to optimum threshold found by Siegel et al., 2014) and the autoregressive polynomial expansion of the RPs (Friston et al., 1996; Lemieux et al., 2007).

Recent work by Beall and Lowe (2014) would suggest that this is an unsurprising finding as they describe the frequently used volumetric measures of motion as not being able to “robustly identify motion corrupted events, especially in the most realistic scenario of sudden head movement”. Furthermore this type of signal change (Fig. 1b) is not uncommon and has been described by Satterthwaite et al. (2013) as the “predominant effect of motion” in a sample ranging from 8 to 23 years old.

Large amplitude signal changes can also be observed due to k-space spikes and are localised to specific slices. Volumetric realignment parameters do not necessarily detect this type of noise. In recognition of potential artefacts from electronic noise as well as from rapid motion, software has been developed to address them such as the ArtRepair toolbox for SPM (<http://cibsr.stanford.edu/tools/human-brain-project/artrepair-software.html>) and 3dDespike by AFNI (<http://afni.nimh.nih.gov/afni/>).

#### Physiological noise

Methods that control for physiological noise have been developed. These include utilising external monitors of physiological processes,

modelling mean white matter and CSF signals, modelling principal components from white matter, grey matter and temporally unstable areas (Glover et al., 2000; Birn et al., 2006; Satterthwaite et al., 2013; Behzadi et al., 2007).

However, the efficacy of these techniques is debated. Power et al. (2014) have described the utility of using PCA to create nuisance regressors from white matter and CSF signals as not producing encouraging results. Whereas Muschelli et al. (2014) have described utilising the aCompCor approach described by Behzadi et al. (2007) as just as effective as full frame-censoring approach (Siegel et al., 2014). The efficacy of external monitors is also debated as it has been suggested that data-driven techniques such as tCompCor, aCompCor and multi-echo ICA denoising provide comparable or more efficient noise reduction without external monitoring (Behzadi et al., 2007; Kundu et al., 2012).

It must be noted that some of these techniques do not purely model physiological noise. For instance modelling signals from white matter, grey matter and temporally unstable areas may also contain information concerning motion. Furthermore techniques such as ICA denoising and Robust Weighted Least Squares (RWLS) may model both physiological noise and motion artefacts (Beckmann & Smith, 2004; Salimi-Khorshidi et al., 2014; Diedrichsen and Shadmehr, 2005).

There are also practical limitations to these methods as well. For instance external monitoring and multi-echo EPI are acquisition dependent and while promising cannot be applied retrospectively. As for ICA denoising the process is not easily automated as it requires a training dataset that may be difficult to create. The focus of this paper is both automatic and retrospective techniques and these methods shall not be discussed further.

#### Methodological aims and hypothesis

Considering there is wide agreement that modelling volumetric RPs is useful (Lund et al., 2005; Power et al., 2014; Satterthwaite et al., 2013) the question becomes how to extend this approach to address the issue of physiological noise and localised large amplitude signal changes (where there is less agreement). We propose to develop a biophysical framework that can be used for identifying and addressing both of these issues. This model will be applicable regardless of sequence, hardware, field strength and subject population. It is hypothesised that a method developed to these specifications will provide improvements in correction of data over existing available methods.

## Theory

### Identifying large amplitude signal changes

We define non-physiological large amplitude signal changes as those that display signal changes larger than the possible BOLD response. To estimate the signal we simulate gradient echo signal decay as follows:

$$S = S_{\max} e^{-TE R2^*} \quad (1)$$

where  $S$  = signal,  $S = 100$  (to describe signal changes in % change),  $TE$  = echo time and

$$R2^* = R2 + R2' \quad (2)$$

where  $R2$  = transverse relaxation rate of grey matter and  $R2'$  is the contribution to the relaxation rate from the local magnetic field inhomogeneities.  $R2$  is estimated using empirical results (Uludag et al., 2009):

$$R2 = 1.74B0 + 7.77 \quad (3)$$

where  $B0$  is the magnetic field strength in Tesla (T). The following theoretical model is adopted for estimating  $R2'$  (Yablonskiy and Haacke, 1994):

$$R2' = \varsigma \delta\omega \quad (4)$$

where  $\varsigma$  = blood volume fraction and  $\delta\omega$  is the characteristic frequency of the object causing the magnetic field inhomogeneities.  $\varsigma$  can be estimated using the following empirical equation presented by Grubb et al. (1974):

$$\varsigma = .8 CBF^{.38} / 100 \quad (5)$$

where  $CBF$  = cerebral blood flow ( $\text{ml } 100 \text{ g}^{-1}$ ). We can estimate  $\delta\omega$  according to Yablonskiy and Haacke (1994) by assuming the static dephasing regime and a random distribution of vessel orientation as:

$$\delta\omega = \gamma B0 \Delta\chi \text{Hct } 4\pi/3 (1-Y) \quad (6)$$

where  $\gamma$  = gyromagnetic ratio of the proton,  $\Delta\chi$  = the volume susceptibility difference between oxygenated and deoxygenated blood,  $\text{Hct}$  = haematocrit, and  $Y$  = fraction of oxygenated blood. To simulate activation we solve the following equation:

$$S = S_{\max} e^{-TE R2^*_{\text{Activation}}} - S_{\max} e^{-TE R2^*_{\text{Baseline}}} \quad (7)$$

The difference between  $R2^*_{\text{Activation}}$  and  $R2^*_{\text{Baseline}}$  is that we allow both  $Y$  and  $\varsigma$  to vary. To calculate the expected response magnitude we allow  $Y$  to vary from .6 to .9. The baseline of .6 is assumed so as to be similar to other studies employing this model (Yablonskiy and Haacke, 1994; Buxton, Wong & Frank, 1998; Winawer et al., 2010; Mildner et al., 2001). The second value of  $Y$  determines the threshold for the detection of non-physiological signal changes. We assume a value of .9 for the second value of  $Y$  as the typical range from rest to strong activation is .6–.8 (Obata et al., 2004). This was chosen to make the threshold robust in children and in clinical populations where there may be greater variation in response magnitudes and to be conservative in terms of how much data is corrected.

For calculating  $\varsigma$  we assume 100% change in  $CBF$  from a baseline of  $55 \text{ ml } 100 \text{ g}^{-1} \text{ min}^{-1}$ . The resulting values estimated for  $\varsigma$  (3.667–4.77%) are in good agreement with empirical values of vascular density which have been measured as having a mean value of ~3.5% and a maximum of ~4.5% (Lauwers et al., 2008). The remaining parameters assumed to calculate  $\delta\omega$  are as follows:  $\Delta\chi = 4\pi \times 1.8 \times 10^{-7}$  (Weisskoff and Kiihne, 1992),  $\gamma = 42:57 \text{ MHz}$ ,  $\text{Hct} = .4$ . Solving

Eq. (7) for a 1.5 T scanner acquiring data with a  $TE$  of 30 ms produces the following threshold of 4.9%. Fig. 2 illustrates this process.

### Identifying physiological noise

The model to identify large amplitude signal changes assumes a low blood volume fraction (typically < 5%). This assumption is violated in brain regions with large vascular density and is therefore not applicable in these areas. However, by using the same framework as was used to identify spurious signal changes we can identify these brain regions and model their effects. If we assume that  $\delta\omega$  stochastically fluctuates with time we can describe the variable by a distribution ( $D$ ) characterised by a centrality ( $C_D$ ) and variability ( $V_D$ ) parameter as follows:

$$\delta\omega(t) \sim D(C_D, V_D). \quad (8)$$

As  $R2' \propto \varsigma$  then  $R2'(t)$  will be time varying with a distribution  $\mathcal{R}$  that is a scaled version of  $D$ :

$$R2'(t) \sim \mathcal{R}(C_{\mathcal{R}}, V_{\mathcal{R}}) \quad (9)$$

where  $C_{\mathcal{R}} = \varsigma C_D$  and  $V_{\mathcal{R}} = \varsigma V_D$ . Rewriting Eq. (1) in terms of  $R2'(t)$  we find that:

$$S(t) = S_{\max} e^{-TE R2 - TE R2'(t)}. \quad (10)$$

To investigate the effect of  $R2'(t)$  on  $S(t)$  we take the partial derivative of Eq. (10) with respect to  $R2'$ . We assume  $R2$  to be constant.

$$\frac{\partial S(t)}{\partial R2'(t)} = -TE S_{\max} e^{-TE R2^*} \quad (11)$$

The partial derivative implies that for positive values of  $R2'$   $S(t)$  will be decreasing. Therefore the resulting voxel time-series will be characterised by a centrality parameter ( $C_S$ ) that decreases and a variability parameter ( $V_S$ ) that increases as  $R2'$  increases. Therefore, an area of higher blood volume could be characterised by signal with lower centrality and higher variability compared to areas with lower blood volume. By taking the ratio of these parameters  $C_S/V_S$  we create a parameter that is sensitive to blood volume and temporal stability.

### Theory implementation

#### Step 1: physiological noise control

Physiological noise is identified automatically using the expectation-maximisation (EM) algorithm to fit a mixture of gaussians on a measure

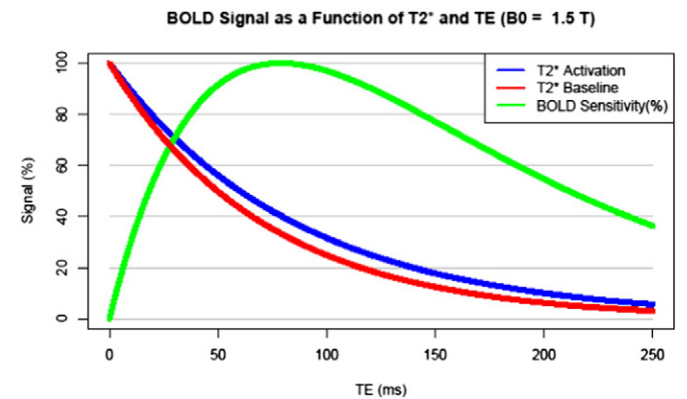


Fig. 2. BOLD contrast. The blue line represents the signal decay of the voxel during activation while the red line represents decay at rest. The green line shows one's sensitivity to the BOLD contrast. The BOLD sensitivity is given in percentage of the maximum value.

of robust temporal signal to noise ratio (rTSNR). We define rTSNR as the median/median absolute deviation of the realigned images (i.e. the  $C_S/V_S$  parameter discussed in section 'Identifying physiological noise') which have slow drifts removed by a high pass filter. The high pass filter consists of a basis set of cosines defaulting to remove frequencies with a period greater than 128 s. However this filter cut-off can be altered to better suit one's study design. Once the mixture model is estimated a mask is created from the voxels that have a  $C_S = V_S$  parameter less than the .05 quantile of the larger distribution. The first 6 principal components (PCs) are then extracted from within the mask. These components can then be included in the first level analysis as a parsimonious noise model. Fig. 3 illustrates this process. This step is similar to temporal component based noise correction (tCompCor) developed by (Behzadi et al., 2007). However, both the image used to identify the physiological noise and methods of segmentation differ.

#### Step 2: large amplitude signal change correction

To identify a time point in a voxel as artefactual we add a measure of noise ( $2 \times$  median absolute deviation) to the theoretically derived threshold for a biologically plausible signal change in the presence of noise. Signal changes relative to the previously calculated median exceeding this threshold are then interpolated using a natural cubic spline with the two time points on either side. If consecutive time points of the same voxel are corrupted the spurious change is replaced with the median to avoid interpolation errors. This method is implemented using the R statistical programming language (R Core team, 2013) and is freely available as an R package (FIACH). It is distributed via the Comprehensive R Archive Network (CRAN) and details for installation and use can be found at the first author's website (<http://www.homepages.ucl.ac.uk/~ucjtjie/index.html>). The following packages were used: Rcpp, RcppArmadillo and RNiftyReg (Eddelbuettel and Francois, 2011; Eddelbuettel and Sanderson, 2014; Clayden, 2015). This method can be applied to both GLM style analysis and rs-fMRI. Fig. 4 graphically explains the analysis pipeline. The only requirement is that the images are realigned prior to usage. Subsequently any fMRI analysis method can be applied with the addition of the derived regressors.

#### Theory summary

The theory proposed for artefact identification can be broken down into two sub-models: One that identifies instances of abrupt temporal instability and one that identifies consistent temporal instability. These models form a two-step procedure for retrospective artefact correction in fMRI. We will refer to this model from now on as the Functional Image Artefact Correction Heuristic (FIACH: The Irish verb to hunt).

## Method

### Participants

Forty two children with no history of psychological, neurological or sensory impairment (21 males) aged 5–16 years (mean = 11, sd = 4) were recruited from local schools via poster advertisement.

### Task and stimuli

The two tasks reported here were performed as part of a larger fMRI task battery investigating the development of language. In each task, twenty four stimuli were presented in a blocked event related design. The four task blocks were separated by 16 s of rest (with fixation to cross hairs). In each task block six stimuli were presented every 6 s, with a total task block duration of 36 s. Each task lasted 3 min and 49 s.

### Sentence comprehension and naming task

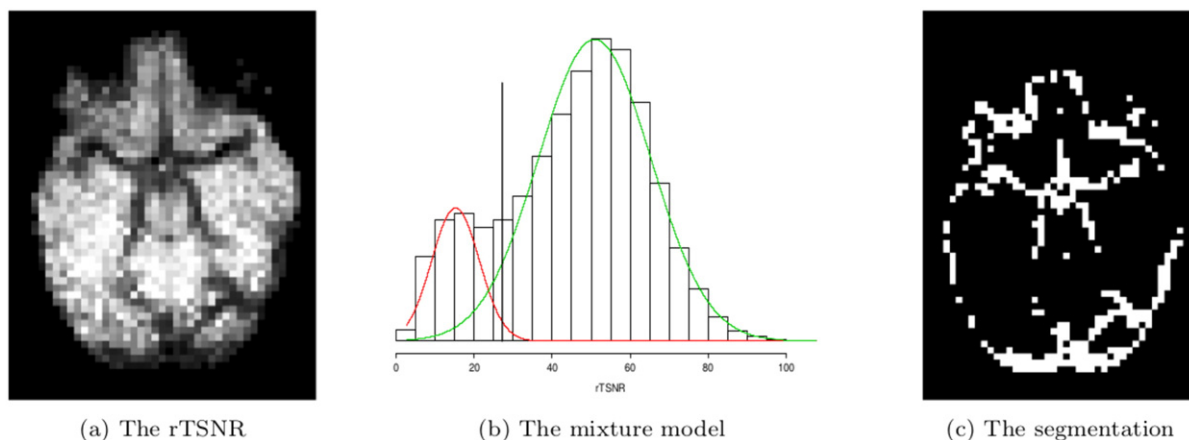
This comprehension task was designed to induce complex auditory processing, sentence level semantic and syntactic processing from phonological input, word retrieval and articulation. Participants were asked to listen to sentence level auditory descriptions of objects and animals and then name the object or animal.

### Sentence generation

This sentence generation task was designed to induce visual processing, syntactic processing from visual input, and articulation. Participants were asked to describe what was happening in a picture, using one simple sentence with a subject–verb–object (S–V–O) structure. Each  $350 \times 350$  pixel colour picture showed a subject and an object, with the subject involved in one of four actions; eating, drinking, jumping or falling.

### Data acquisition

Functional data were acquired using a 1.5 Tesla Siemens Avanto System (Erlangen, Germany), equipped with a 12 channel head coil. Signal changes were measured using a 2D gradient echo sequence (TR = 2160 ms, TE = 30 ms, flip angle = 75 degrees, FoV = 210 mm, 3 mm slice and 1 mm inter-slice gap, slices = 30 (axial acquisition in ascending order), voxel size =  $3.3 \times 3.3 \times 4$  mm). The first two volumes were discarded as dummy scans resulting in 104 volumes per task. Stimuli were presented using Cogent 2000 software; Cogent 2000 team at the Wellcome Trust Centre for Neuroimaging (WTCN) and the UCL Institute of Cognitive Neuroscience ([www.vislab.ucl.ac.uk](http://www.vislab.ucl.ac.uk)). Auditory stimuli were transmitted through MR compatible headphones equipped with



**Fig. 3.** Segmentation of high noise areas. The rTSNR of a lower axial slice (a) is fitted with two gaussians (b) and the high noise areas are segmented (c). The arteries, venous sinuses and the periphery of the brain are typically identified. The black line in (b) is the .05 quantile of the larger distribution where the masks are binarised.

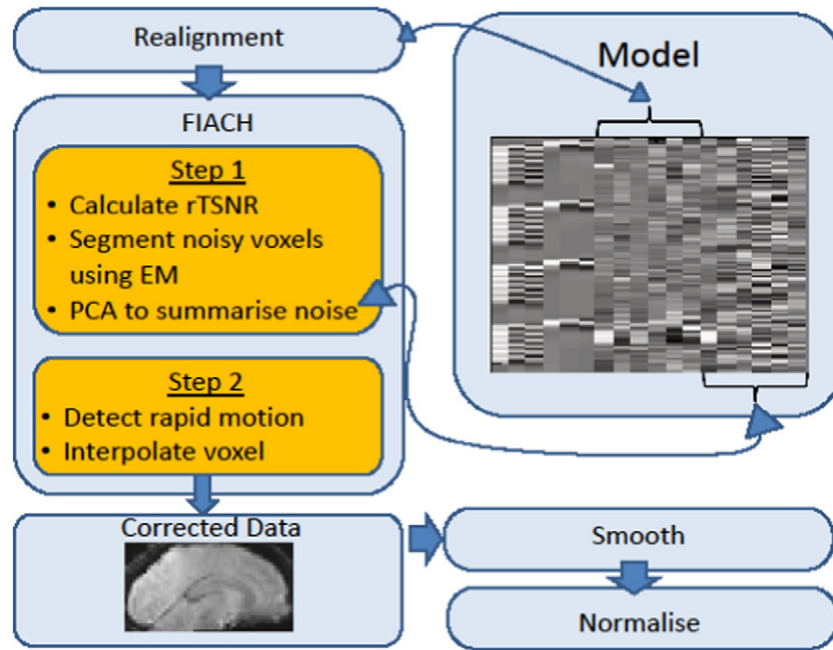


Fig. 4. Flowchart for an analysis pipeline using FIACH. The GLM design matrix includes the task, six realignment parameters and the PCA components.

an active noise cancellation system and speech responses were transmitted using a sensitive head-mounted microphone (<http://www.mrconfon.de/en/technology.html>). Visual stimuli were presented via an MR-compatible wall-mounted screen behind the scanner, and were viewed through a mirror fixed to the head coil.

#### FIACH validation

In order to validate FIACH we analysed data obtained from healthy children while performing two different language tasks. This population was chosen for three reasons; 1) Paediatric populations typically present with large amounts of motion. 2) There is considerable evidence in the literature as to what constitutes valid activations with regards language tasks (Price, 2010). 3) Both tasks induced task correlated motion as the children engaged in overt speech. This approach is similar to the one employed by Lund et al. (2005) for assessing the impact of motion on fMRI data.

To compare the different methods we identified regions of interest (ROI), a priori, which were likely to be active in the sentence comprehension and naming task from Price (2010) and Price (2012). These expected regions consisted of the following: left inferior frontal gyrus, bilateral superior temporal gyrus, bilateral middle temporal gyrus, bilateral primary motor, bilateral somatosensory, bilateral cerebellum, bilateral temporal pole, left hippocampus and supplementary motor area. With regards to the sentence generation task the regions were as follows: bilateral cerebellum, bilateral fusiform gyrus, bilateral middle occipital gyrus, bilateral inferior occipital gyrus, bilateral inferior temporal gyrus, bilateral middle temporal gyrus, supplementary motor area, bilateral primary motor, bilateral somatosensory, bilateral lingual gyrus, bilateral hippocampus, bilateral parahippocampal gyrus, and bilateral superior occipital gyrus. The AAL atlas (Tzourio-Mazoyer et al., 2002) was used to assist in anatomical localisation.

Binomial tests and Wilcoxon signed-rank tests were used to assess whether FIACH produced greater t-values and cluster extents than the other methods in the a priori defined regions of interest. All hypothesis tests, unless stated otherwise, are corrected for multiple comparisons using methods described by Hochberg (1988) at  $p = .05$ . All statistical maps, unless stated otherwise, are thresholded at FWE ( $p < .05$ ) with a cluster extent threshold = 0.

#### Data analysis pipelines

Five other methods of automatic retrospective noise control were also investigated to compare the performance of FIACH. The first method was spatial realignment followed by the regression of the RPs (we will refer to this as the SPM method). We also analysed the autoregressive polynomial expansion of the RPs and scan nulling (Friston et al., 1996; Lemieux et al., 2007). The threshold for scan nulling (also termed frame censoring) was set at 1 mm volume–volume displacement measured using the volumetric realignment parameters. This value is similar to the value of .9 mm described in Siegel et al. (2014) as producing the most optimal results. We will refer to this method as realignment parameter expansion (RPE).

At this point we also calculated the percentage of volumes that were censored for comparison with the percentage of data that the FIACH Step 2 (the filter and the 6 RPs, no physiological noise regression) altered. The fraction of data changed was calculated as follows.

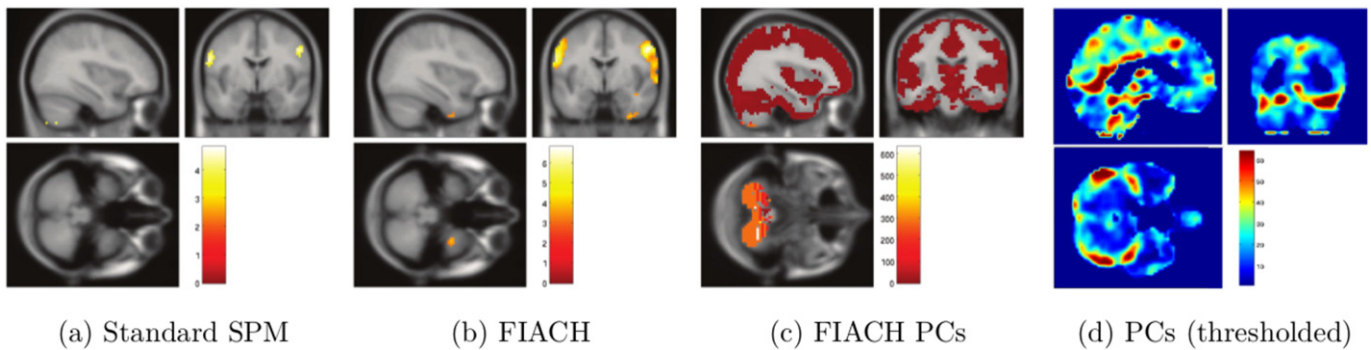
$$\text{No. of imputations} / (\text{No. of voxels in mask} \times \text{No. of volumes}) \quad (12)$$

The third method was motion fingerprint (MFP). This method contrasts with RPE as it reduces the motion model complexity. It also accounts for the varying effect of motion across the brain which RPE does not (Wilke, 2012). We also compared FIACH against two methods not reliant on the RPs, RWLS (Diedrichsen and Shadmehr, 2005) and tCompCor (Behzadi et al., 2007). We did not compare against aCompCor as in the original manuscript tCompCor is described as producing better results.

## Results

#### Illustrative example

We provide examples of both steps; 1) the physiological noise correction and 2) the correction for large amplitude signal changes. First we present a subject performing the sentence comprehension and naming task. This subject displayed no large amplitude signal changes. In Fig. 5 bilateral motor and artefactual activations in the cerebellum can be seen (the lower most portion of the cerebellum was not acquired. This resulted in signal instability due to motion interacting with the normalisation). After applying FIACH the motor cortex activations have



**Fig. 5.** Illustrative example of the physiological noise correction. a) Estimated t-map before correction. b) Estimated t-map after correction. c) F-contrast across the PCs. d) Same as 'c' but display thresholded at  $F = 60$ . All statistical maps are grey matter masked and thresholded at  $p < .001$  ( $k = 10$ ).

increased statistics and additional activations in the right superior temporal gyrus and the inferior temporal lobe are revealed (Fig. 5b). The cerebellar artefact is corrected due to the variance extraction in this area (Fig. 5c, axial image). In Fig. 5d we cap the maximum display values of Fig. 5c at  $F = 60$  to examine the spatial distribution of the regressors effect. Maxima are now seen in the transverse sinus, inferior sagittal sinus, fourth ventricle, brain stem and the inferior temporal lobe. All these areas exhibit large and consistent variability: the sinuses due to blood volume, as predicted in section 'Identifying physiological noise', the brain stem and fourth ventricle due to pulsatile motion. The inferior temporal signal instability is due susceptibility related artefacts that interact with bulk and pulsatile head motion. It is therefore suggested that the improvement (Fig. 5b) is a result of modelling these effects.

In Fig. 6 we show an example of the effect of the large amplitude signal change correction on a corrupted voxel time series. The artefact seen in Fig. 6a is corrected in Fig. 6b. No other time points are altered and neither are spatial locations outside the affected slice (see Fig. 4 for corrected image and Fig. 1 for the corrupted image).

#### Group level results

In the sentence comprehension and naming task FIACH produced the highest t-value in 8 out of the 16 ROIs specified. This was found to be a statistically significant effect using a binomial test ( $p < .05$ ). The probability of the proposed method having the highest t-value in the ROIs = 0.5 (95% CI [0.247, 0.753]). FIACH had the maximum number of voxels in 11 out of the 16 areas specified. This was found to be a statistically significant effect using a binomial test ( $p < .05$ ). The probability of the proposed method having the greatest extent in the region of interest = 0.688 (95% CI [0.413, 0.89]). Fig. 8 and Table 1 summarise the results obtained for the sentence and naming task using all 6 approaches.

In the sentence generation task FIACH had the highest t-value in 19 out of the 25 areas specified. This was found to be a statistically significant effect using a binomial test ( $p < .05$ ). The probability of the proposed method having the highest t-value in the regions of interest =

0.76 (95% CI [0.549, 0.906]). FIACH had the maximum number of voxels in 21 out of the 25 areas specified. This was found to be a statistically significant effect using a binomial test ( $p < .05$ ). The probability of the proposed method having the greatest extent in the regions of interest = 0.84 (95% CI [0.639, 0.955]). Fig. 8 and Table 2 summarise the results obtained for the sentence generation task using all 6 approaches

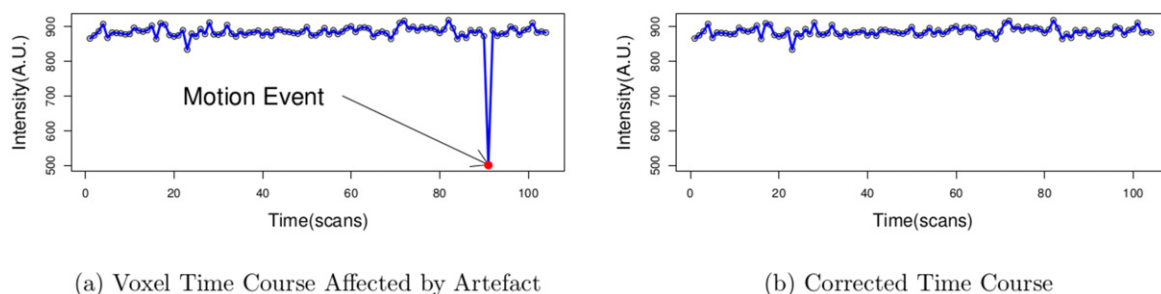
To compare the change in t-values and cluster extent ( $k$ ) in both tasks Wilcoxon signed-rank tests were used. The hypothesis that FIACH would outperform each of the other methods was tested. We also tested the hypothesis that FIACH Step 2 (the filter and the 6 RPs, no physiological noise regression) would provide improvements relative to the SPM approach in order to demonstrate the individual contribution of this step. All these hypotheses were found to be statistically significant at  $p < .05$ . The summaries of these tests are provided in Table 3.

As there was a large discrepancy between the percentage of data changed using FIACH compared to the percentage of data censored using scan nulling (see Fig. 7) we conducted further Wilcoxon signed rank tests to directly compare the FIACH (Step 2) only with RPE. It was found that using the FIACH (Step 2) in combination with 6 RPs resulted in more voxels and greater t-values in the sentence comprehension and naming task compared to the RPE approach ( $V = 103$ ,  $p < .05$ ;  $V = 116:5$ ,  $p < .05$ ). The same result was found in the sentence generation task ( $V = 195$ ,  $p < .05$ ;  $V = 224$ ,  $p < .05$ ).

## Discussion

### Summary of findings

We have evaluated the efficacy of 5 currently available automatic retrospective noise corrections methods in the context of paediatric fMRI where motion control is problematic. We have developed a method that performs favourably in comparison with these methodologies. FIACH produced increased t-value and cluster extent in plausible task related areas. Moreover this effect was largely consistent across brain areas. This was not the case for the other methodologies which had



**Fig. 6.** Example of FIACH step 2. a) Corrupted voxel time course. b) Corrected voxel time course.

**Table 1**  
Sentence comprehension, t-values(cluster extent) for all correction methods.

	FIACH	FIACH (Step 2)	MFP	RPE	RWLS	SPM	tCompCor
STGI	13.66(575)	11.75(494)	12.66(512)	10.92(423)	10.56(483)	10.8(470)	13.24(540)
STGr	13(327)	12.14(334)	12.52(354)	9.41(307)	11.54(314)	11.06(326)	12.31(335)
Cr	8.18(210)	5.92(60)	0(0)	0(0)	7.04(62)	6.22(53)	7.47(174)
MTGI	10.1(170)	8.62(83)	8.23(50)	6.49(37)	7.13(37)	7.38(63)	10.23(116)
Cl	8.65(167)	5.41(4)	5.99(2)	5.44(3)	6.46(15)	5.8(9)	6.08(30)
MTG	10.83(105)	10.85(97)	10.19(94)	8.26(67)	9.33(62)	9.38(80)	11.23(110)
Sr	8.35(101)	6.86(36)	7.56(61)	7.29(19)	6.46(8)	6.06(12)	7.19(55)
SI	7.7(92)	7.04(73)	7.92(61)	7.18(20)	8.2(78)	7.02(32)	7.09(52)
IFGI	7.25(90)	5.94(9)	6.62(30)	0(0)	6.34(15)	6.26(3)	7.95(30)
SMA	8.55(86)	5.97(4)	6.72(18)	6.73(21)	6.33(5)	0(0)	0(0)
PMI	8.38(70)	7.23(25)	7.94(46)	7.06(24)	8.33(59)	7.33(26)	7.28(32)
TPI	9.21(42)	7.41(29)	9.24(83)	7.65(35)	5.74(5)	7.66(25)	8.7(38)
ITGI	6.9(38)	7.41(28)	5.54(1)	0(0)	0(0)	5.47(1)	6.12(13)
HI	7.41(30)	6.49(53)	7.13(35)	7.14(17)	8.42(68)	6.81(57)	7.04(63)
PMr	7.88(29)	6.19(7)	7.41(26)	6.54(14)	6.84(16)	5.67(6)	6.53(11)
TPr	6.95(9)	7.51(8)	7.65(55)	6.57(14)	6.31(5)	7.59(8)	7.58(13)

STGI = left superior temporal gyrus, STGr = right superior temporal gyrus, Cr = right cerebellum, MTGI = left middle temporal gyrus, Cl = left cerebellum, MTGr = right middle temporal gyrus, Sr = right somatosensory, SI = left somatosensory, IFGI = left inferior frontal gyrus, SMA = supplementary motor area, PMI = left primary motor, TPI = left temporal pole, ITGI = left inferior temporal gyrus, HI = left hippocampus, PMr = right primary motor, TPr = right temporal pole. A value of 0 indicates that the area did not survive FWE ( $p < .05$ ).

variable performance across brain areas (see Table 1 & 2 and Fig. 8). We will now discuss some of the key findings from each task.

*Sentence comprehension and naming*

Of substantial interest is the varying performance of the reviewed methodologies in identifying activation in the inferior frontal gyrus. This area is of particular interest as it can be used to assess language laterality in candidates for brain surgery (Liegeois, 2002). It has also been documented that activations in this area correlate most with the results of the Wada test (Lehericy et al., 2000). A potential reason for this areas vulnerability to motion artefact is due to its position (regions further from the neck experience greater displacement when an individual rotates/nods their head). Due to its anterior position a rotation through plane is quite likely to result in the more anterior position of the slice being

excited twice in rapid succession leading to magnetisation saturation and a large signal reduction. It is also possible that the FIACH PCs are modelling the signal variance effects produced by the large vessels located near the gyrus. This would explain the large difference between the improved results of FIACH in this area compared to the other methodologies.

An interesting difference was also observed between the methodologies concerning the left inferior temporal gyrus. In this area the signal is very unstable due to its position at the brain's edge and the local susceptibility induced distortions which are exacerbated by motion. By the application of the Step 2 alone and a substantial activation is yielded (28 voxels, see Table 1). By applying both steps of FIACH the cluster size increases by a factor of ~3 (38 voxels) compared to the next most competitive method (see Table 1). The increased efficacy of FIACH is most likely due to the noise regressors being a reasonable model of the signal instability in these areas (see Figs. 5d & b, axial images). It has previously

**Table 2**  
Sentence generation, t-values(cluster extent) for all correction methods.

	FIACH	FIACH (Step 2)	MFP	RPE	RWLS	SPM	tCompCor
Cr	13.46(478)	9.46(191)	6.81(64)	7.53(76)	7.95(68)	8.46(111)	9.27(180)
FGr	14.3(392)	11.82(300)	7.94(244)	7.26(180)	8.95(197)	8.97(201)	10.72(285)
MOGr	11.48(366)	10.66(274)	10.24(279)	8.7(179)	11.53(301)	10.69(183)	7.83(120)
Cl	12.02(328)	8.73(115)	6.63(20)	6.83(41)	6.28(11)	6.28(41)	7.71(122)
FGI	12.99(320)	9.69(143)	7.85(161)	7.35(89)	8.47(116)	7.59(113)	8.37(124)
IOGI	12.15(256)	12.7(214)	7.9(127)	8.16(75)	10.52(114)	8.98(94)	8.71(183)
ITGI	12.46(217)	9.94(143)	7.38(164)	6.95(100)	9.17(120)	7.73(116)	8.61(121)
IOGr	14.03(195)	13.07(136)	7.7(95)	7.05(59)	8.39(58)	9.36(63)	9.41(112)
MOGI	11.85(170)	11.65(132)	7.98(117)	8.17(35)	11.08(120)	9.06(72)	7.33(28)
MTGI	10.55(134)	8.37(62)	7.06(37)	5.59(7)	7.64(47)	7.44(38)	8.11(21)
ITGr	13.7(116)	11.84(71)	7.75(53)	6.31(12)	9.01(37)	8.19(40)	9.28(61)
SMA	8.41(110)	0(0)	6.71(99)	0(0)	0(0)	5.39(5)	5.85(20)
PMI	7.97(69)	6.37(20)	7.32(46)	7.42(11)	7.02(11)	0(0)	5.58(7)
SI	8.19(62)	6.3(14)	7.1(22)	6.82(7)	7.39(19)	5.1(1)	6.51(21)
MTGr	10.61(60)	8.8(17)	6.15(11)	0(0)	6.98(7)	7.07(13)	7.44(10)
Sr	7.39(32)	0(0)	6.21(11)	5.5(1)	0(0)	0(0)	0(0)
LGr	7.55(27)	5.59(2)	0(0)	0(0)	0(0)	0(0)	5.18(1)
HI	6.56(22)	5.7(1)	6.02(4)	0(0)	7.47(14)	0(0)	0(0)
LGI	9.02(19)	7.52(5)	6.45(5)	6.36(3)	8.12(3)	6.31(4)	6.77(2)
PMr	6.7(18)	0(0)	6.47(20)	5.83(4)	0(0)	0(0)	0(0)
PHGI	6.39(8)	0(0)	6.68(12)	0(0)	5.54(1)	0(0)	0(0)
Hr	6.33(6)	6.36(14)	6.41(13)	5.82(1)	8.95(29)	5.91(3)	5.49(1)
PHGr	7.6(5)	7.35(4)	5.7(6)	0(0)	7.27(5)	6.45(4)	6.37(4)
SOGGr	5.92(5)	0(0)	0(0)	0(0)	6.15(1)	0(0)	0(0)
SOGI	5.73(4)	0(0)	0(0)	0(0)	5.72(1)	0(0)	0(0)

Cr = right cerebellum, FGr = right fusiform gyrus, MOGr = right middle occipital gyrus, Cl = left cerebellum, FGI = left fusiform gyrus, IOGI = left inferior occipital gyrus, ITGI = left inferior temporal gyrus, IOGr = right inferior occipital gyrus, MOGI = left middle occipital gyrus, MTGI = left middle temporal gyrus, ITGr = right inferior temporal gyrus, SMA = supplementary motor area, PMI = left primary motor, SI = left somatosensory, MTGr = right middle temporal gyrus, Sr = right somatosensory, LGr = right lingual gyrus, HI = left hippocampus, LGI = left lingual gyrus, PMr = right primary motor, PHGI = left parahippocampal gyrus, Hr = Right Hippocampus, PHGr = right parahippocampal gyrus, SOGr = right superior occipital gyrus, SOGI = left superior occipital gyrus. A value of 0 indicates that the area did not survive FWE ( $p < .05$ ).

**Table 3**  
Summary of Wilcoxon signed-rank tests for both tasks.

	$\Delta$ Median	Statistic (V)		$\Delta$ Median	Statistic (V)
FIACH(t) > FIACH step 2(t)	1.23	130***	FIACH(t) > FIACH step 2(t)	1.89	320***
FIACH(t) > MFP(t)	0.76	124**	FIACH(t) > MFP(t)	1.42	321***
FIACH(t) > RPE(t)	1.47	136***	FIACH(t) > RPE(t)	2.13	325***
FIACH(t) > RWLS(t)	1.43	129***	FIACH(t) > RWLS(t)	2.09	300***
FIACH(t) > SPM(t)	1.45	134***	FIACH(t) > SPM(t)	2.11	325***
FIACH(t) > tCompCor(t)	0.99	116**	FIACH(t) > tCompCor(t)	1.65	325***
FIACH step 2(t) > SPM(t)	0.22	101*	FIACH step 2 (t) > SPM(t)	1.2	192***
FIACH(k) > FIACH step 2 (k)	58.5	126**	FIACH(k) > FIACH step 2(k)	52	320***
FIACH(k) > MFP(k)	43	110*	FIACH(k) > MFP(k)	47	312***
FIACH(k) > RPE(k)	71.5	135***	FIACH(k) > RPE(k)	62	325***
FIACH(k) > RWLS(k)	64.5	128***	FIACH(k) > RWLS(k)	55	293***
FIACH(k) > SPM(k)	65.5	130***	FIACH(k) > SPM(k)	64	325***
FIACH(k) > tCompCor(k)	46	120**	FIACH(k) > tCompCor(k)	49	325***
FIACH Step 2(k) > SPM(k)	7	108**	FIACH step 2 (k) > SPM(k)	12	185***

(a) Sentence comprehension and naming

(b) Sentence generation

t = t-value, k = cluster extent. Statistic (V) = is the non-parametric equivalent of a t-value in the Wilcoxon-signed rank test.

\*\*\* Significant at  $p < .001$ .

\*\* Significant at  $p < .01$ .

\* Significant at  $p < .05$ .

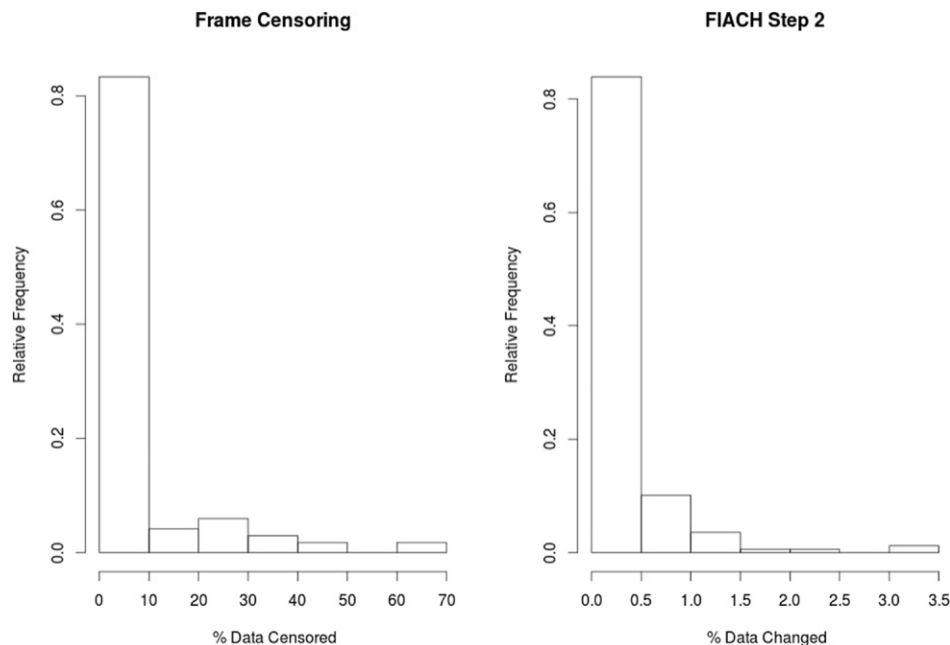
been reported that due to the high magnetic susceptibility effects in the areas next to the air-filled sinuses it would be more appropriate to use small volume corrections for observing activations in these areas (Devlin et al., 2000). Furthermore, in the same study it was shown that these activations are observable with PET and less so with fMRI. Optimised sequences have also been developed for observing activations in these areas (Weiskopf et al., 2006) which produce moderate improvement in signal levels. However, we show that by directly modelling the temporal instability (see Fig. 9a and Table 1) in these areas it is possible to recover the activations at high statistical thresholds. This suggests that it is not sufficient to recover signal in these areas but also necessary to model sources of signal instability to recover activations in these regions.

Another brain region of interest for speech/language is the cerebellum. The cerebellum has been heavily implicated in articulation and language and substantial activations would be expected in this area (Price, 2010). The large difference in activation strength between FIACH and

the other methods is due to the use of the FIACH PCs. These regressors typically model information from the transverse sinuses which are adjacent to the cerebellum. The sinuses cause local disruptions to the main (B0) magnetic field and can therefore corrupt the surrounding area. This problem has been identified previously by Winawer et al. (2010) in the context of fMRI studies of the visual system.

#### Sentence generation

This task was more corrupted by motion as the children overtly produced entire sentences. This also served to increase the correlation of the task with the motion parameters. It is unsurprising therefore that the t-values of the RPE method decrease in 19 of the 25 areas compared to the standard SPM method and that the cluster extent decreases in 21 out of the 25 areas. The overall poor performance of RPE can potentially be explained by the limitations of the RPs. In this task the head motion was task correlated and as a result it is likely that the correction methods

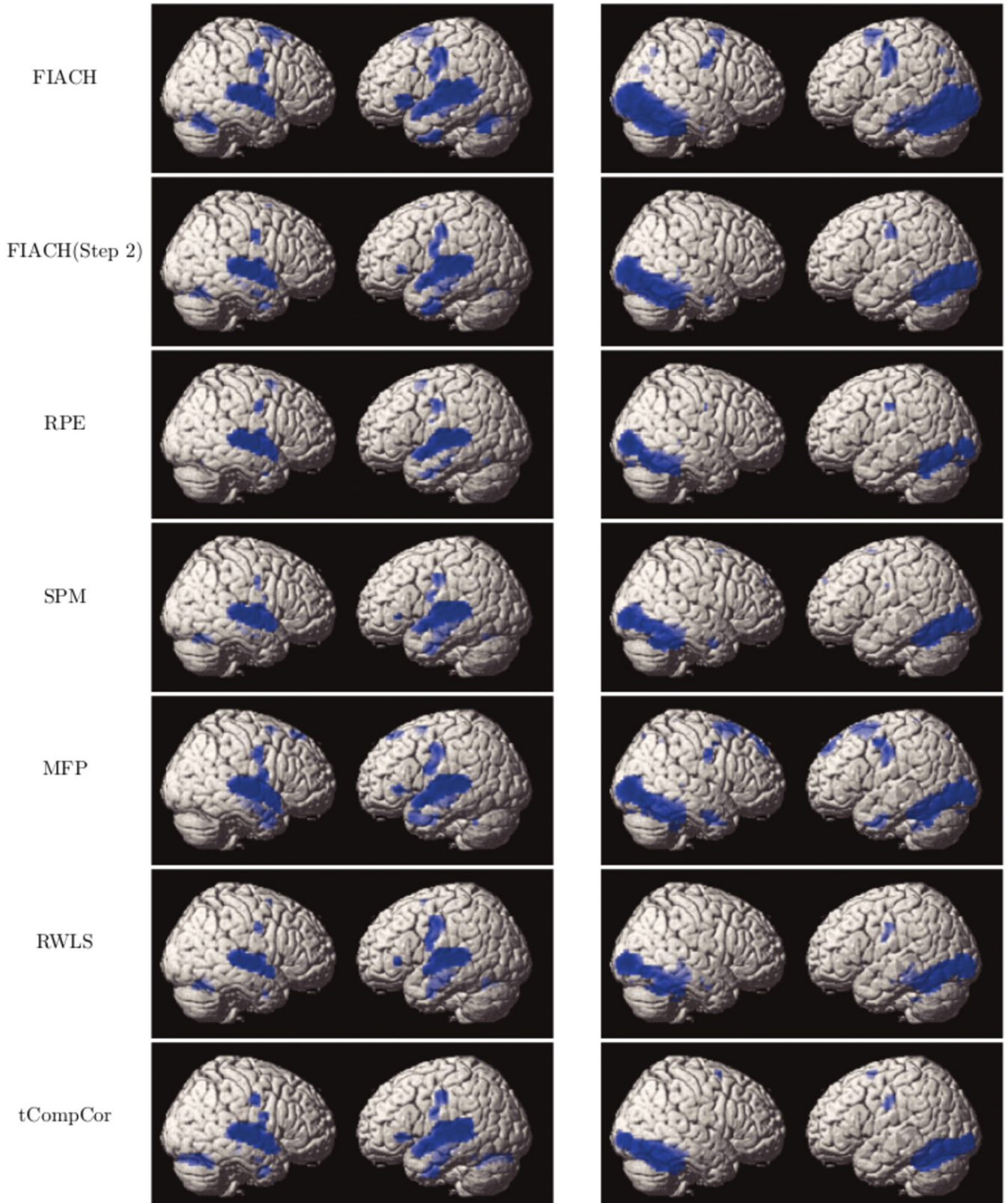


**Fig. 7.** Comparison of FIACH (Step 2) with Frame Censoring across both tasks. The histograms demonstrate that traditional frame censoring using a 1 mm threshold results in on average ~20 times more data changes than using FIACH. Mean(FIACH) = 0.25%, mean (censored) = 6.18%, range (FIACH) = 0–3.41%, range (censored) = 0–65.38%.



(a) Sentence Comprehension and Naming

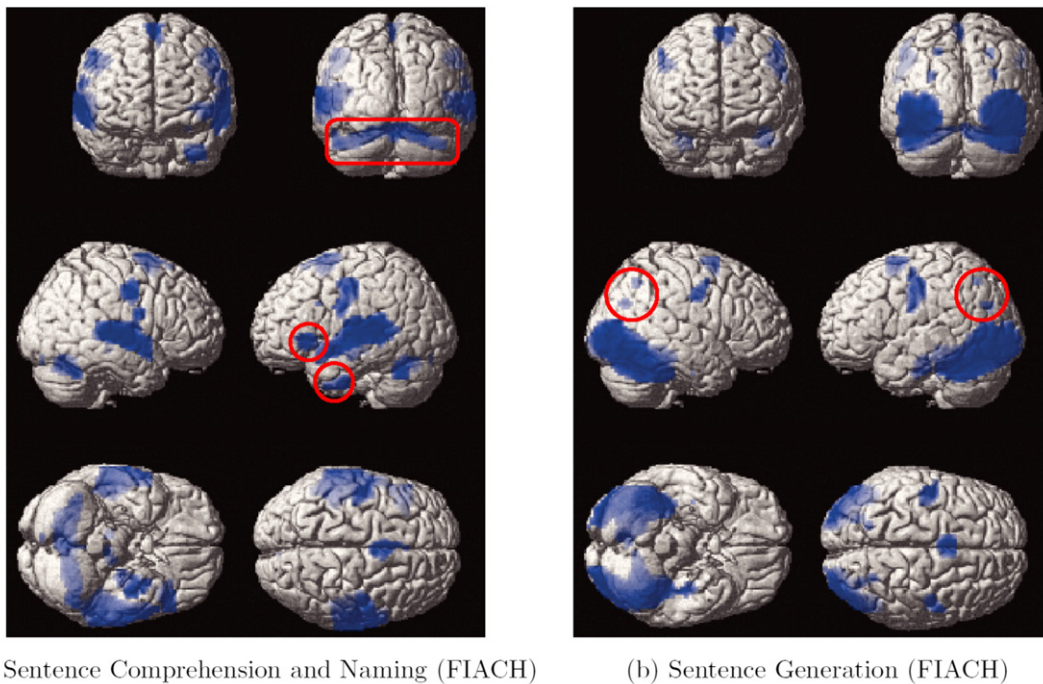
(b) Sentence Generation



**Fig. 8.** t-maps for both tasks (task > rest) rendered on a normalised T1, thresholded at FWE ( $p < .05$ ). Intensity is proportional to the activations distance from the cortical surface.

derived from these parameters regress true information from task related areas. MFP, although similarly derived from motion parameters, reduces model complexity and therefore is more sensitive than RPE. The reduction

in model complexity is not a trivial issue as, on average, the lower bound for the amount of real signal variance removed by chance will be  $\text{rank}(\text{nuisance regressors})/\text{number of scans}$  (Friston et al., 1996).



**Fig. 9.** Task t-maps (task > rest) generated by FIACH rendered on normalised T1, thresholded at FWE ( $p < .05$ ). Areas where FIACH substantially improved statistical power relative to other methods are surrounded in red. Intensity is proportional to the activation distance from the cortical surface.

The difference between FIACH and the other methodologies is much more pronounced in this task (see Table 2). This is because the majority of expected activations (inferior temporal, fusiform, occipital) all are next to areas of pronounced susceptibility (air filled sinuses, transverse sinus). Furthermore, the increased period of overt speech in this task leads to far more through plane movements which FIACH is optimally suited to correct. This is particularly important as the use of overt tasks has been advocated over covert tasks (Croft et al., 2013). The other methods lack the spatial specificity to substantially correct these artefacts.

An interesting and unexpected finding was FIACH's ability to identify activations in the inferior parietal regions where other methods failed to do so (see Fig. 9b). The inferior parietal lobe has recently been implicated in cross-modal semantic processing (Binder et al., 2009; Seghier, 2013). However, parietal activations have only been reported previously in two studies of sentence production in adults (Muller, 1997; Blank et al., 2002). As such, we did not hypothesise activation in these regions. Crucially, the fact that FIACH may be able to enhance activation of these regions is encouraging, as it may enable us to better understand the neural substrates of language production.

Activation in the IFG was not hypothesised for this task as results from previous studies in adults are variable (Awad et al., 2007; Tremblay and Small, 2011). Further, functional neuroanatomical models of language predict tasks with a large semantic load induce activation of the IFG (Price, 2012). However, the semantic and syntactic demands of the sentence generation task were relatively low; participants were asked to construct a sentence based on images which contained animals and objects with high word frequency, using a limited set of only four common verbs.

#### Local and global signal changes.

A large difference was observed between the amount of data that was changed by FIACH step 2 and by frame censoring. This highlights that aggressive cleaning strategies may be sub-optimal. This may be attributed to the most damaging effects of motion being local (due to repeated excitation of the same tissue or in regions of spatially fast

changing signal intensity such as at the edges of the brain) while the average signal changes across the brain are much lower in magnitude. The results of Fig. 7 suggest that when an individual meets the threshold for a frame to be censored the image does not always exhibit widespread signal changes large enough to be discriminated from BOLD signal amplitudes.

This difference could also be attributed to the limited accuracy in using volumetric realignment parameters as models of motion. This has been previously highlighted by Beall and Lowe (2014) who had ground truth measures of motion which we do not have in this study.

It should also be noted that the FIACH model is very conservative in terms of data correction so a direct comparison between FIACH step 2 and RPE is difficult. Less conservative thresholds for FIACH step 2 or more conservative thresholds for censoring will result in more comparable amounts of data being changed/censored. Furthermore FIACH may correct artefacts due to electrical noise which frame censoring (based on RPs) cannot. As such this further limits the comparison. Bearing in mind these limitations the application of FIACH step 2 was still seen to increase t-values and cluster extent relative to the RPE method. This highlights that there may be diminishing returns to using expansions of the realignment parameters.

More widespread signal changes that are lower in magnitude may become more problematic in less efficient designs such as in those seen in studies of epilepsy where the method of censoring used in this paper was developed (Lemieux et al., 2007). As such the combination of FIACH step 2 and censoring may be beneficial in these situations. Furthermore if there are global changes still present in the data after FIACH step 2 (or censoring) a statistically elegant strategy would be to use a weighted least squares approach such as RWLS as a subsequent step to FIACH.

#### Methodological considerations and limitations

A primary limitation of FIACH is the single imputation of values for corrupt time points. While we do show improvements in observed activations a more statistically optimal strategy would be to use a multiple imputation framework (Rubin, 1987). A potential secondary benefit of

multiple imputation would be for spectral analysis of rs-fcMRI which can potentially be affected by single imputation/scrubbing (replacing with mean/median). Future work will focus on the development of this approach in the context of fMRI.

It is also worth considering the choice of model parameters used in this study. In this work we deliberately chose to estimate the maximum possible magnitude of the BOLD response. This was done to deliberately minimise data alteration, and in the knowledge that children and patients may have responses greater than the range typically found in healthy adults. More realistic parameters could be chosen but we demonstrate that FIACH still performs favourably in comparison with other methods even with a conservative approach to data correction. We have demonstrated that applying FIACH produces superior results in both tasks compared to the other methods reviewed.

However, we have not explored the possibility of combining FIACH with other methods for a more optimal method. For instance one could combine the FIACH regressors with motion estimates from Motion Fingerprint and then estimate the model using RWLS. Frame-censoring/scan nulling could also be applied. This is because the results presented are not primarily intended to be comparative but highlight the potential gains from utilising a biophysical framework for extending traditional approaches.

## Conclusion

We have developed a biophysical framework for extending the traditional approach of using volumetric measures of motion to control for noise in fMRI. It can be integrated as a pre-processing step within current standard strategies for mass-univariate GLM based analysis and rs-fcMRI. This method is a two-step procedure involving the identification and correction of large amplitude signal changes and modelling the effects of regions of high temporal instability.

We have demonstrated its efficacy in a paediatric population during overt speech where subject motion is a severely limiting factor in both research and clinical applications of fMRI. We have shown that FIACH reveals additional brain areas involved in language at the group and individual level. It also substantially increases the statistical power in language related areas relative to other methods. Furthermore, this methodology is capable of improving results in regions of low SNR due to local field inhomogeneities (inferior temporal and high blood volume regions). These areas have proven particularly problematic for fMRI and FIACH provides the opportunity to improve knowledge of the function within these areas.

## Acknowledgments

The authors would like to thank Helmut Laufs and the anonymous reviewers for helpful advice with regards to the writing of the manuscript. TT and ES were funded by University College London IMPACT, the James Lewis Foundation, the Child Health Research Appeal Trust, ES additionally by a UCL Overseas Research Scholarship. MC was funded by Action Medical Research (SP4646). The data collection and LWC were funded by Epilepsy Research UK (P1106) and Great Ormond Street Children's Charity (V1213). SP was funded by the NIHR Biomedical Research Centre and Dementia Unit (BRC/U) at South London and Maudsley NHS Foundation Trust and the Institute of Psychiatry, Psychology and Neuroscience, King's College London. This work was undertaken at Great Ormond Street Hospital/UCL Institute of Child Health who received a proportion of funding from the UK Department of Health's NIHR Biomedical Research centres funding scheme.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2015.09.034>.

## References

- Awad, M., Warren, J.E., Scott, S.K., Turkheimer, F.E., Wise, R.J.S., 2007. A common system for the comprehension and production of narrative speech. *The Journal of Neuroscience* 27 (43), 11455–11464.
- Beall, E.B., Lowe, M.J., 2014. SimPACE: generating simulated motion corrupted BOLD data with synthetic-navigated acquisition for the development and evaluation of SLOMOCO: a new, highly effective slice-wise motion correction. *NeuroImage* 101, 21–34.
- Beckmann, C., Smith, S., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on* 23 (2), 137–152.
- Behzadi, Y., Restom, K., Liu, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage* 37 (1), 90–101.
- Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex* 19 (12), 2767–2796.
- Birn, R.M., Diamond, J.B., Smith, M.A., Bandettini, P.A., 2006. Separating respiratory variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. *NeuroImage* 31, 1536–1548.
- Blank, S., Scott, S., Murphy, K., Warburton, E., Wise, R., 2002. Speech production: Wernicke, Broca and beyond. *Brain* 125, 1829–1838.
- Buxton, R.B., Wong, E.C., Frank, L.R., 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magnetic Resonance in Medicine* 39 (6), 855–864.
- Croft, L.J., Rankin, P.M., Liégeois, F., Banks, T., Cross, J.H., Vargha-Khadem, F., Baldeweg, T., 2013. To speak, or not to speak? The feasibility of imaging overt speech in children with epilepsy. *Epilepsy Research* 107 (1–2), 195–199.
- Clayden, J., 2015. RNiftyReg: Image Registration Using the NiftyReg Library. R package version 2.0.1 <https://CRAN.R-project.org/package=RNiftyReg>.
- Desjardins, E.A., Kiehl, K.A., Liddle, P.F., 2001. Removal of confounding effects of global signal in functional MRI analyses. *NeuroImage* 13 (4), 751–758.
- Devlin, J.T., Russell, R.P., Davis, M.H., Price, C.J., Wilson, J., Moss, H.E., Matthews, P.M., Tyler, L.K., 2000. Susceptibility-induced loss of signal: comparing PET and fMRI on a semantic task. *NeuroImage* 11 (6), 589–600.
- Diedrichsen, J., Shadmehr, R., 2005. Detecting and adjusting for artifacts in fMRI time series data. *NeuroImage* 27 (3), 624–634.
- Eddelbuettel, D., Francois, R., 2011. Rcpp: seamless R and C++ integration. *J. Stat. Softw.* 40 (8), 1–18.
- Eddelbuettel, D., Sanderson, C., 2014. RcppArmadillo: accelerating R with high-performance C++ linear algebra. *Comput. Stat. Data Anal.* 71, 1054–1063.
- Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S., Turner, R., 1996. Movement-related effects in fMRI time-series. *Magn. Reson. Med.* 35 (3), 346–355.
- Glover, G.H., Li, T.Q., Ress, D., 2000. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn. Reson. Med.* 44 (March), 162–167.
- Grubb, R.L., Raichle, M.E., Eichling, J.O., Ter-Pogossian, M.M., 1974. The effects of changes in PaCO<sub>2</sub> cerebral blood volume, blood flow, and vascular mean transit time. *Stroke* 5 (5), 630–639.
- Hochberg, Y., 1988. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75, 800–803.
- Kundu, P., Inati, S.J., Evans, J.W., Luh, W.M., Bandettini, P.A., 2012. Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage* 60 (3), 1759–1770.
- Lauwers, F., Cassot, F., Lauwers-Cances, V., Puwanarajah, P., Duvernoy, H., 2008. Morphometry of the human cerebral cortex microcirculation: general characteristics and space-related profiles. *NeuroImage* 39 (3), 936–948.
- Lehericy, S., Cohen, L., Bazin, B., Samson, S., Giacomini, E., Rougetet, R., ... Baulac, M., 2000. Functional MR evaluation of temporal and frontal language dominance compared with the Wada test. *Neurology* 54 (8), 1625–1633.
- Lemieux, L., Salek-Haddadi, A., Lund, T.E., Laufs, H., Carmichael, D., 2007. Modelling large motion events in fMRI studies of patients with epilepsy. *Magn. Reson. Imaging* 25 (6), 894–901.
- Liegeois, F., 2002. A direct test for lateralization of language activation using fMRI: comparison with invasive assessments in children with epilepsy. *NeuroImage* 17 (4), 1861–1867.
- Lund, T.E., Norgaard, M.D., Rostrup, E., Rowe, J.B., Paulson, O.B., 2005. Motion or activity: their role in intra- and inter-subject variation in fMRI. *NeuroImage* 26 (3), 960–964.
- Mildner, T., Norris, D.G., Schwarzbauer, C., Wiggins, C.J., 2001. A qualitative test of the balloon model for BOLD-based MR signal changes at 3 T. *Magn. Reson. Med.* 46 (5), 891–899.
- Muller, R.-A., Rothermel, R.D., Behen, M.E., Muzik, O., Mangner, T.J., Chugani, H.T., 1997. Receptive and expressive language activations for sentences. *NeuroReport* 8 (17), 3767–3770.
- Murphy, K., Birn, R.M., Handwerker, D. a, Jones, T.B., Bandettini, P.A., 2009. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *NeuroImage* 44, 893–905.
- Murphy, K., Birn, R.M., Bandettini, P.A., 2013. Resting-state fMRI confounds and cleanup. *NeuroImage* 80, 349–359.
- 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.
- Obata, T., Liu, T.T., Miller, K.L., Luh, W.-M., Wong, E.C., Frank, L.R., Buxton, R.B., 2004. Discrepancies between BOLD and flow dynamics in primary and supplementary motor areas: application of the balloon model to the interpretation of BOLD transients. *NeuroImage* 21 (1), 144–153.

- Power, J.D., Barnes, K. a, Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2011. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage* 59, 2142–2154.
- Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2014a. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage* 84, 320–341.
- Power, J.D., Schlaggar, B.L., Petersen, S.E., 2014b. Recent progress and outstanding issues in motion correction in resting state fMRI. *NeuroImage* 1–16.
- Price, C.J., 2010. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann. N. Y. Acad. Sci.* 1191, 62–88.
- Price, C.J., 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage* 62 (2), 816–847.
- R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria <https://www.R-project.org/>.
- Rubin, D.B., 1987. Multiple Imputation for Nonresponse in Surveys. Wiley, New York.
- 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.
- Satterthwaite, T.D., Elliott, M. a, Gerraty, R.T., Ruparel, K., Loughhead, J., Calkins, M.E., Wolf, D.H., 2013. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage* 64, 240–256.
- Seghier, M.L., 2013. The angular gyrus: multiple functions and multiple subdivisions. *The Neurosci.: Rev. J. Bringing Neurobiol., Neurol. Psychiatry* 19 (1), 43–61.
- Siegel, J., Power, J., Dubis, J., 2014. Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. *Hum. Brain Mapp.* 35 (5), 1981–1996.
- 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. *Cereb. Cortex* 20 (12), 2852–2862.
- Tremblay, P., Small, S.L., 2011. Motor response selection in overt sentence production: a functional MRI study. *Front. Psychol.* 2 (September), 253.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15 (1), 273–289.
- Uladag, K., Muller-Bierl, B., Ugurbil, K., 2009. An integrative model for neuronal activity-induced signal changes for gradient and spin echo functional imaging. *NeuroImage* 48 (1), 150–165.
- Van Dijk, K.R. a, Sabuncu, M.R., Buckner, R.L., 2011. The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage* 59 (1), 431–438.
- Weiskopf, N., Hutton, C., Josephs, O., Deichmann, R., 2006. Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: a whole-brain analysis at 3 T and 1.5 T. *NeuroImage* 33 (2), 493–504.
- Weisskoff, R., Kiihne, S., 1992. MRI susceptometry: Image-based measurement of absolute susceptibility of MR contrast agents and human blood. *Magn. Reson. Med.* 383, 375–383.
- Wilke, M., 2012. *NeuroImage*: an alternative approach towards assessing and accounting for individual motion in fMRI timeseries. *NeuroImage* 59 (3), 2062–2072.
- Winawer, J., Horiguchi, H., Sayres, R., 2010. Mapping hV4 and ventral occipital cortex: the venous eclipse. *J. Vis.* 10 (5), 1–22.
- Yablonskiy, D., Haacke, E., 1994. Theory of NMR signal behavior in magnetically inhomogeneous tissues: the static dephasing regime. *Magn. Reson. Med.* 32, 749–763.