ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 31 (2006) 1104 – 1115

Physiological noise reduction for arterial spin labeling functional MRI

Khaled Restom,[a,b] Yashar Behzadi,[a,b] and Thomas T. Liu[a,*]

[a]*Center for Functional Magnetic Resonance Imaging and Department of Radiology, UCSD Center for Functional MRI, 9500 Gilman Drive, MC 0677, La Jolla, CA 92093-0677, USA*
[b]*Department of Bioengineering, University of California San Diego, La Jolla, CA, USA*

Received 20 July 2005; revised 5 December 2005; accepted 24 January 2006
Available online 13 March 2006

Three methods for the reduction of physiological noise in arterial spin labeling (ASL) functional magnetic resonance imaging (fMRI) are presented and compared. The methods are based upon a general linear model of the ASL measurement process and on a previously described retrospective image-based method (RETROICOR) for physiological noise reduction in blood oxygenation level dependent fMRI. In the first method the contribution of physiological noise to the interleaved control and tag images that comprise the ASL time series are assumed to be equal, while in the second method this assumption is not made. For the third method, it is assumed that physiological noise primarily impacts the perfusion time series obtained from the filtered subtraction of the control and tag images. The methods were evaluated using studies of functional activity in the visual cortex and the hippocampal region. The first and second methods significantly improved statistical performance in both brain regions, whereas the third method did not provide a significant gain. The second method provided significantly better performance than the first method in the hippocampal region, whereas the differences between methods were less pronounced in visual cortex. The improved performance of the second method in the hippocampal region appears to reflect the relatively greater effect of cardiac fluctuations in this brain region. The proposed methods should be particularly useful for ASL studies of cognitive processes where the intrinsic signal to noise ratio is typically lower than for studies of primary sensory regions.

Introduction

Arterial spin labeling (ASL) is a non-invasive magnetic resonance imaging (MRI) method for the measurement of cerebral blood flow (CBF) (Detre et al., 1992; Williams et al., 1992; Kim, 1995). Because CBF is tightly linked to neural activity, the use of ASL to measure CBF can enhance the interpretation of functional (fMRI) experiments. As compared to the blood oxygenation level dependent (BOLD) measures that are typically used in fMRI studies, ASL measures have the potential of providing better localization of the sites of neural activity (Luh et al., 2000; Duong et al., 2001) since they reflect the delivery of blood to the capillary beds, whereas BOLD can reflect deoxyhemoglobin in venules that are far downstream from the sites of activity. In addition, ASL methods have some practical advantages, including (i) an inherent insensitivity to the low-frequency fluctuations commonly observed in fMRI experiments (Aguirre et al., 2002; Wang et al., 2003b), (ii) the ability to take advantage of imaging sequences (e.g. spin-echo) that are insensitive to susceptibility induced off-resonance effects (Wang et al., 2004) and (iii) the potential for less inter-subject variability (Aguirre et al., 2002; Wang et al., 2003b; Tjandra et al., 2005).

Despite their potential advantages, the use of ASL methods for fMRI studies has been rather limited, especially for studies of cognitive processes. A significant limitation of ASL methods is their relatively low sensitivity as compared to BOLD-based methods (Buxton, 2002). Improved sensitivity, either through an increase in intrinsic signal or a reduction of noise, would greatly enhance the utility of ASL for a broader spectrum of fMRI studies. In this paper, we examine the improvement in sensitivity that can be obtained through the reduction of noise due to physiological sources.

Physiological fluctuations have been shown to be a significant source of noise in BOLD fMRI experiments, especially at higher field strengths (Kruger and Glover, 2001). A number of methods have been developed for the reduction of physiological noise in fMRI experiments. These include image based retrospective correction (RETROICOR), *k*-space based retrospective correction (RETROKCOR) and navigator echo based correction (DORK) (Hu et al., 1995; Josephs et al., 1997; Glover et al., 2000; Pfeuffer et al., 2002b). At present the application of these methods has been focused primarily on BOLD experiments with a brief mention in (Pfeuffer et al., 2002a) of the application of the DORK method to ASL data acquired at 7 Tesla.

In this work, we consider several extensions of the RETROICOR method for the reduction of physiological noise in ASL data. We first present the extensions within the framework of a general linear model (GLM) for ASL experiments and define a metric for assessing

* Corresponding author. Fax: +1 858 822 0605.
E-mail address: ttliu@ucsd.edu (T.T. Liu).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2006.01.026

statistical significance. We then compare the performance of the methods using experimental data obtained from functional ASL studies of the visual cortex and hippocampal region.

Theory

General linear model for arterial spin labeling

In an arterial spin labeling (ASL) experiment, a series of control images and tag images in which arterial blood is either fully relaxed or magnetically inverted, respectively, is acquired (Golay et al., 2004). Typically, the control and tag images are acquired in an interleaved fashion and a perfusion time series is formed from the filtered subtraction of control and tag images. In a previous paper (Liu and Wong, 2005) we showed that ASL time series data could be modeled as

$$y[n] = b[n]\left(s_M M_0 + s_q q[n]\right) + (-1)^{n+1} b[n]q[n]\alpha e^{-TI/T_{1B}} + e[n] \quad (1)$$

where $y[n]$ represents interleaved tag (even indices) and control images (odd indices), $b[n]$ represents multiplicative BOLD weighting, $q[n]$ is a term that is proportional to cerebral blood flow (CBF), and $e[n]$ represents additive noise. The remaining parameters are defined as follows: T_1 and T_{1B} are the longitudinal time constants of tissue and arterial blood, respectively; M_0 represents the magnetization of static tissue in the imaging volume of interest; TI and TI_P represent inversion and pre-saturation times, respectively; α denotes inversion efficiency; β denotes the action of an optional pre-saturation pulse; and $s_M = 1-\beta e^{-TI_p/T_1}$ and $s_q = 1-\alpha e^{-TI/T_{1B}}$.

For statistical analysis, it is convenient to rewrite Eq. (1) in matrix form as

$$\mathbf{y} = s_m M_0 \mathbf{b} + s_q \mathbf{q} \circ \boldsymbol{b} + \alpha\exp(-TI/T_{1B})\boldsymbol{M}(\mathbf{q} \circ \mathbf{b}) + \mathbf{U}_c\mathbf{D}_c\mathbf{P}\mathbf{c}_c + \mathbf{U}_t\mathbf{D}_t\mathbf{P}\mathbf{c}_t + \mathbf{e} \quad (2)$$

where $\mathbf{y}$, $\mathbf{b}$, $\mathbf{q}$, and $\mathbf{e}$ are the vector equivalents (each with dimension $N \times 1$ where N is the number of data points) of $y[n]$, $b[n]$, $q[n]$, and $e[n]$, respectively, $\mathbf{M}$ is a $N \times N$ diagonal matrix consisting of alternating -1 and 1 s along the diagonal, and $\mathbf{q}\circ\mathbf{b}$ denotes the Hadamard product of the two vectors (i.e. element-by-element multiplication). The alternating -1 and 1 s in the matrix $\mathbf{M}$ represent the modulation of the magnetization of arterial blood in the interleaved tag and control images, respectively. We have also added physiological noise terms $\mathbf{U}_c\mathbf{D}_c\mathbf{P}\mathbf{c}_c + \mathbf{U}_t\mathbf{D}_t\mathbf{P}\mathbf{c}_t$ where $\mathbf{P}$ is a $N \times m$ matrix containing m regressors and $\mathbf{c}_c$ and $\mathbf{c}_t$ are unknown regressor weights for the control and tag images, respectively. The calculation of these regressors follows the approach presented in (Glover et al., 2000) for the RETROICOR method, and is described in detail in the appendix. The term $\mathbf{U}_c\mathbf{D}_c\mathbf{P}\mathbf{c}_c$ represents physiological noise contributions to the control images where $\mathbf{D}_c$ is a downsampling matrix that picks out every odd sample of $\mathbf{P}\mathbf{c}_c$ and $\mathbf{U}_c$ is an upsampling matrix that inserts zeros between samples of $\mathbf{D}_c\mathbf{P}\mathbf{c}_c$. The matrices $\mathbf{D}_t$ and $\mathbf{U}_t$ are defined similarly, with $\mathbf{D}_t$ picking out even samples of $\mathbf{P}\mathbf{c}_t$. Examples of downsampling and upsampling matrices are given in (Liu et al., 2002).

Most perfusion estimates are based upon filtered subtraction methods that attenuate the un-modulated term $s_m M_0\mathbf{b} + s_q\mathbf{q}\circ\mathbf{b}$ while preserving the modulated term $\alpha\exp(-TI/T_{1B})\mathbf{M}(\mathbf{q}\circ\mathbf{b})$. As shown in (Liu and Wong, 2005), this process can be expressed as a modulation operation followed by a low pass filtering operation

$$\hat{\mathbf{q}} = \mathbf{G}\mathbf{M}\mathbf{y} \approx \mathbf{G}\left(\mathbf{q}\circ\tilde{\mathbf{b}}\right) + \mathbf{G}\mathbf{U}_c\mathbf{D}_c\mathbf{P}\mathbf{c}_c - \mathbf{G}\mathbf{U}_t\mathbf{D}_t\mathbf{P}\mathbf{c}_t + \mathbf{n} \quad (3)$$

where $\hat{\mathbf{q}}$ is a BOLD-weighted and filtered version of the true perfusion signal $\mathbf{q}$, $\mathbf{G}$ is a low pass filtering matrix, $\tilde{\mathbf{b}} = \alpha e^{-TI/T_{1B}}\mathbf{b}$ is the BOLD-weighting term scaled by the constant term $\alpha e^{-TI/T1B}$, and $\mathbf{n}$ is the additive noise after modulation and filtering. Note that we have made use of the identities $\mathbf{M}\mathbf{U}_c\mathbf{D}_c\mathbf{P}\mathbf{c}_c = \mathbf{U}_c\mathbf{D}_c\mathbf{P}\mathbf{c}_c$ and $\mathbf{M}\mathbf{U}_t\mathbf{D}_t\mathbf{P}\mathbf{c}_t = -\mathbf{U}_t\mathbf{D}_t\mathbf{P}\mathbf{c}_t$. For the types of low-frequency noise that are commonly observed in fMRI experiments, the covariance matrix of the noise is well approximated as $\mathbf{V} = \sigma^2\mathbf{G}\mathbf{G}^T$ (Liu and Wong, 2005) where σ^2 is assumed to be unknown. While the filtered subtraction approach presented here should be nearly optimal for block experimental designs that we address in this paper, alternate approaches that directly solve a version of Eq. (2) may be preferable for event-related designs, and are addressed in the Discussion section (Liu et al., 2002; Hernandez-Garcia et al., 2005).

To form a general linear model, we model the BOLD-weighted perfusion term $\mathbf{q}\circ\tilde{\mathbf{b}}$ as the sum of a constant term $\mathbf{q}_0$, which is proportional to baseline CBF, and a dynamic term $\mathbf{X}\mathbf{h}$ representing the time-varying portion of $\mathbf{q}\circ\tilde{\mathbf{b}}$, where $\mathbf{X}$ is a $N \times k$ design matrix and $\mathbf{h}$ is a $k \times 1$ vector of hemodynamic parameters. Note that this model does not make any assumptions about the magnitude of the BOLD-weighting in the $\mathbf{q}\circ\tilde{\mathbf{b}}$ term, so that the statistical tests described below will assess the significance of activation related changes in BOLD-weighted perfusion. However, given the short echo time acquisitions used in this paper (TE = 9.1 ms and 2.8 ms for the visual and hippocampal studies, respectively), the observed BOLD percent changes (0.3 to 0.7% at TE = 9.1 ms and 0.1 to 0.4% at TE = 2.8 ms) are relatively small compared to the observed functional perfusion changes of 40 to 100%. As a result, the statistical tests will primarily reflect changes in perfusion (Liu and Wong, 2005).

In the case of a block design, $\mathbf{X}$ reduces to a vector containing the smoothed stimulus pattern and $\mathbf{h}$ reduces to a scalar representing the unknown amplitude. The constant term $\mathbf{q}_0$ plus other low frequency confounds are modeled as $\mathbf{S}\mathbf{d}$, where $\mathbf{S}$ is a $N \times l$ matrix comprised of l nuisance model functions and $\mathbf{d}$ is an $l \times 1$ vector of nuisance parameters. Integrating these components into Eq. (3) yields the general linear model (GLM)

$$\hat{\mathbf{q}} = \mathbf{G}\mathbf{X}\mathbf{h} + \mathbf{S}\mathbf{d} + \mathbf{G}\mathbf{U}_c\mathbf{D}_c\mathbf{P}\mathbf{c}_c - \mathbf{G}\mathbf{U}_t\mathbf{D}_t\mathbf{P}\mathbf{c}_t + \mathbf{n} \quad (4)$$

Physiological noise reduction methods

Within the framework of the GLM, physiological noise is reduced by removing an estimate of the physiological noise terms $\mathbf{G}\mathbf{U}_c\mathbf{D}_c\mathbf{P}\mathbf{c}_c - \mathbf{G}\mathbf{U}_t\mathbf{D}_t\mathbf{P}\mathbf{c}_t$ from the perfusion time series $\hat{\mathbf{q}}$. This process can improve sensitivity by reducing the residual error that would otherwise be included in the additive noise term $\mathbf{n}$. The efficacy of the noise removal process depends on how well the noise is modeled. There are a number of plausible methods that correspond to different model assumptions about the structure of the physiological noise regressor matrix $\mathbf{P}$ and the regressor weights $\mathbf{c}_c$ and $\mathbf{c}_t$.

In this paper we consider three methods, where each method reflects an assumption about the primary effect of the physiological noise on the ASL time series and a corresponding assumption

about the regressor weights in the model. The methods are defined below and in Table 1.

Method 1

We assume that the primary effect of the noise is on the acquisition of the tag and control images, and that this effect does not differ between tag and control images. The corresponding assumption about the regressor weights is that $\mathbf{c}_c = \mathbf{c}_t = \mathbf{c}$ where $\mathbf{c}$ denotes the common value. This method may be considered a direct extension of the RETROICOR method to the ASL tag and control time series data.

Method 2

Here we assume that the relative weighting of the physiological noise differs between the tag and control images. For example, cardiac pulsations may affect how much blood is being tagged, and thus have a greater effect on the tag images than on the control images. The corresponding assumption about the regressor weights is that $\mathbf{c}_c \neq \mathbf{c}_t$.

Method 3

We assume that the dominant effect of physiological noise is due to the direct modulation of the perfusion time series, as opposed to a modulation of the image acquisition process. This assumption corresponds to a model of the form $\hat{\mathbf{q}} = \mathbf{GXh} + \mathbf{Sd} + \mathbf{GPc} + \mathbf{n}$, which is equivalent to the GLM in Eq. (4) with the constraint that $\mathbf{c}_t = -\mathbf{c}_c = -\mathbf{c}$. This method may be considered a direct extension of the RETROICOR method to the perfusion time series obtained from the filtered subtraction of the control and tag images.

To understand the relative performance of the different methods, we will also find it useful to consider a variation of Method 2 in which we assume that the regressor weights are the same for one component (e.g. for the respiratory component $\mathbf{c}_{t,R} = \mathbf{c}_{c,R} = \mathbf{c}_R$ with corresponding physiological noise matrix $\mathbf{P}_R$) but are different for the other component (e.g. for the cardiac component $\mathbf{c}_{t,C} \neq \mathbf{c}_{c,C}$ with corresponding physiological noise matrix $\mathbf{P}_C$) where the subscripts R and C denote respiratory and cardiac components, respectively. This variation is referred to as Method 2a and is further described in Table 1.

Statistical tests

To compare the performance of the different methods, we use two complementary metrics: (i) the multiple correlation coefficient and (ii) the statistical significance (P value) of an activation. The multiple correlation coefficient (defined below in Eq. (8)) is a measure of how well the model $\mathbf{Xh}$ fits the data $\hat{\mathbf{q}}$ after nuisance terms and physiological components have been projected out of both the model and the data, whereas the P value represents the probability of the data given that the null hypothesis of no functional activation is true. In this work, an F-statistic is used to compute the P value (see Eqs. (6) and (7) below). Methods with better performance will tend to yield higher correlation coefficients and lower P values. However, a method that achieves a higher correlation coefficient may result sometimes in a higher P value if there is an increase in the number of physiological regressors used. This is because an increase in the number of regressors reduces the degrees of freedom assigned to the additive noise component, leading to a decrease in both the F-statistic and the effective degrees of freedom of the F-distribution used in the computation of the P value. As an example, an increase in the correlation coefficient obtained with Method 2 as compared to Method 1 would reflect a better fit of the model to the data after physiological noise has been removed. However, a modest improvement in the model fit could be offset by the reduction in the degrees of freedom, resulting in an increase in the P value.

To compute the desired metrics, we note that the models corresponding to the different methods can be written in the form

$$\hat{\mathbf{q}} = \tilde{\mathbf{X}}\mathbf{h} + \mathbf{Za} + \mathbf{n} \tag{5}$$

where $\tilde{\mathbf{X}} = \mathbf{GX}$, $\mathbf{Z} = [\mathbf{S}\ \mathbf{P}_{TOT}]$, $\mathbf{a} = [\mathbf{d}^T\ \mathbf{c}_{TOT}^T]^T$, and $\mathbf{P}_{TOT}$ and $\mathbf{c}_{TOT}$ for the different methods and variations are defined in Table 1. The F-statistic is then defined as

$$F = \frac{Trace(\mathbf{RV})\hat{\mathbf{q}}^T\mathbf{R}_0\hat{\mathbf{q}}}{Trace(\mathbf{R}_0\mathbf{V})\hat{\mathbf{q}}^T\mathbf{R}\hat{\mathbf{q}}} \tag{6}$$

where $\mathbf{R}_0 = (\mathbf{P}_{\tilde{\mathbf{X}}\mathbf{Z}} - \mathbf{P}_{\mathbf{Z}})$ and $\mathbf{R} = (\mathbf{I} - \mathbf{P}_{\tilde{\mathbf{X}}\mathbf{Z}})$, and the projection matrices are defined as $\mathbf{P}_{\mathbf{Z}} = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T$ and $\mathbf{P}_{\tilde{\mathbf{X}}\mathbf{Z}} = \mathbf{Q}(\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T$ with $\mathbf{Q} = [\tilde{\mathbf{X}}\ \mathbf{Z}]$ (Worsley and Friston, 1995; Kiebel et al., 2003). Although not required, the use of projection matrices simplifies notation and also has a nice geometric interpretation as described in (Liu et al., 2001). Briefly, the projection matrix $\mathbf{R}_0$ is used to project the data $\hat{\mathbf{q}}$ onto the subspace that is both spanned by the columns of $\tilde{\mathbf{X}}$ and orthogonal to the columns of $\mathbf{Z}$, thus giving the portion of the data that can only be explained by $\tilde{\mathbf{X}}$ (i.e. changes in perfusion). Similarly, $\mathbf{R}$ projects the data onto the subspace that is orthogonal to the subspace spanned by $\mathbf{Q}$, thus yielding a residual noise term that represents the part of the data that cannot be explained by either changes in perfusion ($\tilde{\mathbf{X}}$) or physiological noise and other nuisance terms ($\mathbf{Z}$). Probability values are computed using an $F_{v_0,v}$ distribution, where the degrees of freedom taking into account the noise covariance are defined using a Satterthwaite approximation (Worsley and Friston, 1995; Kiebel et al., 2003) as

$$v_0 = \frac{Trace(\mathbf{R}_0\mathbf{V})^2}{Trace(\mathbf{R}_0\mathbf{V}\mathbf{R}_0\mathbf{V})};\ v = \frac{Trace(\mathbf{RV})^2}{Trace(\mathbf{RVRV})}, \tag{7}$$

where $\mathbf{V} = \sigma^2\mathbf{G}\mathbf{G}^T$, as previously noted below Eq. (3). We compute the multiple correlation coefficient (Seber and Lee, 2003) as follows:

$$R = \left(\frac{\hat{\mathbf{q}}^T\mathbf{R}_0\hat{\mathbf{q}}}{\hat{\mathbf{q}}(\mathbf{R}_0 + \mathbf{R})\hat{\mathbf{q}}}\right)^{1/2}. \tag{8}$$

Note that in contrast to the F-statistic, the terms in the multiple correlation coefficient are not normalized by terms $Trace(\mathbf{RV})$ and $Trace(\mathbf{R}_0\mathbf{V})$ that change with the number of physiological

Table 1
Assumptions and definitions of matrices for the different methods

Method	Assumptions	$\mathbf{P}_{TOT}$	$\mathbf{c}_{TOT}$
1	$\mathbf{c}_t = \mathbf{c}_c = \mathbf{c}$	$\mathbf{G}(\mathbf{U}_c\mathbf{D}_c - \mathbf{U}_t\mathbf{D}_t)\mathbf{P}$	$\mathbf{c}$
2	$\mathbf{c}_c \neq \mathbf{c}_t$	$[\mathbf{GU}_c\mathbf{D}_c\mathbf{P}\ -\mathbf{GU}_t\mathbf{D}_t\mathbf{P}]$	$[\mathbf{c}_c\ \mathbf{c}_t]$
2a	$\mathbf{c}_{c,R} = \mathbf{c}_{t,R}\ \mathbf{c}_{c,C} \neq \mathbf{c}_{t,C}$	$[\mathbf{G}(\mathbf{U}_c\mathbf{D}_c - \mathbf{U}_t\mathbf{D}_t)\mathbf{P}_R$ $\mathbf{GU}_{c,C}\mathbf{D}_{c,C}\mathbf{P}_C\ -\mathbf{GU}_{t,C}\mathbf{D}_{t,C}\mathbf{P}_C]$	$[\mathbf{c}_R\ \mathbf{c}_{c,C}\ \mathbf{c}_{t,C}]$
3	$\mathbf{c}_t = -\mathbf{c}_c = -\mathbf{c}$	$\mathbf{GP}$	$\mathbf{c}$

Method 2a is shown for the case of additional cardiac regressors. Swapping the R and C subscripts yields Method 2a with additional respiratory regressors. The subscript TOT (standing for total) is used to denote the physiological regressor matrix $\mathbf{P}_{TOT}$ and coefficient vector $\mathbf{c}_{TOT}$ that contain all the necessary elements for each method.

regressors. This is reflected in the following relation between the multiple correlation coefficient and the F-statistic

$$R = \left(\frac{F \cdot Trace(\mathbf{R}_0\mathbf{V})/Trace(\mathbf{RV})}{1 + F \cdot Trace(\mathbf{R}_0\mathbf{V})/Trace(\mathbf{RV})} \right)^{1/2}. \quad (9)$$

For brain regions such as the hippocampus where the intrinsic signal-to-noise ratio (SNR) is low, multiple runs can be analyzed together to improve the SNR. For these cases, Eq. (5) is expanded to allow for different regressors for each run. For example, for two experimental runs, the expanded GLM is

$$\begin{bmatrix} \hat{\mathbf{q}}_1 \\ \hat{\mathbf{q}}_2 \end{bmatrix} = \begin{vmatrix} \tilde{\mathbf{X}} \\ \tilde{\mathbf{X}} \end{vmatrix} \mathbf{h} + \begin{bmatrix} \mathbf{Z}_1 & 0 \\ 0 & \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} + \mathbf{n}, \quad (10)$$

where the subscripts 1 and 2 correspond to the first and second runs, respectively.

Methods

Imaging system

All experimental imaging data were collected on a GE Signa Excite 3 Tesla whole body system with a body transmit coil and an eight channel receive coil.

Visual cortex studies

Six healthy subjects (5 male; age 31 ± 6 years old) participated in the study after giving informed consent. During each imaging session, the subject was presented with a visual stimulus consisting of "on" periods of a full-field, full-contrast radial 8 Hz flickering checkerboard with a small white fixation square at the center of the screen and "off" periods of a black background with white fixation square. Each imaging section consisted of four runs of a block design paradigm with a 30s initial off period followed by four cycles of 20s on and 40s off. Three oblique 8 mm slices through the calcarine sulcus were imaged in a sequential fashion from inferior to superior. Functional ASL data were acquired using a PICORE-QUIPSS II tagging scheme (Wong et al., 1998) with a dual echo spiral readout (TE1 = 9 ms, TE2 = 30 ms), interleaving of tag and control images, TR = 2 s, number of repetitions = 130, TI1/TI2 = 700/1400 ms, flip angle = 90, FOV 24 cm, and 64 × 64 matrix. Small bipolar crusher gradients were included to remove signal from large vessels (b = 2 s/mm^2).

Hippocampal studies

Seven healthy subjects (6 female; age 33 ± 17 years) participated in the study after giving informed consent. During each imaging session, the subject was presented with a memory encoding task in which a series of familiar and novel landscape scenes were viewed (Stern et al., 1996). At the start of each session, the subject viewed 4 landscape images (2 with vertical and 2 with horizontal aspect ratio), which served as the familiar images for the experiment. Each experimental run used a block design paradigm, consisting of 5 cycles of 10 familiar and 10 novel images. Images were shown for 2 seconds with a 0.5 seconds gap between images. The sequence of familiar images was changed between cycles, and no familiar image occurred twice in a row. To maintain attention, the subject was asked to decide whether each image had a horizontal or vertical aspect ratio and to indicate their response with a two-button response box. A total of three experimental runs were performed. Two runs from subject 2 and one run from each of subjects 4 through 7 were excluded from analysis due to excessive subject motion. Five oblique 6 mm thick slices aligned with the hippocampus were imaged in a sequential fashion from inferior to superior. Functional ASL data were acquired using a PICORE-QUIPSS II sequence (Wong et al., 1998) with a dual echo spiral readout (TE1 = 2.4 ms, TE2 = 24 ms), interleaving of tag and control images, and TR = 2 s, number of repetitions = 125, TI1/TI2 = 800/1400 ms, flip angle = 90, FOV 24 cm, and 64 × 64 matrix. High resolution structural scans were acquired with a magnetization prepared 3D fast spoiled gradient acquisition in the steady state (FSPGR) sequence (TI 450 ms, TR 7.9 ms, TE 3.1 ms, 12 degree flip angle, FOV 25 × 25 × 16 cm, matrix 256 × 256 × 124).

Physiological data

Cardiac pulse and respiratory effort data were monitored using a pulse oximeter (InVivo) and a respiratory effort transducer (BIOPAC), respectively. The pulse oximeter was placed on the subject's right index finger. The respiratory effort belt was placed around the subject's abdomen. Physiological data were sampled at 40 samples per second using a multi-channel data acquisition board (National Instruments). In addition to the physiological data, scanner TTL pulse data (10 ms duration, 5 volt pulse per slice acquisition) were recorded at 1 kHz. The TTL pulse data were used to synchronize the physiological data to the acquired images. Cardiac, respiratory and TTL data were used to calculate the physiological noise regressors in **P**. A detailed description is given in the appendix.

Data analysis

All images were co-registered using AFNI software (Cox, 1996). For the hippocampal data sets, the hi-resolution anatomical scans were used to delineate regions of interest (ROI) encompassing the hippocampus and the parahippocampal gyrus.

Probability values (P values), F-statistics (F), and multiple correlation coefficients (R) were calculated on a per-voxel basis using the different methods described in the theory section. For the hippocampal data sets, only voxels within the anatomical ROI were analyzed. Estimates $\hat{\mathbf{q}}$ of the perfusion time series were obtained from the first echo of the ASL time series using a lowpass filtering matrix **G** (see Eq. (3)), where **G** is the matrix form of $g[n] = [1\ 2\ 1]/2$ (Liu et al., 2002). The first four images in each time series were not included in the analysis. The design matrix **X** in Eq. (4) was assumed to be a vector obtained from convolving the block design stimulus pattern with a gamma density function of the form $h(t) = (\tau\, n!)^{-1}((t-\Delta t)/\tau)^n \exp(-(t-\Delta t)/\tau)$ for $t \geq \Delta t$ and 0 otherwise, with $\tau = 1.2$, $n = 3$, and $\Delta t = 1$. A constant and a linear trend term were included as terms in the nuisance matrix **S**.

For each subject and experimental run, the number of activated voxels obtained with each method was calculated using either a threshold on the multiple correlation coefficients (R > 0.54 and R > 0.35 for visual cortex and hippocampus, respectively) or on the P values ($P < 0.00001$ and $P < 0.02$ for visual cortex and hippocampus, respectively). These thresholds were chosen to yield approximately the same number of activated voxels in the two brain regions. The multiple correlation coefficient thresholds were obtained from the P value thresholds using the number of degrees of freedom for the uncorrected data. Differences between methods in

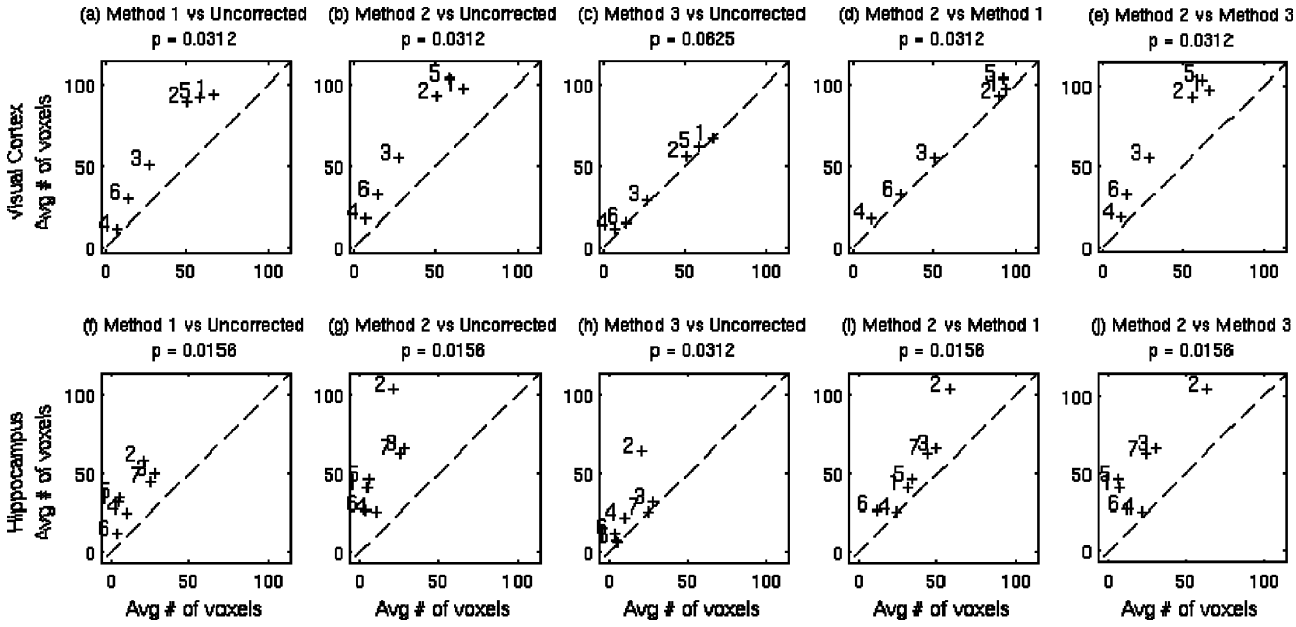

Fig. 1. Comparisons of the average number of activated voxels obtained with the various methods using correlation coefficient thresholds of 0.54 and 0.35 for the visual cortex and hippocampal region, respectively. The upper row (a–e) shows data from the visual cortex, while the bottom row (f–j) shows data from the hippocampal region. The title of each plot indicates the methods that are compared (e.g. (a) compares Method 1 (y-axis) to Uncorrected (x-axis)). The P value indicating the significance of the difference between methods as assessed with a Wilcoxon signed rank test is also shown in each title. Subject numbers are shown for each data point. The dashed line indicates the line of unity.

the average number of activated voxels were assessed with a Wilcoxon signed rank test.

To compare the methods on a per subject basis, we defined a functional region of interest (ROI) for each subject using either the multiple correlation or P value thresholds described above. The ROI consisted of the union of all voxels that passed the threshold using either no correction or the application of method 1 or method 2. Comparisons were performed with both paired and unpaired two-tailed t-tests on either the multiple correlation coefficients or the logarithms of the P values across all voxels within the respective functional ROIs. The unpaired t-tests assess the significance of differences in the sample means, which is related to the difference in the number of activated voxels. The paired t-tests assess whether per-voxel differences across the ROI are

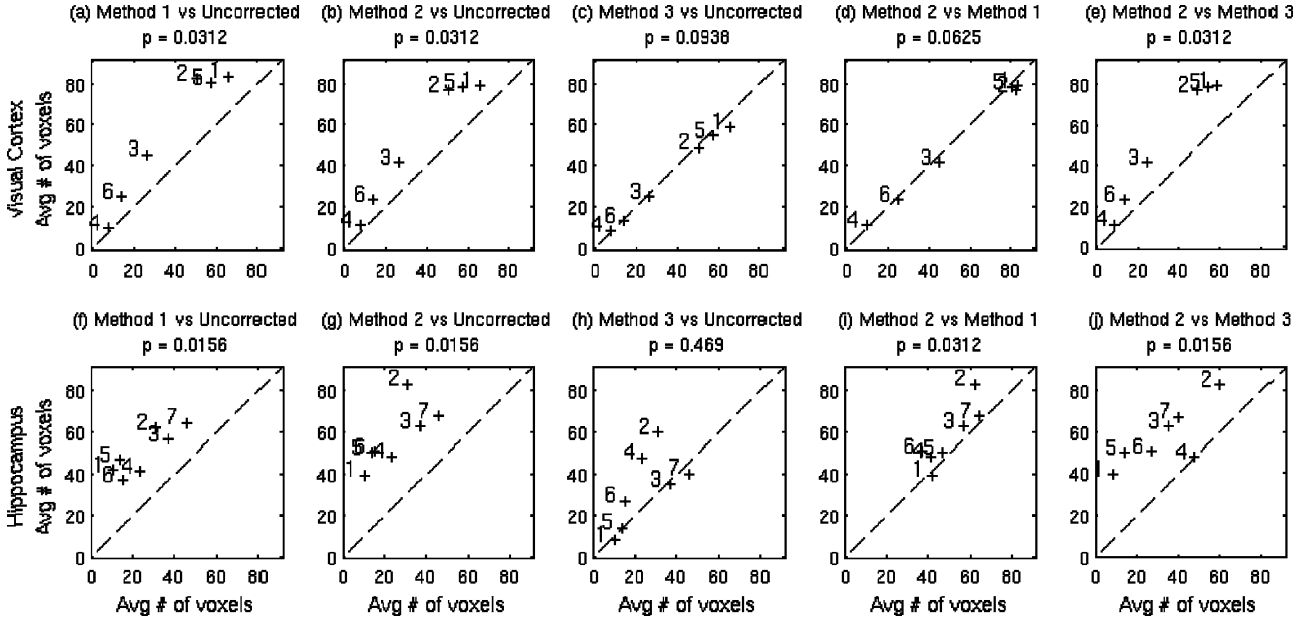

Fig. 2. Comparisons of the average number of activated voxels obtained with the various methods using P value thresholds of 0.0001 and 0.02 for the visual cortex and hippocampal region, respectively. The upper row (a–e) shows data from the visual cortex, while the bottom row (f–j) shows data from the hippocampal region. The correction methods compared in each case (see text for definitions) and the statistical significance of the difference between methods using a Wilcoxon signed rank test are shown above each plot. Subject numbers are shown for each data point. The dashes indicate the line of unity.

significantly different than zero. As shown in the Results, per-voxel differences can be significantly different than zero without a significant difference in the sample means.

To assess the relative contributions of cardiac and respiratory fluctuations, we divided the energy of either the cardiac or respiratory model components by the sum of the energies of the cardiac, respiratory, and stimulus-related model components. For example, the fractional energy of the cardiac regressors is given by $\|\mathbf{P}_{TOT,C}\hat{\mathbf{c}}_{TOT,C}\|^2/(\|\tilde{\mathbf{X}}\hat{\mathbf{h}}\|^2 + \|\mathbf{P}_{TOT,C}\hat{\mathbf{c}}_{TOT,C}\|^2 + \|\mathbf{P}_{TOT,R}\hat{\mathbf{c}}_{TOT,R}\|^2)$ where the subscripts R and C denotes the respiratory and cardiac regressors, respectively. For each subject, we used a paired t-test (two-tailed) over all voxels in the functional ROI to compare the fractional energies of the cardiac and respiratory regressors.

Results

As shown in Figs. 1 and 2, for both visual cortex and the hippocampal region, the use of methods 1 and 2 resulted in a significant ($P < 0.04$) improvement in the number of activated voxels as compared to no correction. Examples of the improvement obtained with Method 2 are shown in Figs. 3 and 4 for a representative subject from each study. For the hippocampal study, the activation map was obtained using the multi-run model presented in Eq. (10). Increases in the number of activated voxels and the reduced noise in the time series after correction are evident for both studies.

For method 3, a significant increase ($P < 0.04$) as compared to no correction was observed in the number of activated voxels based on a correlation threshold in hippocampus, but other comparisons did not indicate significant differences ($P > 0.05$). In addition, Method 2 was found to yield significantly more activated voxels ($P < 0.05$) than Method 3 for the two brain regions. Because of its lower performance, the performance of Method 3 is not further addressed in this section.

In the visual cortex, Method 2 yielded significantly more ($P < 0.04$) voxels than Method 1 with a correlation coefficient threshold, but no significant difference was found ($P > 0.06$) for voxels with a P value threshold. The significance of the increase in activated voxels with the P value threshold for Method 1 was highly dependent on the threshold used for selecting activated voxels. At higher thresholds (e.g. $P < 0.001$), the difference in the visual cortex was not significant ($P > 0.05$). For the hippocampal region, Method 2 yielded significantly more ($P < 0.04$) voxels than

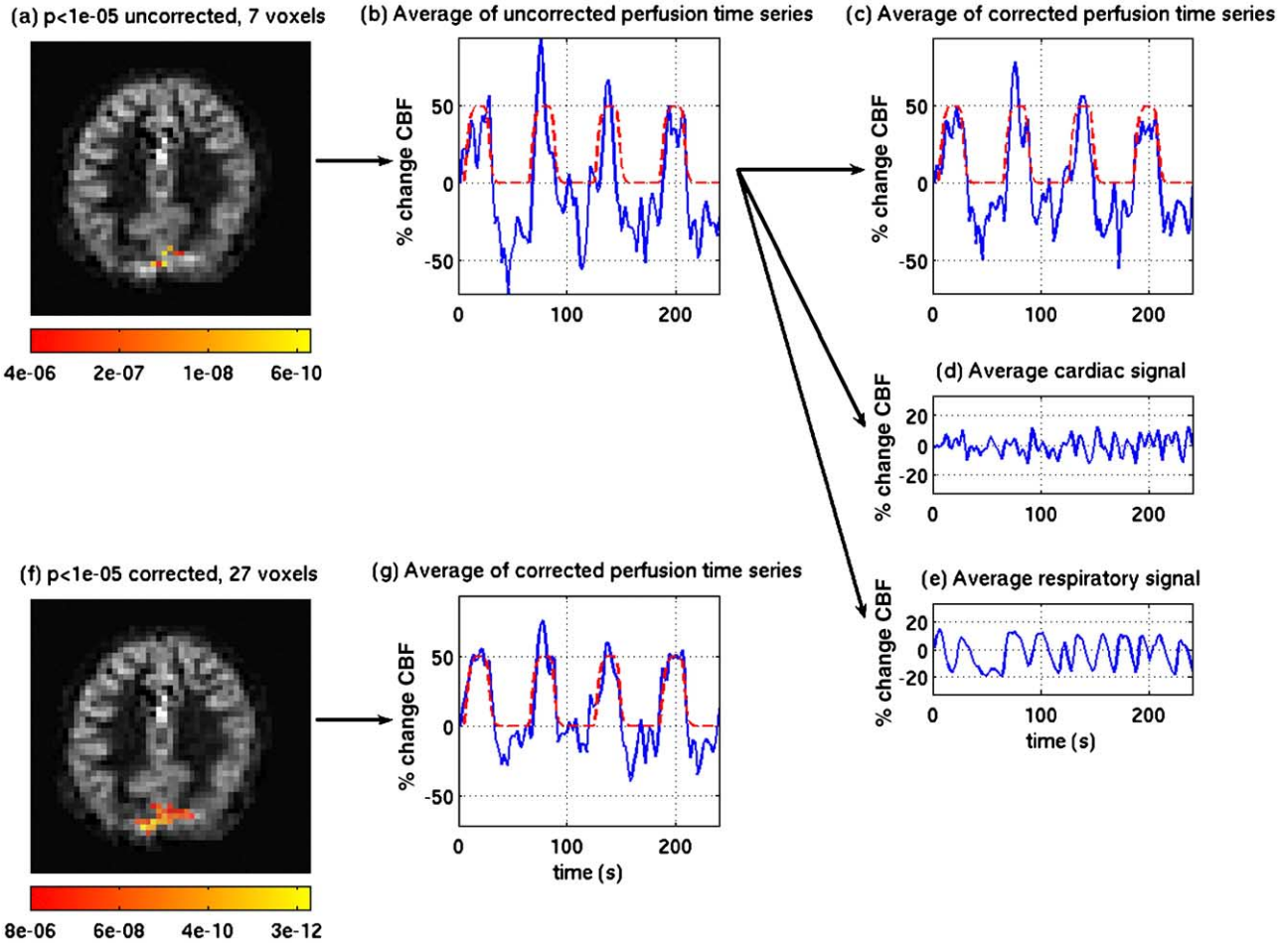

Fig. 3. Example single subject data from visual cortex. Statistical significance maps (thresholded at $P < 0.00001$ with nearest neighbor clustering and no smoothing) showing voxels with significant activation both prior to panel (a) and after correction using Method 2 (panel (f)). These maps are overlaid on mean CBF images. Uncorrected and corrected average CBF time courses corresponding to the activated region in panel (a) are shown in panels (b) and (c), respectively, with the cardiac and respiratory components shown in panels (d) and (e), respectively. The average corrected CBF time course corresponding to the activated region in panel (f) is shown in panel (g).

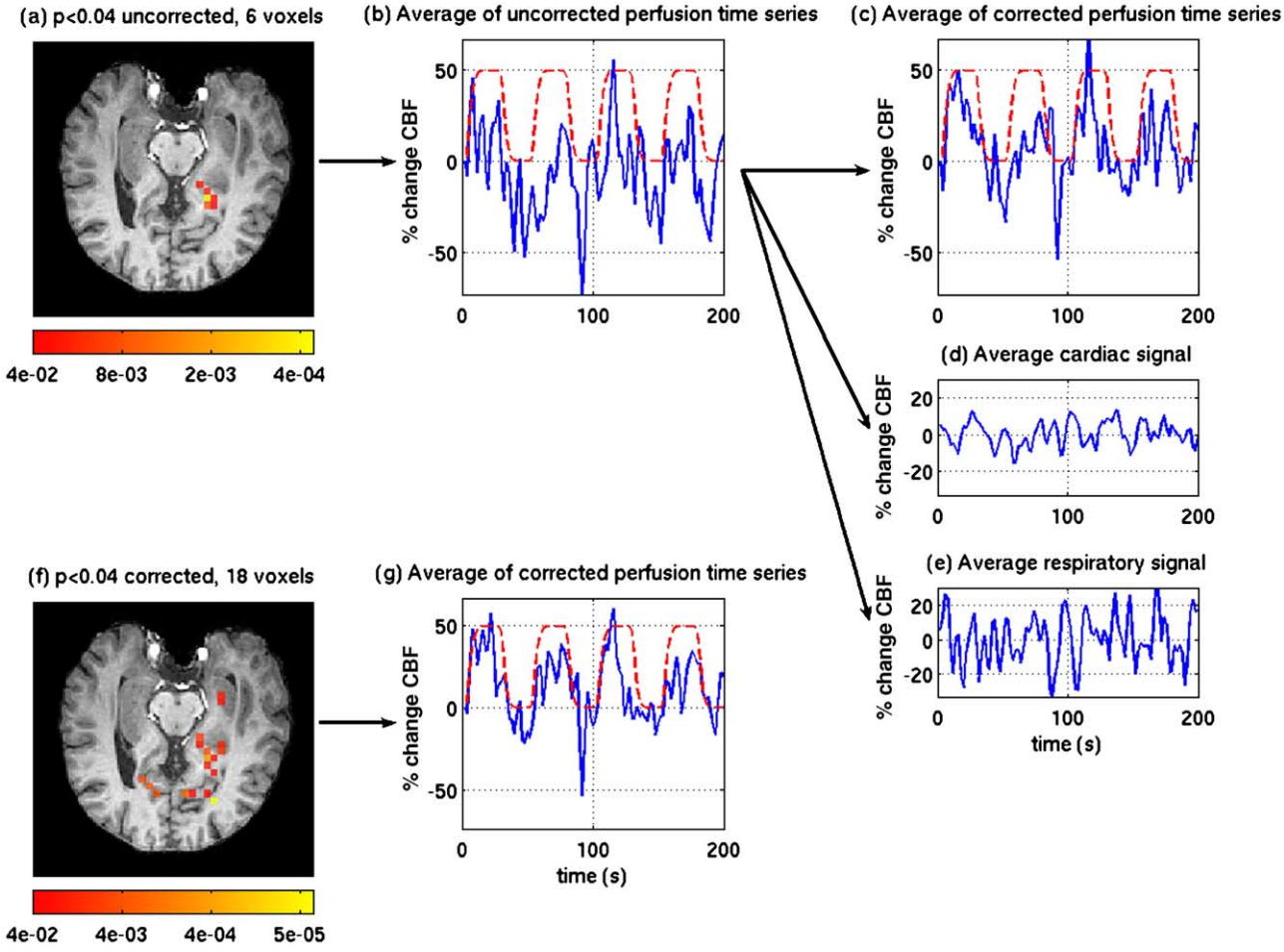

Fig. 4. Example single subject data from hippocampus data. Statistical significance maps (thresholded $P < 0.04$ with nearest neighbor clustering and no smoothing) showing voxels with significant activation both prior to panel (a) and after correction using Method 2 (panel (f)). These maps are overlaid on high resolution anatomical images. Uncorrected and corrected average CBF time courses corresponding to the activated region in panel (a) are shown in panels (b) and (c), respectively, with the cardiac and respiratory components shown in panels (d) and (e), respectively. The average corrected CBF time course corresponding to the activated region in panel (f) is shown in panel (g).

Method 1 using either the correlation coefficient threshold or the P value threshold. The improved performance of Method 2 in the hippocampal region is due primarily to the inclusion of the cardiac regressors. This is reflected in the observation (data not shown) that Method 2a with additional cardiac regressors yielded significantly more activated voxels than Method 1 ($P < 0.04$ for correlation and P value thresholds), while Method 2a with additional respiratory regressors did not yield significantly more activated voxels ($P > 0.2$ and $P > 0.8$ for correlation and P value thresholds, respectively).

The means and standard deviations of the multiple correlation coefficients and logarithmic P values are summarized in Tables 2–5. Based on paired comparisons, Methods 1 and 2 resulted in a significant ($P < 0.0001$) improvement in statistical significance as compared to no correction for all subjects and brain regions. With unpaired comparisons, Methods 1 and 2 provided a significant

Table 2
Mean and standard deviation of multiple correlation coefficients for voxels within functional ROI (union of voxels with R > 0.54) in visual cortex

	Multiple correlation coefficients			Comparison of Methods 1 and 2		N
	Uncorrected	Method 1	Method 2	Paired	Unpaired	
1	0.57 ± 0.14	0.64 ± 0.09	0.65 ± 0.08	0.28	0.75	421
2	0.55 ± 0.14	0.65 ± 0.09	0.66 ± 0.09	1×10^{-14}	0.14	387
3	0.52 ± 0.12	0.61 ± 0.07	0.62 ± 0.07	1×10^{-12}	0.18	229
4	0.53 ± 0.08	0.57 ± 0.07	0.60 ± 0.06	4×10^{-11}	0.001	75
5	0.55 ± 0.13	0.63 ± 0.09	0.64 ± 0.09	2×10^{-27}	0.04	433
6	0.48 ± 0.15	0.60 ± 0.06	0.61 ± 0.05	2×10^{-9}	0.15	140

Statistical comparisons were performed with both two-tailed paired and unpaired t-tests across voxels within the ROI, where the number of samples N denotes the number of voxels times the number of experimental runs. The significance values of the comparisons between Methods 1 and 2 are shown. For all subjects, the multiple correlation coefficients for Methods 1 and 2 were significantly higher than those for uncorrected data ($P < 1 \times 10^{-11}$).

Table 3
Mean and standard deviation of multiple correlation coefficients for voxels within functional ROI (union of voxels with R > 0.35) in hippocampal region

	Multiple correlation coefficients			Comparison of Methods 1 and 2		N
	Uncorrected	Method 1	Method 2	Paired	Unpaired	
1	0.25 ± 0.07	0.36 ± 0.05	0.39 ± 0.04	3×10^{-8}	1×10^{-4}	139
2	0.29 ± 0.07	0.34 ± 0.07	0.40 ± 0.06	6×10^{-16}	2×10^{-11}	115
3	0.32 ± 0.08	0.38 ± 0.06	0.40 ± 0.05	3×10^{-19}	3×10^{-6}	215
4	0.24 ± 0.11	0.36 ± 0.07	0.35 ± 0.08	0.26	0.30	83
5	0.23 ± 0.10	0.37 ± 0.05	0.39 ± 0.04	2×10^{-6}	0.002	101
6	0.29 ± 0.05	0.34 ± 0.05	0.39 ± 0.03	6×10^{-14}	2×10^{-9}	54
7	0.28 ± 0.10	0.32 ± 0.10	0.37 ± 0.11	4×10^{-5}	1×10^{-5}	182

Statistical comparisons were performed with both two-tailed paired and unpaired *t*-tests across voxels within the ROI, where the number of samples *N* denotes the number of voxels times the number of experimental runs. The significance values of the comparisons between Methods 1 and 2 are shown. For all subjects, the multiple correlation coefficients for Methods 1 and 2 were significantly higher than those for uncorrected data ($P < 1 \times 10^{-9}$).

improvement ($P < 0.005$) in all subjects, except for subject 4 in the visual cortex study. As an example of the improvement obtained, Figs. 5 and 6 present comparisons of the per voxel *P* values obtained with Method 2 versus the *P* values obtained with no correction, for each subject in the visual and hippocampal study, respectively.

As shown in Table 2 for visual cortex, paired tests of the correlation coefficients showed that Method 2 yielded significantly higher correlation coefficients ($P < 10^{-5}$) than Method 1 for 5 out of 6 subjects. With unpaired tests, only 2 out of 6 subjects showed a significant difference ($P < 0.05$), with Method 2 providing higher correlation coefficients. In contrast, with the paired comparisons of the logarithmic *P* values summarized in Table 4, Method 1 obtained significantly ($P < 10^{-6}$) lower *P* values than Method 2 for 5 out of 6 subjects in the visual cortex study, while *P* values obtained with Method 2 were significantly lower for the remaining subject ($P < 0.02$). However, only subject 1 showed a significant difference at $P < 0.05$ when using an unpaired comparison. This observation is consistent with the observation noted above that the difference in the number of activated voxels for the two methods in visual cortex is just significant at the $P < 0.0001$ threshold, and is not significant for higher thresholds.

As shown in Table 3 for the hippocampal study, both paired and unpaired comparisons of the correlation coefficients showed significant differences ($P < 10^{-4}$ and $P < 0.003$ for paired and unpaired, respectively) for 6 out of 7 subjects, with higher correlation coefficients for Method 2. For the comparisons of *P* values shown in Table 5, paired tests indicated that Method 2 achieved significantly lower ($P < 0.01$) *P* values than Method 1 for 3 out of 7 subjects, with no significant difference in *P* values for the remaining 4 subjects. With unpaired comparisons, 2 of the 7 subjects showed significant differences ($P < 0.005$).

The cardiac and respiratory fractional energies averaged across all voxels in the functional ROI are listed in Table 6, along with the results of paired t-tests. In the visual cortex, the fractional energy of the respiratory component was significantly higher ($P < 0.03$) than the energy of the cardiac component for all subjects. In the hippocampal region, the energy of the respiratory components was higher ($P < 0.03$) in 5 out of 7 subjects, but significantly lower ($P < 0.005$) in 2 out of 7 subjects. While there was no significant difference ($P > 0.15$) between the visual cortex and the hippocampal region in the mean fractional energies of the respiratory component, the mean fractional energy of the cardiac component was significantly higher ($P < 0.001$) in the hippocampal region as compared to the visual cortex.

Maps of the cardiac and respiratory fractional energies for both the visual cortex and hippocampus (one subject from each study) are shown in Fig. 7. Consistent with previous studies examining physiological noise in BOLD data (Dagli et al., 1999; Glover et al., 2000), the cardiac components are primarily localized around sulci and vessels (such as the superior sagittal sinus), whereas respiratory components are more diffusely distributed across the brain. Overall, the cardiac fluctuations are more pronounced in the hippocampal data as compared to the visual data. The reduction in cardiac fluctuations around the superior sagittal sinus in the visual data most likely reflects the use of small diffusion gradients, which

Table 4
Means and standard deviations of the logarithms of *P* values for voxels within functional ROI (union of voxels with $P < 0.00001$) in visual cortex

	Logarithmic *P* values			Comparison of Methods 1 and 2		N
	Uncorrected	Method 1	Method 2	Paired	Unpaired	
1	−6.63 ± 3.33	−7.72 ± 2.55	−7.24 ± 2.38	1.4×10^{-21}	0.007	377
2	−6.14 ± 3.08	−8.07 ± 2.75	−7.87 ± 2.70	5.1×10^{-9}	0.347	343
3	−5.22 ± 2.26	−6.89 ± 1.95	−6.63 ± 1.84	1.2×10^{-14}	0.186	193
4	−5.54 ± 1.72	−6.02 ± 1.77	−6.35 ± 1.68	0.014	0.336	50
5	−6.43 ± 3.05	−7.78 ± 2.88	−7.61 ± 2.91	5.4×10^{-7}	0.457	353
6	−4.84 ± 2.35	−6.35 ± 1.34	−6.08 ± 1.29	4.9×10^{-7}	0.121	111

Statistical comparisons were performed with both two-tailed paired and unpaired *t*-tests across voxels within the ROI, where the number of samples *N* denotes the number of voxels times the number of experimental runs. The significance values of the comparisons between Methods 1 and 2 are shown. Log *P* values for both Methods 1 and 2 were significantly lower than for uncorrected data on the basis of paired comparisons ($P < 10^{-4}$) and unpaired comparisons ($P < 0.005$) for all subjects except subject 4 for whom the unpaired comparison with Method 1 was not significant ($P = 0.17$) and the unpaired comparison with Method 2 was significant at $P = 0.02$.

Table 5
Means and standard deviations of the logarithms of P values for voxels within functional ROI (union of voxels with $P < 0.02$) in hippocampal region

	Logarithmic P values			Comparison of Methods 1 and 2		N
	Uncorrected	Method 1	Method 2	Paired	Unpaired	
1	−1.13 ± 0.56	−1.91 ± 0.39	−1.93 ± 0.44	0.46	0.66	162
2	−1.45 ± 0.56	−1.79 ± 0.51	−2.03 ± 0.55	2.1×10^{-5}	0.001	105
3	−1.65 ± 0.63	−2.00 ± 0.51	−2.06 ± 0.52	0.006	0.21	231
4	−1.25 ± 0.67	−1.78 ± 0.55	−1.76 ± 0.59	0.71	0.75	154
5	−1.23 ± 0.67	−1.94 ± 0.50	−1.97 ± 0.41	0.48	0.67	129
6	−1.43 ± 0.43	−1.81 ± 0.44	−1.95 ± 0.36	4.0×10^{-5}	0.005	121
7	−1.40 ± 0.71	−1.66 ± 0.66	−1.73 ± 0.79	0.27	0.25	233

Statistical comparisons were performed with both two-tailed paired and unpaired t-tests across voxels within the ROI, where the number of samples N denotes the number of voxels times the number of experimental runs. The significance values of the comparisons between Methods 1 and 2 are shown. Log P values for both Methods 1 and 2 were significantly lower than for uncorrected data on the basis of paired comparisons ($P < 10^{-5}$) and unpaired comparisons ($P < 10^{-4}$) for all subjects.

were not applied in the hippocampal experiments. The large cardiac fluctuations in the anterior portion of the hippocampal slices are similar to those observed by (Dagli et al., 1999) and most likely reflect the presence of arteries and large arterioles in this region of the brain, which is close to the Circle of Willis.

Discussion

We have examined three extensions of a retrospective image based correction method (RETROICOR) for the reduction of physiological noise in ASL fMRI studies. The proposed methods are derived from a general linear model of the estimated perfusion time series. Method 1 assumes that physiological noise affects the tag and control images equally, while Method 2 does not impose this constraint. In contrast, Method 3 assumes that the dominant effect of the physiological noise occurs through modulation of cerebral blood flow as opposed to contamination of the measured tag and control images.

We found that Methods 1 and 2 provided significant improvements in performance for both brain regions examined, as compared to both Method 3 and no correction. The relatively poor performance of Method 3 appears to suggest that the effect of physiological noise on the acquisition of the tag and control images is greater than the effect on perfusion. It is possible that the performance of method 3 could be improved with the use of measurements of end tidal carbon dioxide ($ETCO_2$), since it has been shown that fluctuations in arterial velocities and the BOLD signal show a correlation with $ETCO_2$ (Wise et al., 2004). For example, a hybrid approach that utilizes Method 2 for the respiratory and cardiac regressors and Method 3 for $ETCO_2$ regressors may provide gains in performance.

Method 2 was significantly better than Method 1 for functional studies of the hippocampal region. The difference in the methods was not as pronounced in the visual cortex, where Method 2 performed better than Method 1 on the basis of correlation coefficients but was not significantly different on the basis of P values. Although the use of P values is preferred when a more

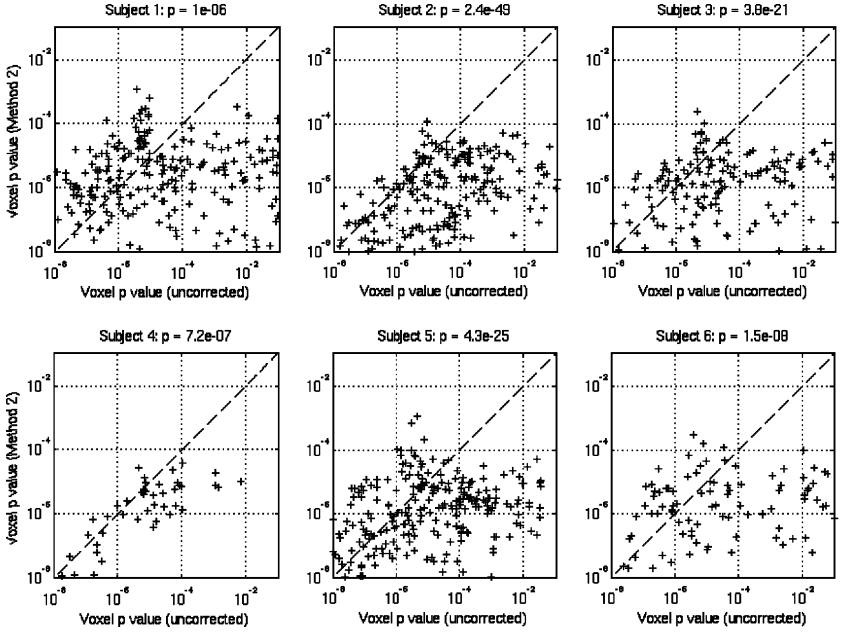

Fig. 5. Comparison of voxel-wise P values prior to (x-axis) and after correction (y-axis) using Method 2 for visual cortex data in each of 6 subjects. The statistical significance (two-tailed paired t-test) of the difference in P values is indicated above each plot.

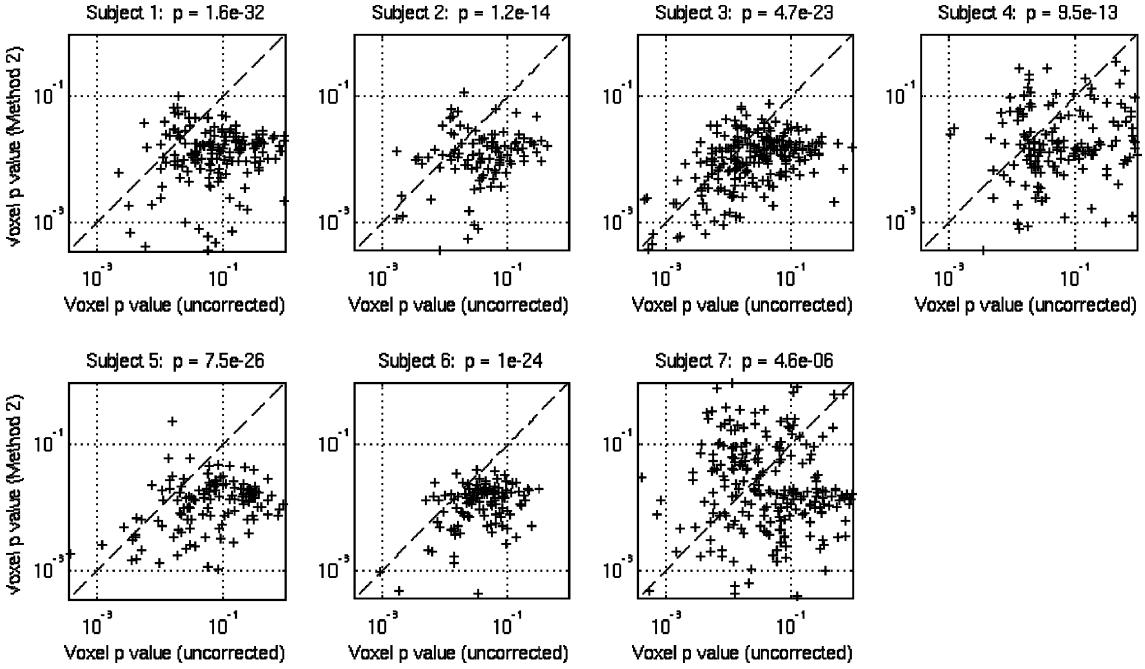

Fig. 6. Comparison of voxel-wise P values prior to (x-axis) and after correction (y-axis) using Method 2 for hippocampal data in each of 7 subjects. The statistical significance (two-tailed paired t-test) of the difference in P values is indicated above each plot.

rigorous metric of statistical significance is desired, we have included correlation coefficients in our presentation to facilitate a qualitative comparison with a number of previous investigations that have used correlation coefficients to assess the impact of physiological noise reduction in BOLD experiments (Hu et al., 1995; Chuang and Chen, 2001; Pfeuffer et al., 2002b).

The quantitative comparison of the two methods is consistent with the following qualitative conclusion that has emerged from our hands-on experience with the data: Method 2 should be used for hippocampal studies, while the difference between Methods 1 and 2 for the visual cortex is relatively small so that either method can be used with confidence. The differences in the performances of the two methods may reflect the significant increase in the fractional energy of cardiac fluctuations in the hippocampal region as compared to the visual cortex. As mentioned in the Results section, the increased cardiac fluctuations in the hippocampal region most likely reflect the greater presence of large vessels in this brain area. It has been shown that the amount of blood that is tagged and delivered to the imaging region depends on the temporal position of the tagging pulse relative to the cardiac cycle (Wu et al., 2005). As a result, cardiac fluctuations are likely to affect tag images differently than control images. Method 2 models this effect by allowing for different regressor weights for the tag and control images.

Table 6
Fractional energies of the respiratory and cardiac components for (a) visual cortex and (b) hippocampal region

Subject #	Respiratory	Cardiac
(a) Visual cortex		
1	0.30	0.076[a]
2	0.30	0.085[a]
3	0.22	0.165[a]
4	0.20	0.132[a]
5	0.28	0.127[a]
6	0.40	0.115[a]
(b) Hippocampal region		
1	0.41	0.33[a]
2	0.36	0.20[a]
3	0.18	0.32[b]
4	0.30	0.40[b]
5	0.47	0.26[a]
6	0.36	0.20[a]
7	0.41	0.25[a]

[a] Respiratory and cardiac fractional energies are significantly different ($P < 0.03$) with larger mean respiratory fractional energy.
[b] Respiratory and cardiac fractional energies are significantly different ($P < 0.005$) with larger mean cardiac fractional energy.

In presenting the results of the reduction of physiological noise in ASL fMRI studies, there is an implicit assumption that the lowering of P values with the application of the algorithm indicates an improvement in statistical performance. In general, the removal of deterministic physiological components should tend to make the residuals more Gaussian, leading to a more accurate model and less biased statistics. In some cases, however, it is possible that the more accurate model may yield higher P values than a less accurate model (e.g. with non-Gaussian residuals). While our experimental results (e.g. Figs. 5 and 6) suggest that this is not a prevalent effect, further studies using an analysis in which the Gaussian assumption is relaxed (e.g. a Bayesian analysis with sampling of the posterior distribution (Woolrich et al., 2004)) could provide a more accurate understanding of the performance of the proposed methods.

In this paper, we have focused on extensions of a retrospective image-based noise reduction method (RETROICOR) that has been previously applied to BOLD fMRI studies. The proposed methods can be readily combined with navigator-based methods (Pfeuffer et al., 2002b), and an investigation of the performance of the combination of methods would be useful. Similarly, retrospective k-space methods (Hu et al., 1995) are also likely to provide some

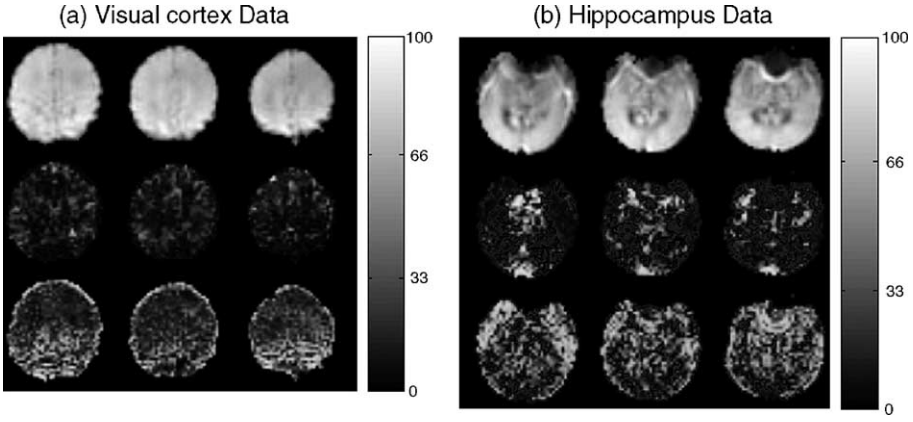

Fig. 7. Maps of fractional cardiac and respiratory energies in the visual cortex (a; subject 1) and the hippocampal region (b; Subject 1). Low-resolution anatomical images are shown in the top row. The middle and bottom rows show cardiac and respiratory energy maps (scaled between 0 and 100%), respectively.

degree of noise reduction for ASL experiments. Both the navigator-based and k-space methods would most likely be used as pre-processing steps, since their integration into the framework of a general linear model does not appear to be straightforward.

An alternative approach to reducing physiological noise in ASL measurements is the use of background suppression methods that attenuate the static tissue (Ye et al., 2000; Kemeny et al., 2005). While these methods have been shown to reduce the standard deviation of ASL images (Ye et al., 2000), one study has reported that the gains achieved for functional ASL studies are slight but not significant (St Lawrence et al., 2005). This is in contrast to the significant gains presented here with image-based correction methods. However, in a preliminary study focused on visual cortex, we have found that background suppression offers gains similar to those achieved with the methods presented here. In addition, the combination of background suppression and image-based correction methods provides significantly better performance than either method alone. Further study into the optimal application of background suppression and image-based correction methods is therefore likely to lead to significant improvements in the sensitivity of functional ASL experiments.

The noise correction methods examined in this paper have been based on a general linear model (Eq. (4)) of the perfusion signal, which is estimated using a filtered subtraction of the tag and control images. The filtered subtraction approach is standard in the literature and introduces relatively minimal bias for the block designs that we have used in this paper (Aguirre et al., 2002; Wang et al., 2003a; Liu and Wong, 2005). However, for event-related designs, the filter subtraction approach leads to a significant filtering of the estimated impulse response (Liu et al., 2002; Liu and Wong, 2005) and is also less statistically efficient than methods based on the unfiltered model in Eq. (2) (Hernandez-Garcia et al., 2005). The extension of the methods presented here to the unfiltered model should therefore prove to be useful for event-related ASL experiments.

In conclusion, we have shown that the application of image-based physiological noise reduction methods can significantly improve the sensitivity of functional ASL experiments. This improvement in sensitivity will be particularly important for cognitive fMRI studies, where activations tend to be less robust than those found in sensory studies and where the use of event-related paradigms can result in reduced detection power.

Acknowledgments

This work was supported by a Clinical Hypotheses in Imaging Grant from the Dana Foundation and a Biomedical Engineering Research Grant from the Whitaker Foundation. We would like to thank Katie Bangen, Duke Han, and Mark Bondi for their assistance with the hippocampal experiments.

Appendix A

Each row of the matrix **P** representing physiological regressors in Eq. (2) is defined as $[\cos(C_n)\ \cos(2C_n)\ \sin(C_n)\ \sin(2C_n)\ \cos(R_n)\ \cos(2R_n)\ \sin(R_n)\ \sin(2R_n)]$ where $C_n = \varphi_c[n]$ is the cardiac phase, $R_n = \varphi_r[n]$ is the respiratory phase, and n indexes the image data acquired at time $t = nTR$. The columns of **P** thus form the terms of a 2nd order Fourier series expansion in terms of cardiac phase (columns 1 through 4) and a 2nd order Fourier series expansion in terms of respiratory phase (columns 5 through 8).

For each image, cardiac phase is defined as $\varphi_c[n] = 2\pi\frac{t - t_1}{t_2 - t_1}$, where t is the time at which the image is acquired, t_1 is the time of the cardiac peak immediately prior to t, and t_2 is the time of the cardiac peak immediately after t (Hu et al., 1995). Following (Glover et al., 2000), respiratory phase is defined as

$$\varphi_r[n] = \pi \frac{\sum_{b=1}^{rnd\left(100\cdot\frac{R[t]}{R\max}\right)} H[b]}{\sum_{b=1}^{100} H[b]} \operatorname{sgn}(dR/dt)$$

where *rnd* denotes an integer rounding operation, $R[t]$ is the amplitude of the signal from the respiratory belt normalized from 0 to $R_{\max}$, where $R_{\max}$ is the maximum of $R[t]$, and $H[b]$ denotes the histogram of the number of occurrences of respiratory amplitude values that occur at bin value b. Bin values span $0.01R_{\max}$ thru $R_{\max}$ with intervals of $0.01R_{\max}$. The term $\operatorname{sgn}(dR/dt)$ is the sign of dR/dt, with a value of 1 during inspiration and -1 during exhalation. To calculate $\operatorname{sgn}(dR/dt)$, we used a sliding window that spans two consecutive respiratory signal peaks. A value of -1

is assigned to all data points preceding the minimum data point, whereas a value of +1 is assigned to the rest of the points within the window. This process is continued for each succeeding window. When the sgn(dR/dt) value is positive (inhalation), $\varphi_r[n]$ spans 0 to π, whereas when sgn(dR/dt) is negative (exhalation), $\varphi_r[n]$ spans 0 to −π.

References

Aguirre, G.K., Detre, J.A., Zarahn, E., Alsop, D.C., 2002. Experimental design and the relative sensitivity of BOLD and perfusion fMRI. NeuroImage 15, 488–500.

Buxton, R.B., 2002. Introduction to Functional Magnetic Resonance Imaging. Cambridge University Press, Cambridge.

Chuang, K.H., Chen, J.H., 2001. IMPACT: image-based physiological artifacts estimation and correction technique for functional MRI. Magn. Reson. Med. 46 (2), 344–353.

Cox, R.W., 1996. AFNI-software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.

Dagli, M.S., Ingeholm, J.E., Haxby, J.V., 1999. Localization of cardiac-induced signal change in fMRI. Neuroimage 9 (4), 407–415.

Detre, J.A., Leigh, J.S., Williams, D.S., Koretsky, A.P., 1992. Perfusion imaging. Magn. Reson. Med. 23, 37–45.

Duong, T.Q., Kim, D.S., Ugurbil, K., Kim, S.G., 2001. Localized cerebral blood flow response at submillimeter columnar resolution. Proc. Natl. Acad. Sci. U. S. A. 98, 10904–10909.

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 (1), 162–167.

Golay, X., Hendrikse, J., Lim, T.C., 2004. Perfusion imaging using arterial spin labeling. Top. Magn. Reson. Imaging 15 (1), 10–27.

Hernandez-Garcia, L., Nichols, T.E., Lee, G.R., Noll, D.C., 2005. Estimation efficiency and bias in ASL-based functional MRI: frequency response of differencing strategies. Proceedings of the ISMRM 13th Scientific Meeting, Miami Beach, p. 1579.

Hu, X., Le, T.H., Parrish, T., Erhard, P., 1995. Retrospective estimation and correction of physiological fluctuation in functional MRI. Magn. Reson. Med. 34 (2), 201–212.

Josephs, O., Howseman, A.M., Friston, K., Turner, R., 1997. Physiological noise modelling for multi-slice EPI fMRI using SPM. Proceedings of the 5th ISMRM Scientific Meeting, Vancouver, p. 1682.

Kemeny, S., Ye, F.Q., Birn, R., Braun, A.R., 2005. Comparison of continuous overt speech fMRI using BOLD and arterial spin labeling. Hum. Brain Mapp. 24 (3), 173–183.

Kiebel, S.J., Glaser, D.E., Friston, K.J., 2003. A heuristic for the degrees of freedom of statistics based on multiple variance parameters. Neuroimage 20 (1), 591–600.

Kim, S.-G., 1995. Quantification of relative cerebral blood flow change by flow-sensitive alternating inversion recovery (FAIR) technique: application to functional mapping. Magn. Reson. Med. 34, 293–301.

Kruger, G., Glover, G.H., 2001. Physiological noise in oxygenation-sensitive magnetic resonance imaging. Magn. Reson. Med. 46 (4), 631–637.

Liu, T.T., Wong, E.C., 2005. A signal processing model for arterial spin labeling functional MRI. Neuroimage 24 (1), 207–215.

Liu, T.T., Frank, L.R., Wong, E.C., Buxton, R.B., 2001. Detection power, estimation efficiency, and predictability in event-related fMRI. Neuroimage 13 (4), 759–773.

Liu, T.T., Wong, E.C., Frank, L.R., Buxton, R.B., 2002. Analysis and design of perfusion-based event-related fMRI experiments. Neuroimage 16 (1), 269–282.

Luh, W.M., Wong, E.C., Bandettini, P.A., Ward, B.D., Hyde, J.S., 2000. Comparison of simultaneously measured perfusion and BOLD signal increases during brain activation with T(1)-based tissue identification. Magn. Reson. Med. 44 (1), 137–143.

Pfeuffer, J., Adriany, G., Shmuel, A., Yacoub, E., Moortele, P.-F.V.D., Hu, X., Ugurbil, K., 2002a. Perfusion-based high-resolution functional imaging in the human brain at 7 tesla. Magn. Reson. Med. 47, 903–911.

Pfeuffer, J., Van de Moortele, P.F., Ugurbil, K., Hu, X., Glover, G.H., 2002b. Correction of physiologically induced global off-resonance effects in dynamic echo-planar and spiral functional imaging. Magn. Reson. Med. 47 (2), 344–353.

Seber, G.A.F., Lee, A.J., 2003. Linear Regression Analysis. John Wiley and Sons, New York.

St Lawrence, K.S., Frank, J.A., Bandettini, P.A., Ye, F.Q., 2005. Noise reduction in multi-slice arterial spin tagging imaging. Magn. Reson. Med. 53 (3), 735–738.

Stern, C.E., Corkin, S., Gonzalez, R.G., Guimares, A.R., Baker, J.R., Carr, C.A., Sugiura, R.M., Vadantham, V., Rosen, B.R., 1996. The hippocampal formation participates in novel picture encoding: evidence from functional magnetic resonance imaging. Proc. Natl. Acad. Sci. U. S. A. 93, 8600–8665.

Tjandra, T., Brooks, J.C., Figueiredo, P., Wise, R., Matthews, P.M., Tracey, I., 2005. Quantitative assessment of the reproducibility of functional activation measured with BOLD and MR perfusion imaging: implications for clinical trial design. Neuroimage.

Wang, J., Aguirre, G.K., Kimberg, D.Y., Detre, J.A., 2003a. Empirical analyses of null-hypothesis perfusion fMRI data at 1.5 and 4 T. NeuroImage 19, 1449–1462.

Wang, J., Aguirre, G.K., Kimberg, D.Y., Roc, A.C., Li, L., Detre, J.A., 2003b. Arterial spin labeling perfusion fMRI with very low task frequency. Magn. Reson. Med. 49, 796–802.

Wang, J., Li, L., Roc, A.C., Alsop, D.C., Tang, K., Butler, N.S., Schnall, M.D., Detre, J.A., 2004. Reduced susceptibility effects in perfusion fMRI with single-shot spin-echo EPI acquisitions at 1.5 Tesla. Magn. Reson. Imaging 22 (1), 1–7.

Williams, D.S., Detre, J.A., Leigh, J.S., Koretsky, A.P., 1992. Magnetic resonance imaging of perfusion using spin inversion of arterial water. Proc. Natl. Acad. Sci. U. S. A. 89, 212–216.

Wise, R.G., Ide, K., Poulin, M.J., Tracey, I., 2004. Resting fluctuations in arterial carbon dioxide induce significant low frequency variations in BOLD signal. Neuroimage 21 (4), 1652–1664.

Wong, E.C., Buxton, R.B., Frank, L.R., 1998. Quantitative imaging of perfusion using a single subtraction (QUIPSS and QUIPSS II). Magn. Reson. Med. 39 (5), 702–708.

Woolrich, M.W., Jenkinson, M., Brady, J.M., Smith, S.M., 2004. Fully Bayesian spatio-temporal modeling of FMRI data. IEEE Trans. Med. Imaging 23 (2), 213–231.

Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited-again. Neuroimage 2 (3), 173–181.

Wu, W.C., Wong, E.C., Mazaheri, Y., 2005. The effect of flow dispersion in arterial spin labeling perfusion. Proceedings of the 13th ISMRM Scientific Meeting, Miami Beach, p. 1149.

Ye, F.Q., Frank, J.A., Weinberger, D.R., McLaughlin, A.C., 2000. Noise reduction in 3 D perfusion imaging by attenuating the static signal in arterial spin tagging (ASSIST). Magn. Reson. Med. 44 (1), 92–100.