

Using pulse oximetry to account for high and low frequency physiological artifacts in the BOLD signal

Timothy D. Verstynen^{a,*}, Vibhas Deshpande^b

^a Keck Center for Integrative Neuroscience, University of California, San Francisco, San Francisco, CA, USA

^b Siemens Medical Solutions USA, Inc., San Francisco, CA, USA

ARTICLE INFO

Article history:

Received 1 April 2010

Revised 8 November 2010

Accepted 4 January 2011

Available online 9 January 2011

ABSTRACT

The BOLD signal not only reflects changes in local neural activity, but also exhibits variability from physiological processes like cardiac rhythms and breathing. We investigated how both of these physiological sources are reflected in the pulse oximetry (PO) signal, a direct measure of blood oxygenation, and how this information can be used to account for different types of noise in the BOLD response. Measures of heart rate, respiration and PO were simultaneously recorded while neurologically healthy participants performed an eye-movement task in a 3T MRI. PO exhibited power in frequencies that matched those found in the independently recorded cardiac and respiration signals. Using the phasic and aperiodic properties of these signals as nuisance regressors, we found that the different frequency components of the PO signal could be used to identify different types of physiological artifacts in the BOLD response. A comparison of different physiological noise models found that a simple, down-sampled version of the PO signal improves the estimation of task-relevant statistics nearly as well as more established noise models that may run the risk of over-parameterization. These findings suggest that the PO signal captures multiple sources of physiological noise in the BOLD response and provides a simple and efficient way of modeling these noise sources in subsequent analysis.

© 2011 Elsevier Inc. All rights reserved.

Introduction

The blood oxygenation level dependent (BOLD) response in functional MRI (fMRI) is a powerful, albeit indirect, method for measuring task-related changes in neural activity (Logothetis et al., 2001). Unfortunately the BOLD signal is inherently variable (Aguirre et al., 1998a,b), with a major component of this variance arising from non-neural, physiological processes. In fact, depending on the acquisition parameters, physiological noise can account for 20–70% of the variance observed in fMRI data collected at 3T (Triantafyllou et al., 2005). Two major sources of physiological noise in the BOLD signal are artifacts induced by respiration cycles and contractions of the heart. These events can corrupt the task-related BOLD signal through B0 field modulations, T1 inflow and pulsatile motion, and direct modulation of blood oxygenation itself (Glover et al., 2000; Lund et al., 2006; Birn et al., 2008; Brooks et al., 2008; Chang and Glover, 2009a).

Most researchers choose to ignore the influence of these secondary, physiological processes because it is assumed that they are not time-locked to task events or because the frequency of physiological processes is generally higher than that of task-related BOLD responses. Unfortunately, there are several ways by which these processes can correlate with, or at least influence the interpretation of, task-related

changes in the BOLD signal. For example, systematic heart rate changes are known to occur in many motor control tasks (Jennings et al., 2002) as well as during feedback processing in cognitive tasks (Crone et al., 2006). This can result in the interpretation of neural changes in the BOLD signal that are, in fact, caused by secondary physiological processes induced by the task itself. Even if these signals are not directly time-locked to the task of interest, there are several ways by which physiological events can influence the estimation of task-related activity. First, the time of slice acquisition (TR) aliases the higher frequency physiological signals, resulting in lower frequency fluctuations that can sometimes match those of task-related events (Lund et al., 2006). Second, spontaneous fluctuations in respiration rate have a long-lasting effect on the BOLD signal (Birn et al., 2008). A portion of this influence is related to a high degree of correlation between respiration rate and changes in end-tidal carbon dioxide, a strong vasodilator (Chang and Glover, 2009b).

Accounting for the quasi-periodic fluctuations of cardiac and respiration cycles has been shown to improve the detection of underlying neural dynamics (Glover et al., 2000; Deckers et al., 2006; Harvey et al., 2008; Chang and Glover, 2009a). Similarly, accounting for aperiodic variations in both breathing volume (Birn et al., 2006, 2008) and heart rate (Chang et al., 2009) also accounts for a significant portion of physiological noise in the BOLD response. With increased interest in cognitive states that do not have easily discernible onsets, for example resting-state network activity, distinguishing relevant neural patterns from physiological noise becomes even more important (Shmueli et al.,

* Corresponding author. Learning Research and Development Center, Room 631, 3939 O'Hara Street, University of Pittsburgh, Pittsburgh, PA 15260, USA.

E-mail address: timothyv@pitt.edu (T.D. Verstynen).

2007; Chang and Glover, 2009a). Accounting for these physiological artifacts can also be critical when making comparisons of BOLD changes across populations with different degrees of cardiovascular integrity. For example, the ability to detect intracerebral blood vessels reduces with age (Bullitt et al., 2010), resulting in significant changes in evoked BOLD responses (Huettel et al., 2001). These changes make comparisons of the degree of task-related activation between aging cohorts problematic.

Typical measures of cardiac and respiration signals, such as electrocardiography (ECG) and pneumatic belts, can be useful in accounting for physiological artifacts in the analysis of functional brain imaging data. These, however, are indirect measures of the mechanical events that mediate fluctuations in the BOLD signal. Often there can be great variability across individuals in how these events map onto changes in global blood pressure and oxygenation. In contrast, pulse-oximetry (PO), through the use of a photoplethysmograph, provides a direct assessment of changes in global blood oxygenation by measuring the absorption of infrared light transmission through blood infused tissue (Mannheimer, 2007). Greater light absorption reflects a greater density of oxygenated hemoglobin in the underlying vessels. Being a direct measure of fluctuations in blood oxygenation, PO likely captures multiple components of the physiological noise that is reflected in the BOLD response, making it a particularly appealing method for optimizing the analysis of brain imaging data.

PO is particularly appealing over ECG because it affords a more flexible means of application. Obtaining a clean PO signal only requires a certain degree of skin translucency for recording. This can be found on many body surfaces (e.g., finger pad, ear lobe, etc.) making PO a viable alternative for individuals where ECG, or even pneumatic belts, may not be easily adapted or cleanly recorded (e.g., infants, elderly, high magnetic fields, etc.). Indeed, for this reason many researchers have generally adopted the photoplethysmograph as an alternative to MR-compatible ECG recordings of heart rate (Glover et al., 2000; Birn et al., 2006, 2008; Chang and Glover, 2009a, b; Chang et al., 2009). It is not yet clear whether PO could also supplement mechanical measures of respiration by extracting the low frequency changes in global blood oxygenation induced by breathing.

Given that both the BOLD and PO signals reflect changes in blood oxygenation, we set out to evaluate the full utility of using the spectral information contained in the PO signal to characterize and remove physiological artifacts from the BOLD response. We begin by confirming that PO captures both the low and high frequency fluctuations induced from respiration and cardiac events respectively (Nilsson et al., 2007). These filtered components are then used to produce physiological noise models of both cardiac and respiration artifacts (Glover et al., 2000; Birn et al., 2006, 2008; Chang and Glover, 2009a,b; Chang et al., 2009) and compared against similar models generated using the independently recorded measures of heart rate (ECG) and breathing (pneumatic belt). Finally, we evaluate how using novel physiological noise model, simply consisting of a down-sampled version of the raw PO amplitude, can improve the detection of task-related BOLD activity compared to more established phasic and variation models of physiological noise.

Materials and methods

Participants

Ten healthy, right-handed participants took part in this study (5 male, age range: 22–32 years). All procedures were approved by the Committee for Human Research at the University of California, San Francisco.

Task

Participants performed a block-designed saccadic eye-movement paradigm. Participants were instructed to maintain fixation on a centrally presented cross (white symbol on a black background). During each

block of trials, participants were presented with 5 movement trials (each trial having a 2 s duration). Trials began with the presentation of a red target at a randomly selected position about the fixation cross (5 visual degree distance). Participants were instructed to plan, but not execute, a saccade to the target. After 500 ms, the target color changed to green, instructing the participant to execute the saccade to the target and remain fixated on it. After 750 ms, the target disappeared and participants were instructed to return their gaze to the fixation cross. The next target would appear after a fixed 750 ms delay. Each block of trials lasted 10 s, with a 10 s inter-block interval during which participants maintained fixation. All participants received twelve blocks of trials resulting in a 4 min and 2 s total testing duration.

Physiological recording

All physiological recording was performed using the integrated Siemens Physiological Monitoring Unit. During scanning, heart rate was monitored using dermal electrodes placed over the chest (400 Hz sampling rate), that recorded the changes in electrical signal across the heart during each beat. Respiration was recorded using a pneumatic belt placed across the abdomen, just beneath the diaphragm (50 Hz sampling rate). Finally, PO was recorded using a photoplethysmograph with an infra-red emitter placed under the pad of the left index finger (50 Hz sampling). Participants were instructed to keep this hand still during the task in order to maintain a consistent signal throughout the scan. All signals were transmitted wirelessly from the scanning bed to the control room. Recording was automatically time-locked to the beginning of the first volume acquisition.

Physiological signal analysis

Processing of all physiology data was performed using the PhLEM Toolbox (<http://sites.google.com/site/phlemtoolbox/>). For the ECG data, the time series was band-pass filtered (0.6–2.0 Hz; Butterworth filter) to remove gradient artifacts from the data. Data from the pneumatic belt was up-sampled to the faster recording rate of the ECG signal using a cubic spline interpolation and then low pass filtered with a Gaussian kernel (500 ms FWHM). Finally, the PO data was similarly up-sampled to the recording rate of the ECG signal. No smoothing or filtering was performed on the PO signal prior to estimating relevant physiological regressors. For estimating phase regressors, the high frequency information in the PO signal was band-pass filtered at the same frequencies as the ECG signal (0.6–2.0 Hz; Butterworth filter). Low frequency PO phase information was filtered using a Gaussian smoothing kernel with a 400 ms FWHM.

Two types of physiological-noise regressors were generated for this study. First we estimated the dominant *phase* of quasi-periodic fluctuations induced by heart rate and respiration rate using the RETROICOR algorithm (Fig. 1; Glover et al., 2000). For each signal type (i.e., ECG, pneumatic belt, high frequency PO, low frequency PO) individual events in the signal were identified using a peak detection algorithm (*peakdet* function in Matlab; <http://www.billauer.co.il/peakdet.html>) that locates local maxima in an oscillating signal (vertical lines in Fig. 1A). In the ECG signal, these events reflected the R component of the QRS complex. In the pneumatic belt signal, these events corresponded to peak expansion of the diaphragm. For both the high and low frequency filtered PO signals, event peaks indicated individual maxima in local blood oxygenation. Once the time of events in a signal was identified, a phase time estimate between each successive event was determined by re-coding the signal between events as one full phase cycle: i.e., recoding each event as being equally spaced between zero and 2π . The cumulative summation of this phase-time signal was taken beginning with the first even in the series. This provides an estimate of the dominant Fourier series of the signal. Doubling the phase time between successive events (i.e., increasing from 2π to 4π) provides an estimate of the next higher

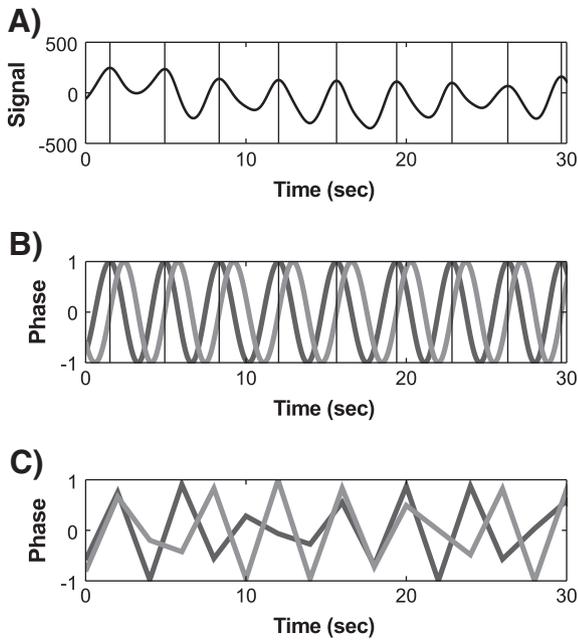


Fig. 1. A) Schematic representation of the RETROICOR algorithm. Individual events in the signal were identified using a peak detection procedure (vertical lines). B) Samples between each event were recategorized to being between 0 and 2π . The cumulative summed phase time served as an estimate of the primary Fourier series for the signal of interest. The sine (gray lines) and cosine (black lines) of this phase were then computed. C) These phase regressors were down-sampled to the TR interval and saved as covariate regressor terms in the GLM.

harmonic of the Fourier series. Fig. 1B shows the sine and cosine components of these signals. Finally, in order to generate the covariate regressors to include in the general linear model (GLM; see below), the sine and cosine components of each dominant Fourier series were then down-sampled to the sampling rate of the EPI sequence (TR = 0.5 Hz; Fig. 1C).

In addition to these phasic signals, we also generated nuisance regressors for variations in breathing rate/volume (RV; Birn et al., 2006, 2008) and heart rate (HR; Chang et al., 2009) to account for physiological noise. RV was calculated by taking the standard deviation of the respiration signal (both from the pneumatic belt and from the low-pass filtered PO signal) between each successive TR. HR was calculated by taking the average deviation in inter-event interval (i.e., ms between R-components of the QRS complex), per second, between each TR. Each signal type was then z-score normalized (gray lines in Figs. 2A and D). These variance signals were convolved with the “respiration response function” (RRF; Birn et al., 2008, Eq. (3); Fig. 2B) and “cardiac response function” (CRF; Chang et al., 2009, Eq. (5); Fig. 2E) to create model respiration (RV + RRF) and cardiac (HR + CRF) variation regressors respectively for the GLM (Figs. 2C and F).

fMRI acquisition

A single block of 121 functional images was acquired with a 3T Siemens Tim Trio System (Siemens AG, Germany) using a gradient echo-planar (EPI) pulse sequence (26 slices, $2.0 \times 2.0 \times 3.0$ mm voxel resolution, 128×128 matrix, gap = 0.45 mm, TR = 2000 ms; TE = 32 ms; FOV = 250×100 mm). All images were acquired with a Siemens 12-channel head-coil using a GRAPPA parallel imaging method (acceleration factor = 2). An online prospective motion correction adjustment (PACE) was performed during acquisition to accommodate small head movements (Thesen et al., 2000). Slice acquisition occurred in the axial plane. At the end of the scanning session a high resolution structural 3-D MPRAGE scan was acquired (1 mm³ voxel resolution; $240 \times 256 \times 160$ matrix; TR = 2300 ms; TE = 2.96 ms).

fMRI data processing and analysis

All images were reconstructed from k-space using standard Siemens software. Functional data were converted to Nifti format and analyzed using SPM8 (Wellcome Institute of Cognitive Neurology, London, UK). The EPI images for each subject were corrected for differences in the slice acquisition time and realigned to the first

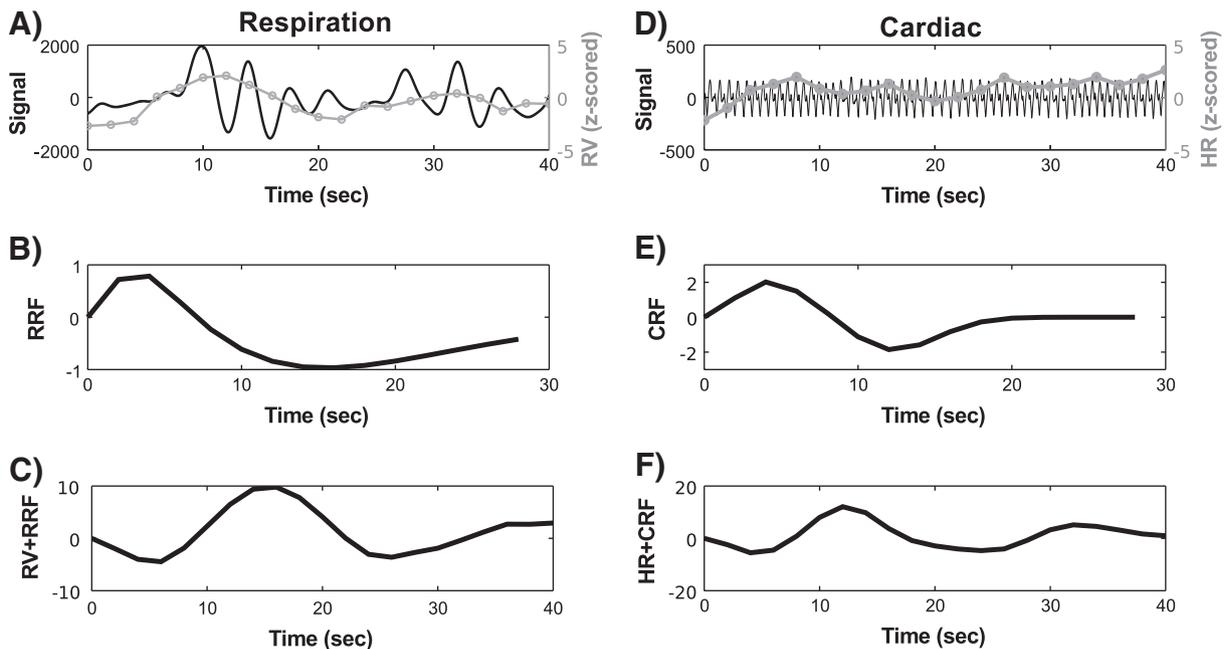


Fig. 2. A) The respiration variance (RV) was calculated by taking the standard deviation of the respiration signal (black line; either from the pneumatic belt or PO) between each successive TR (gray line). B) The respiration response function (RRF) was simulated using parameters derived elsewhere (Birn et al., 2008; Eq. (3)). C) The RV was convolved with the RRF to produce a model regressor. D) The heart-beat variation (HR; gray line) was calculated as the average deviation in inter-event interval for pulsations between each TR (black line). E) The cardiac response function (CRF) was simulated according to Chang et al. (2009) (Eq. (5)). F) The HR was then convolved with the CRF to produce a model regressor.

image in the series. These images were then registered into MNI space by normalizing to the EPI MNI template using a $7 \times 8 \times 7$ parameter nonlinear transform (Ashburner et al., 1998) and smoothed using a Gaussian kernel (4 mm FWHM).

Estimates of both task and physiology-related changes were performed using a general linear model (GLM; Kiebel and Holmes, 2004). Baseline estimates of eye movement activation (Saccade – Rest) were first determined by modeling task-related activity using a simple model with 6 head motion regressors (x, y, z, pitch, roll, yaw) included as covariates, but without physiological regressors. Areas with consistent saccade-related activity across subjects were determined using a whole-brain random effects analysis by performing a one-sample t-test on the contrast maps. Clusters of significantly active voxels ($t > 3.5$) were then segmented into 11 separate individual regions of interest (ROIs): superior colliculus, putamen (left/right), bilateral calcarine fissure (i.e., V1), superior parietal lobule (SPL; left/right), dorsal lateral precentral sulcus (dIPSC; left/right), dorsal medial precentral sulcus (dmPSC; left/right) and bilateral medial superior frontal gyrus (mSFG). Subsequent analysis of different model structures on task-relevant statistics was performed with these ROIs. It should be noted that since the relevant analysis involved looking at changes in the efficacy of first-level parameter estimates on any task-related voxels, no secondary analysis was performed. Therefore, the “double-dipping” problem (Kriegeskort et al., 2009) does not apply to the analysis in this study.

Comparisons of physiological noise models

Within-model assessments

In order to directly compare how well the inclusion of each physiologically-based regressor improves the estimation of task-relevant statistics, we adopted a bootstrap permutation approach (Manly, 1997) on data from the 11 ROIs described above. For each ROI, we first extracted the average voxel time-series, mean subtracted the data, and used an ordinary least squares fit to obtain regression coefficients, β s, for the input parameters described in one of the three design matrices shown in Fig. 3A. These are the design matrices used to produce the maps in Figs. 7A–C (blue voxels). Next we performed a t-contrast on the regression coefficient for the eye-movement task regressor (β_{Task}) as follows

$$T_{\text{Task}} = \beta_{\text{Task}} / \text{SE}(\beta_{\text{Task}}).$$

Efficiently accounting for physiological noise should optimize the estimate of task-related statistics better than what would be expected

simply using arbitrarily random regressors (i.e., chance). To test this hypothesis, we randomly scrambled the image volume order, i.e., rows in the design matrix, of each physiological term, i.e., relevant column, and re-estimated β_{Task} and T_{Task} . This process was repeated 10,000 times for each subject and ROI, using a with-replacement sampling method, in order to generate a probability distribution reflecting the values expected to be obtained by chance (Fig. 3B). This type of randomization eliminates the precise temporal sequence of signal changes that characterize the underlying physiological events. The area under curve between the observed T_{Task} and the mean of the simulated distribution (μ_{Sim} ; see Equation in Fig. 3B) provides an estimate of the percent change, either improvement (positive values) or impairment (negative values), that the true physiological regressors provide compared to chance. Finally, these estimates from each ROI were compared across subjects using a one-sample t-test.

Between-model assessments

As a traditional goodness-of-fit comparison across models with different numbers of regressors we first adopted a likelihood ratio test (LRT; Stuart and Ord, 1987). In this test, the difference in the log likelihood of the model with more parameters, $\log(p(y | \theta_{\text{Model1}}))$, and the model with fewer parameters, $\log(p(y | \theta_{\text{Model0}}))$, falls off according to a χ^2 distribution. The degrees-of-freedom of this distribution is the difference in the number of parameters in each model. For our comparisons, each GLM that included one of the three physiological noise models (i.e., downsampled PO, variation or phasic model), was compared against the standard GLM model containing only task and head motion regressors. A summary of the p-values observed from this test for all 11 ROIs is shown in Table 3.

As an alternative test of the goodness-of-fit that does not require nested models, we utilized another permutation test approach. First, we simulated the effect of adding random parameters to a simple GLM model that only initially included task and head motion regressors. For this analysis, we chose to use the residual square error (i.e., the deviance of the GLM for normally distributed data; Stuart and Ord, 1987) as a measure of goodness-of-fit. For each ROI, a Gaussian white noise regressor, $\mu = 0$ and $\sigma = 1$, was added to the design matrix and the model parameters refit using an ordinary least squares fit method (*glmfit* in Matlab). The difference between the residual sums-of-squares (RSS) error of this model and the original model was then calculated. This simulation was repeated 2000 times for each ROI and for up to 20 Gaussian noise parameters. The result of these simulations is shown in Fig. 9A.

These simulations were then used to normalize the goodness-of-fit for each of three different physiological noise models (see Results, Improving detection of task-relevant responses section). For each ROI

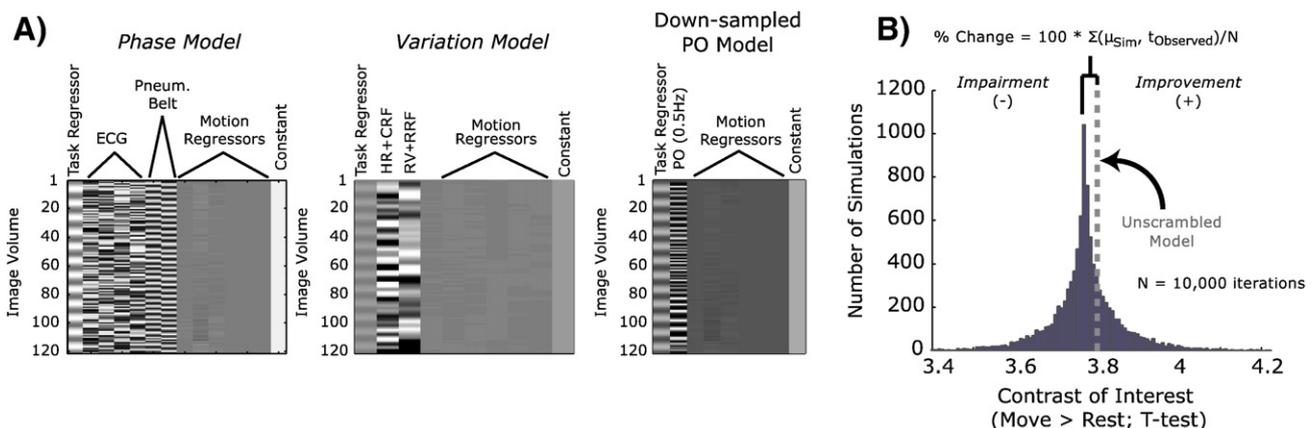


Fig. 3. A) Design matrices from a single subject of the three GLM models being compared. The Phase model has the principle Fourier series of the ECG signal as well as one phase expansion, along with the first-order Fourier series for the pneumatic belt data (see Glover et al., 2000). The variation model includes both the convolved RV + RRF and the HR + CRF vectors. The PO model includes only the pulse-oximetry data down sampled to the TR frequency. B) Histogram of the contrast values (movement–rest) for the simulated models where the physiology regressors are scrambled to reflect random regressor values. The measure of improvement or impairment was estimated by measuring the area under the curve between the mean of the simulated distribution (μ_{Sim}) and the observed test statistics (t_{Observed}). Dashed line shows the real contrast value from the unscrambled model (from the Phase model shown in A).

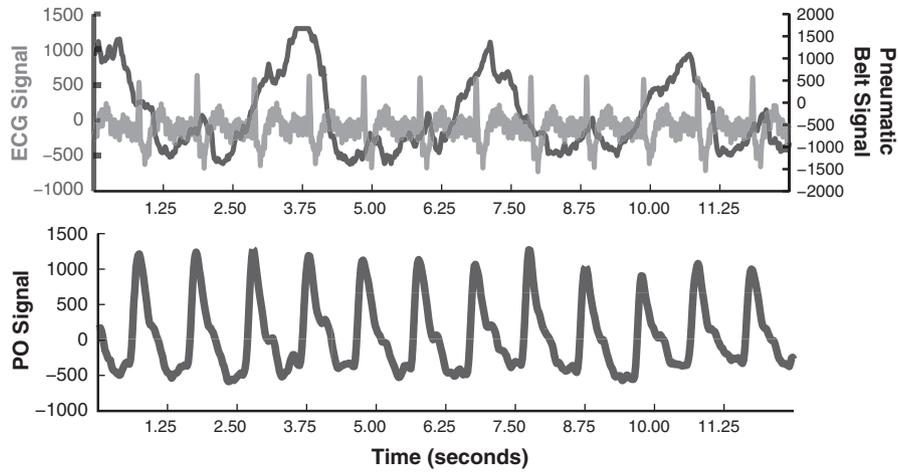


Fig. 4. The time-locked, unfiltered time series for the three signals from a representative participant.

and noise model type, we calculated the distance, as a z-score, between the observed RSS from the physiological noise model and the distribution of RSSs obtained using a Gaussian white noise model with the same number of parameters. These z-scores were averaged across subjects to produce a mean normalized change in the goodness-of-fit per ROI. The difference in the normalized goodness-of-fit between the different physiological noise models was then determined using paired-sample t-tests on the z-scored data across subjects.

Since the BOLD signal has an autoregressive structure to it (Biswal et al., 1995), this bootstrap analysis was repeated using nuisance regressors that were temporally correlated (i.e., non-white noise). For this analysis, an auto-correlative structure was added to the noise model terms by simulating an ARMA(1,1) process using the *armax* function in Matlab with the parameters $A = [1 \ -0.5]$, $B = [0 \ 0.2]$ and $C = [1 \ 0.6]$ (Brockwell and Davis, 2002). The resulting time series had an average lag-1 autocorrelation of $r=0.5$. All other analysis procedures were kept the same.

Results

Comparison of physiological signals

We were able to reliably record from all three signal sources in all participants. An example of the unfiltered time series for each signal

type is shown in Fig. 4. To determine the optimal bandwidth of information in each signal, we performed a power spectral density analysis on each time series using the Welch method (*pwelch* in MATLAB). Each signal was then normalized to the peak power spectral density of the time series and averaged across subjects. Fig. 5A confirms that the PO signal contains multiple peaks that overlap with the low frequency respiration cycles (0.2–0.5 Hz) and high frequency cardiac cycles (1–1.5 Hz and 2–3 Hz). The spectral analysis of the ECG signal reveals significant power across a broader range of frequencies; however much of this has to do with the complexity of the cardiac waveform itself (inset Fig. 5B; Hara et al., 1999). Nonetheless, the high frequency peak in the PO signal (0.6–2.0 Hz) corresponds to a normal resting heart rate of 45 to 90 bpm, confirming that this frequency band contains information reflecting the cardiac component of oscillations in global blood oxygenation.

Given the differences in the nature of each signal type (i.e., mechanical events vs. oxygenated hemoglobin content), there should be a phase lag between peaks in the ECG/pneumatic belt data and phases in the components of the PO signal. To measure this we filtered the PO signal to isolate the individual high and low frequency components thought to reflect cardiac and respiration events. Each of these filtered signals was then phase locked to the peak of the ECG (for high frequency PO information) and pneumatic belt (for low frequency PO information) waveforms. When aligned to the R component of the

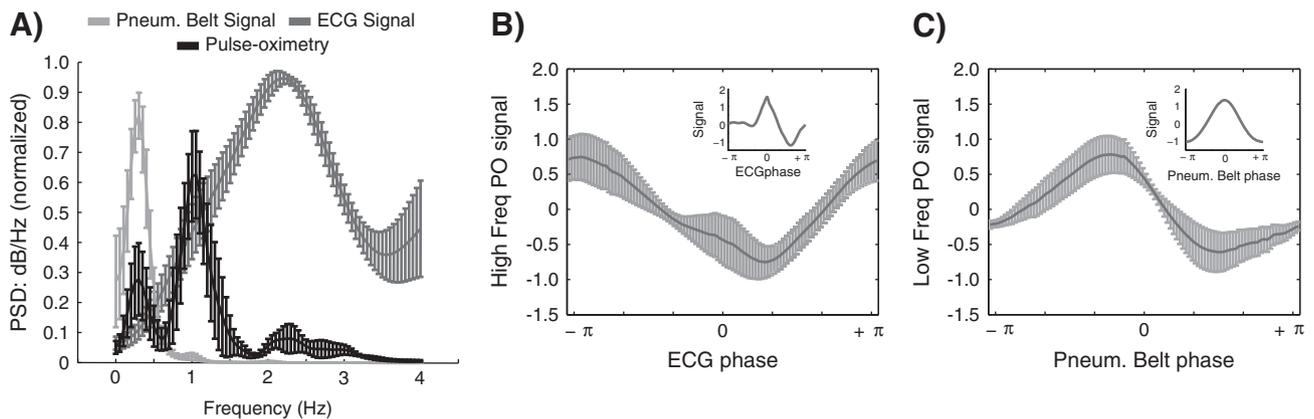


Fig. 5. A) Spectral density functions for all three signal types (mean \pm standard error). The respiration signal from the pneumatic belt (light gray) peaks at about 15 breaths per minute. The cardiac signal from the ECG (dark gray) covers a broad range of frequencies ranging from 0.5 to 3.5 Hz. The pulse oximetry (PO) signal (black) also has a peak corresponding to 15 events per minute, as well as a peak at approximately 60 events per minute (see also Nilsson et al., 2007). A minor high frequency peak can be found at about 2.25 Hz as well. B) Phase-locking the filtered high frequency PO (0.6–2.0) signal to the peak of the cardiac signal events. The PO signal appears to be phase shifted 180° from the peak of the cardiac signal. C) Same phase-locking analysis for the low frequency PO signal (<1 Hz) to the peak of the respiration waveform. Here the PO signal appears to be nearly completely phase shifted in time.

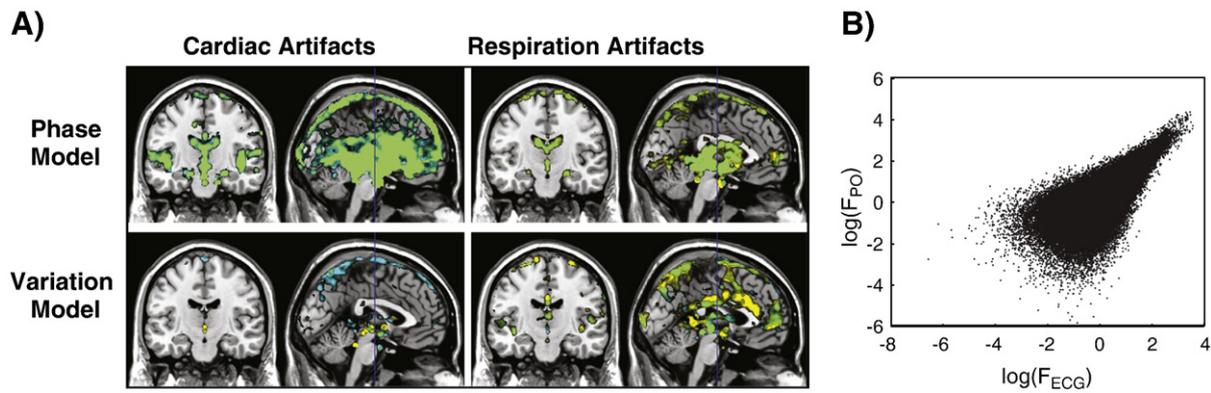


Fig. 6. A) A comparison of estimated physiological artifacts shown as the mean extra sums of squares (ESS) accounted for by the physiological noise terms across all subjects. For the Phase model voxels are thresholded for ESS at 10 and for the Variation model, voxels are threshold at 5. Voxels with physiological artifacts are generally clustered in regions with large vascular structures. Green areas reflect overlap of artifacts estimated from the indirect, mechanical recordings (ECG and pneumatic belt; yellow voxels) and the PO signal (cyan voxels), with ESS values both greater than threshold. B) An example of the correlation of log F-values for a different subject. The mean correlations, averaged across subjects, are shown in Table 1.

QRS complex, the high frequency PO signal is phase shifted about 180° from the peak of each cardiac event (Fig. 5B). Thus, the high frequency peaks in PO signal that reflect the cardiac component (see peaks in Fig. 5A) occur in between each ventricular contraction. The low frequency component of the PO signal consistently peaked just before the point of maximum diaphragm expansion, meaning that this signal is almost completely phase shifted from the point of the previous peak inhalation (Fig. 5C). Therefore, it appears that changes in blood oxygenation at distal vasculature lag behind the mechanical events that trigger them. It remains unclear how similar this lag at finger vasculature systems is to the lags that occur in neural vasculature. Given the difference in arterial distances (Fischer et al., 2010), it is presumed that this lag at the finger vasculature is longer than lags that would occur in the brain. We explore this further in the **Time locking of the PO signal** section below.

This analysis confirms that the PO signal reflects both cardiac and respiration components of blood oxygenation modulation (see also Nilsson et al., 2007). Using spectral filtering, it was possible to isolate these components, which appeared to be shifted from the time of the mechanical events that cause them. This allows for the possibility of using the different spectral components of the PO signal to identify both cardiac and respiration artifacts in the BOLD signal.

Artifact identification in the BOLD signal

Using data from a blocked eye movement task (see Methods, Task section) we set out to see how these measures of cardiac and respiration processes can capture physiological artifacts in the BOLD signal. Artifacts were determined by generating two different classes of noise models using each recording method: a phasic RETROICOR model (Glover et al., 2000; see Methods, Physiological signal analysis section) and a physiological variation model for heart rate (HR + CRF; Chang et al., 2009) and breathing (RV + RRF; Birn et al., 2008). Fig. 6A shows the distribution of artifacts identified from each physiological

noise model and signal type. For this analysis, each model was computed separately. The models generated using the filtered PO signal (cyan voxels) appeared to capture a slightly larger distribution of cardiac artifacts across the brain than artifacts identified using the ECG signal (yellow voxels). Yet, the cardiac phase regressors from both ECG and high frequency PO signals were also able to pick up a significant amount of overlapping artifacts (indicated as green voxels), particularly in areas of large vasculature (for example voxels in the Circle of Willis). For respiration artifacts, the data from the pneumatic belt was able to detect a broader distribution of artifacts, particularly in cortical gray matter, than the low frequency filtered PO data. As with the cardiac artifacts, there was still a significant amount of overlap between respiration artifacts identified using the different signal types, particularly in subcortical areas regions with larger vasculature.

To quantify the consistency of artifact detection between these signal types, we looked at the voxel-by-voxel correlation of log-transformed F-values (Fig. 6B) using a Spearman rank order correlation test. Mean results across subjects are summarized in Table 1, with comparisons of interest highlighted in gray. Phasic model artifacts that were calculated using the high frequency PO signal were primarily correlated with those that were estimated using the ECG signal. Similarly the low frequency PO component was primarily correlated with the pneumatic belt signal. Artifacts from the variation model showed a similar pattern of correlation, i.e., the strongest correlations occurred between ECG and high frequency PO signals and pneumatic belt with low frequency PO signals; however these correlations are weaker than those observed with the phasic physiological noise model.

The correlation analysis in Table 1 illustrates how changes in the magnitude of F-statistics compare across signal type and physiological noise model. For mapping the distribution of voxels that are significantly influenced by physiological signals, it is in some ways more appropriate to identify voxels whose contrast statistic surpasses a given significance threshold. Therefore, we also looked at the

Table 1
Correlations between log-valued artifact estimates (mean \pm standard deviation).

	Phase regressors			Variation regressors			
	RESP	PO-high	PO-low	ECG	RESP	PO-high	PO-low
ECG	0.03 \pm 0.07	0.57 \pm 0.27	0.02 \pm 0.06	-0.01 \pm 0.05	-0.01 \pm 0.06	-0.02 \pm 0.05	0.01 \pm 0.06
	RESP	0.06 \pm 0.06	0.48 \pm 0.26	0.01 \pm 0.06	0.02 \pm 0.07	0.02 \pm 0.07	0.01 \pm 0.05
		PO-high	0.04 \pm 0.06	0.01 \pm 0.08	0.01 \pm 0.07	-0.02 \pm 0.06	0.03 \pm 0.10
			PO-low	-0.01 \pm 0.01	-0.01 \pm 0.06	-0.01 \pm 0.05	-0.00 \pm 0.04
				ECG	0.06 \pm 0.06	0.17 \pm 0.24	0.05 \pm 0.05
					RESP	0.10 \pm 0.13	0.21 \pm 0.23

Table 2
Overlap of statistically significant physiological voxels ($p < 0.05$, mean \pm standard deviation).

	Phase regressors			Variation regressors			
	RESP	PO-high	PO-low	ECG	RESP	PO-high	PO-low
ECG	1% \pm 2	48% \pm 33	1% \pm 2	1% \pm 2	2% \pm 3	1% \pm 2	2% \pm 3
	RESP	3% \pm 4	40% \pm 26	4% \pm 4	6% \pm 7	6% \pm 7	5% \pm 5
		PO-high	3% \pm 5	3% \pm 3	5% \pm 5	3% \pm 2	5% \pm 4
			PO-low	2% \pm 5	4% \pm 6	3% \pm 2	3% \pm 4
				ECG	12% \pm 5	23% \pm 18	13% \pm 7
					RESP	18% \pm 13	29% \pm 21

overlap of significantly “active” voxels, i.e., those voxels exhibiting statistically significant physiological noise. For each regressor type (phasic vs. variation), we identified all voxels with F-contrasts reflecting $p < 0.05$ from both the PO and non-PO signals and then calculated the percent overlap between each set of voxels. This analysis is shown in Table 2. As with the changes in artifact magnitude, we found that the percentage of overlapping physiologically significant voxels was strongest between ECG and high frequency PO signals and between pneumatic belt and low frequency PO signals. Again, this effect was strongest for phasic noise regressors than for variation noise regressors.

Our analysis has replicated the observation that, using models of both the phasic and variance properties of physiological processes, it is possible to capture the major noise components in the BOLD response (Hu et al., 1995; Glover et al., 2000; Birn et al., 2006; Deckers et al., 2006; Birn et al., 2008; Harvey et al., 2008; Chang and Glover, 2009a,b; Chang et al., 2009). As expected, the different frequency components of the PO signal provided qualitatively similar estimates of physiological artifacts as the more indirect, mechanical measures. This further supports the notion that PO signal contains relevant information about the two major physiological noise sources in the BOLD response. In this way, the slow and fast modulations of PO amplitude capture the convolved modulations in global blood oxygenation induced by respiration and cardiac processes. This, along with the fact that both the BOLD and PO signals are direct measures of regional blood oxygenation, suggests that the moment-by-moment oscillations of the PO signal may also be reflected in the dynamic modulations of the BOLD signal. If true, then simply using the raw PO signal as a model of physiological noise may improve estimates of task-related activity better than what is expected by chance and possibly as efficiently as existing physiological noise models.

Improving detection of task-relevant responses

For most researchers, the utility of accounting for physiological noise is that it allows for a more optimal estimation of task-related parameters (Harvey et al., 2008; Chang and Glover, 2009a). When using a physiological noise model that accounts for the phasic properties of both cardiac and respiration signals (blue voxels, Fig. 7A), we see a reduction in the number of task-related voxels ($t > 3.5$) in ventral areas that express more physiological artifacts (see Fig. 6A). In contrast, motor planning areas (e.g., parietal motor planning regions) exhibit an increase in the number of task-related voxels when the phasic physiological noise terms are included. Including the phasic noise model in the GLM reduced the residual mean square error of the model fit, averaged across all voxels in the brain, from 1.03 ± 0.06 (mean \pm standard error of the mean) to 0.93 ± 0.04 . This change was consistent across subjects when tested using a 1-sample t-test ($t(9) = 4.86$, $p < 0.001$).

A similar pattern of changes in task-related voxels is observed when using a full variation regression model (RV + RRF and HR + CRF; Fig. 7B). In this case, including the variation model reduced the residual mean square error only slightly to 1.02 ± 0.05 . While this

reduction in variance did change the pattern of significantly active voxels, the overall change in residual error from the simple GLM model was not consistent across subjects ($t(9) = 0.77$, $p = 0.23$).

If the raw PO signal contains the combined information of cardiac and respiration processes, then presumably including it as a nuisance regressor it should produce a similar pattern of task-related voxels as the larger phasic and variation models. Indeed, with this novel physiological noise model we observed a pattern of task-related voxels that resembled those found with the more traditional noise models (Fig. 7C). The down-sampled PO model reduced the residual mean square error across voxels to 0.99 ± 0.05 , which was a consistent change observed in all subjects ($t(9) = 3.77$, $p = 0.002$).

At first glance the change in the distribution of task-relevant voxels would suggest that all three models improve the estimation of task-related BOLD responses in roughly equivalent ways. However, in every comparison, the models that included noise regressors have more parameters than the simple model that does not include these terms. Therefore the change in the distribution of task-related voxels may simply reflect over-parameterization (McCullagh and Nelder, 1989). To illustrate this point, we performed the same comparison but with a Gaussian white noise term instead of a regressor linked to any physiological process (Fig. 7D). As expected, random white noise model produced patterns of activation nearly identical to those seen using the physiologically-based regressors. This is particularly true when compared to the down-sampled PO model that has the same number of model terms. The random white noise model even reduced the mean residual error across voxels to 1.00 ± 0.06 , which is a significant reduction from the simple GLM across all subjects ($t(9) = 2.76$, $p = 0.01$). It should be emphasized, that these effects are observed even though this model does not directly map on to any physiological processes. Thus, the distribution of task-related voxels that we modeled cannot differentiate optimal noise fitting from simple over-parameterization.

To get a sense of how including each noise model improves the overall GLM fit, we first used the traditional likelihood ratio test (LRT; Stuart and Ord, 1987) to compare the goodness-of-fits across model types. Each LRT compared the residual squared-error for a model that includes physiological terms against a baseline model with only task and head-motion terms. The difference in residual error should fall off with the difference in number of model terms according to a χ^2 distribution. Table 3 shows the mean (\pm standard error) p-value across subjects for a set of LRTs using data from 11 different task relevant ROIs (see Methods, Physiological signal analysis and Comparisons of physiological noise models sections). For this comparison we tested the different physiological noise models shown in Figs. 7A–C: the phase regressor model using data from the ECG and pneumatic belt inputs (Phase model; see Harvey et al., 2008), the variation model using data from the same signal sources (Variation model), and the model using the down-sampled PO signal (PO Model). Using this approach, none of the physiological noise terms significantly improved the residual variance over the simpler model with fewer terms.

However, this null result is in some ways expected and uninformative. It is well known that the LRT can be insensitive to

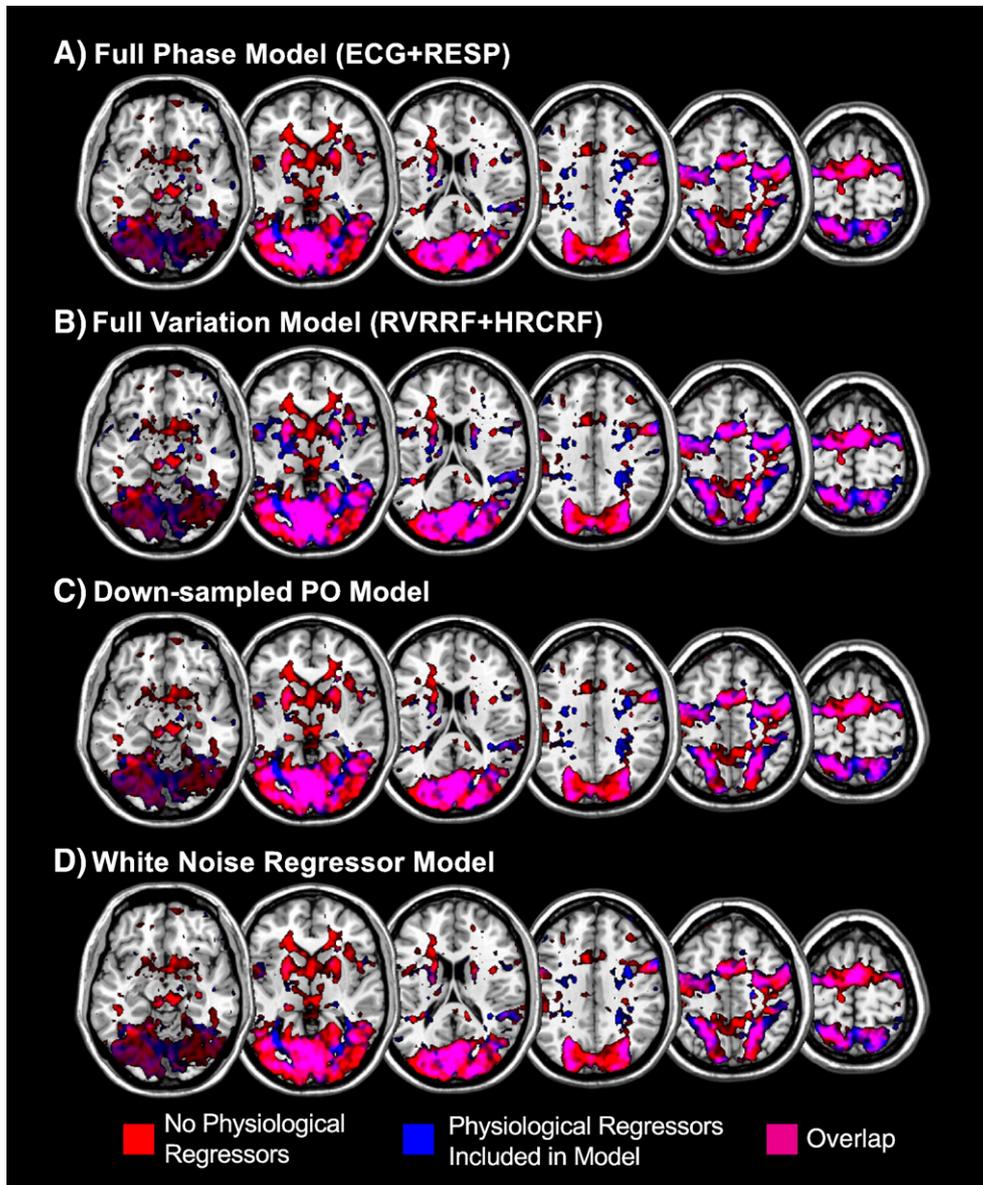


Fig. 7. Changes in task-related voxels ($t > 3.5$) from a model with no physiological noise regressors (red voxels) to a model that includes these regressors (blue voxels) for all subjects. A) A full model of heart-rate phase (plus 1 phase expansion) and respiration phase regressors using data from the ECG and pneumatic belt. B) A full variation model (RV + RRF and HR + CRF) using data from the ECG and pneumatic belt. C) A model using the down-sampled PO signal as a regressor. D) A model using a single white-noise vector as a nuisance regressor.

small differences in model size (i.e., number of parameters) when the sample size is small (Lehmann, 2004; Geweke and Singleton, 1980). The 121 samples used to fit each model did not allow for sufficient

Table 3
LRT p-values across ROIs ($\mu \pm$ SEM).

Region of interest	PO model	Variation model	Phase model
Left dmPSC	0.64 \pm 0.07	0.88 \pm 0.04	0.97 \pm 0.01
Right dmPSC	0.59 \pm 0.06	0.80 \pm 0.07	0.93 \pm 0.04
Left dlPSC	0.60 \pm 0.10	0.72 \pm 0.10	0.82 \pm 0.07
Right dlPSC	0.60 \pm 0.08	0.68 \pm 0.10	0.77 \pm 0.10
mSFG	0.68 \pm 0.09	0.72 \pm 0.12	0.82 \pm 0.08
Left SPL	0.78 \pm 0.06	0.88 \pm 0.06	0.88 \pm 0.06
Right SPL	0.68 \pm 0.07	0.77 \pm 0.06	0.86 \pm 0.06
V1	0.63 \pm 0.09	0.79 \pm 0.07	0.88 \pm 0.06
Sup. colliculus	0.45 \pm 0.12	0.53 \pm 0.12	0.51 \pm 0.12
Left putamen	0.61 \pm 0.07	0.79 \pm 0.07	0.97 \pm 0.01
Right putamen	0.70 \pm 0.06	0.81 \pm 0.08	0.96 \pm 0.01

sensitivity to detect small differences in fits, particularly with such a restricted range of degrees-of-freedom (ranging from 1 to 5 in the tests shown in Table 3). Indeed, this might be a reason why the LRT has not been used in previous studies of physiological noise, where investigators instead adopted other means of comparing across models (Birn et al., 2006, 2008; Chang et al., 2009; Chang and Glover, 2009a; Glover et al., 2000). Permutation test approaches have sometimes been shown to be more sensitive to differences of model fits than closed-form, asymptotic model selection tests similar to the LRT (Zhao et al., 2010; Preacher and Hayes, 2008). These approaches use sampling-with-replacement permutation techniques to generate a hypothetical null distribution using the observed data itself.

We devised two types of bootstrap tests. The first was a within-model test designed to determine how well each physiological noise model influenced the estimation of task-related statistics compared to what would be expected by chance. This analysis was performed on data from the same ROIs and noise models shown in Table 3. All models exhibited positive percent change values (Figs. 8A and B),

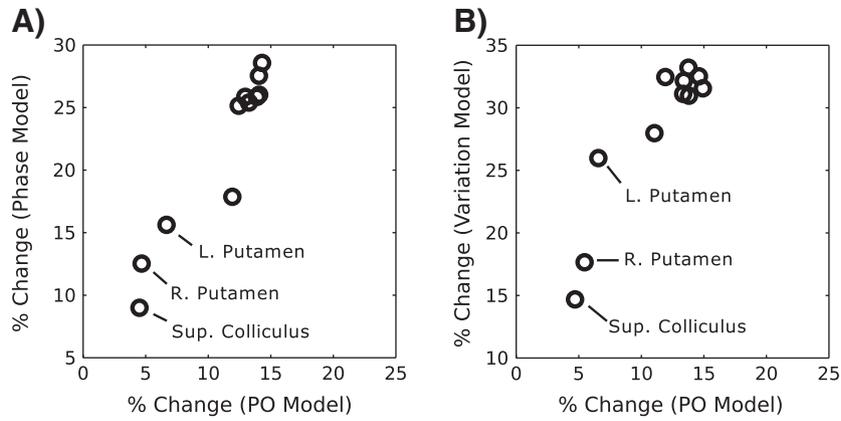


Fig. 8. A) Group permutation test results compared across all 11 ROIs for the full Phase model and the down-sampled PO (see Fig. 3A). Assessment of within-region significance was determined using a 1-sample t-test across all subjects. All regions had a statistically significant shift (see Results, Improving detection of task-relevant responses section). Subcortical ROIs are highlighted. The unlabeled datapoint near the left putamen is the left dorsal medial precentral sulcus (dmPCS). B) Similar comparison as in A for the full variation model (RV + RRF and HR + CRF) versus the down-sampled PO model. As with A, the data point outside the cortical ROI cluster is for the dmPCS ROI.

indicating that the inclusion of physiological noise terms improved the overall estimates of task-related activity better than chance. This was significant across subjects in every cortical ROI (all $p < 0.011$, minimum $t(9) = 4.23$). While the effect was somewhat weaker in the three subcortical ROIs, all three were nonetheless significant across subjects according to 1-sample t-tests (all $p < 0.0156$, minimum $t(9) = 2.55$).

More importantly, there was a strong correlation between the established physiological noise models and the down-sampled PO model (Fig. 8; $r = 0.96$ and $r = 0.74$ for the phasic and variation models respectively). Yet despite such strong correlations, the overall effect size is much larger for the phasic and variation models than for the PO model. This difference in magnitude, however, may again be explained by over-parameterization. For example, the phasic model includes 6 physiological regressors while the PO model has only a single physiological term.

To illustrate this effect further, we conducted a simulation where a series of arbitrary white noise regressors were added to a GLM that initially only included task and head motion parameters. For this simulation, we looked at how the residual sums of squares (RSS) changed with the inclusion of each additional random noise term. As is shown in Fig. 9A, the RSS of any GLM drops linearly with the inclusion of each random regressor. Thus it can be problematic to compare the goodness-of-fit of a model with 6 physiological parameters to a model with only 1 or 2 physiological terms.

Therefore we used a second permutation approach to compare the goodness-of-fit across model types. This evaluation reflects an

alternative to the LRT described above and shown in Table 3 (see Methods, Comparisons of physiological noise models section). For this analysis we calculated the change in RSS observed from each physiological noise model with the change of RSS expected from a distribution of simulated models with the same number of random Gaussian noise parameters. The resulting z-score serves as a transformation to normalize the size of the model fits across noise models with different numbers of parameters. We specifically chose white noise regressors because they should have none of the temporal or spectral structure qualities of actual physiological noise. Therefore, the improvement introduced by these regressors (see Fig. 9A) should be purely due to chance, as opposed to capturing any part of the physiological noise structure that is also contained in the physiology regressors.

Similar to our observations using the within-model bootstrap comparison, the down-sampled PO and the phasic noise models in this test were both significantly better at reducing residual error than comparable random white noise models (PO: $t(9) = 2.50$, $p = 0.017$; Phasic: $t(9) = 4.45$, $p = 0.001$). The variation model was on average better than including white noise regressors, but this did not reach statistical significance ($t(9) = 1.42$, $p = 0.095$). Thus the improvement in estimated task statistics we observed above (Fig. 8) reflects, in part, a reduction in the overall residual squared error of each model. Comparing between model types, we found that, after accounting for the differences in the number of model parameters, there was not a significant difference in goodness-of-fit between the down-sampled PO model and the variation noise model ($t(9) = 0.49$, $p = 0.32$).

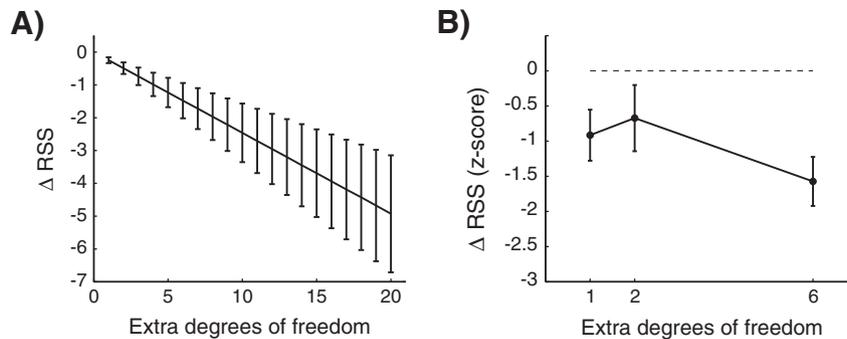


Fig. 9. A) Change in the residual sums of squares (i.e., the model RSS) when Gaussian white noise terms are added to a simple task GLM. 2000 simulations were run for each number of added model noise terms. Data shown as mean and standard deviation from a set of simulations for one ROI in an example subject. B) The change in RSS for each of the physiological noise models normalized according to the distribution of changes expected from a Gaussian noise model with the same number of parameters. Data shown as mean and standard error of the mean across subjects.

Table 4
Correlation between regression vectors.

	ECG cosine (1)	ECG sine (2)	ECG cosine (2)	RESP sine	RESP cosine	ECG: HR+CRF	RESP: RV+RRF	PO-high sine (1)	PO-high cosine (1)	PO-high sine (2)	PO-high cosine (2)	PO-low sine	PO-low cosine	PO: HR+CRF	PO: RV+RRF	Raw PO	Random noise	Task	X	Y	Z	Pitch	Roll	Yaw
ECG sine (1)	-0.02	-0.01	-0.07	-0.02	-0.02	-0.01	-0.04	0.12	-0.01	0.03	-0.02	0.00	0.00	0.04	0.00	0.01	0.01	-0.04	0.03	-0.02	-0.04	0.05	-0.01	0.04
ECG cosine (1)	-0.02	-0.01	-0.01	0.02	0.01	-0.06	-0.06	0.12	0.03	0.03	-0.02	-0.02	0.02	0.02	0.08	0.01	0.05	-0.02	0.02	0.08	-0.06	-0.06	0.00	0.04
ECG sine (2)	-0.01	-0.04	-0.02	-0.02	0.00	-0.04	0.04	0.17	-0.29	0.01	0.01	-0.04	0.00	-0.05	0.03	0.01	0.04	0.03	0.01	-0.03	0.00	-0.02	0.02	0.02
ECG cosine (2)	-0.03	0.00	0.01	0.00	-0.01	-0.03	0.30	0.18	-0.02	-0.03	-0.02	0.01	-0.01	0.01	0.01	0.04	0.03	0.01	0.01	-0.03	0.04	0.04	0.06	0.06
RESP sine	-0.01	0.01	0.04	0.00	0.04	-0.02	0.00	0.33	0.21	0.04	0.05	0.03	-0.05	0.07	-0.31	0.01	0.14	-0.03	0.01	0.14	-0.03	0.06	-0.08	-0.08
RESP cosine	-0.01	-0.02	0.01	-0.02	0.00	0.00	0.00	-0.30	0.35	-0.04	0.01	0.07	0.04	0.02	0.36	-0.08	-0.08	0.01	0.01	0.01	0.01	0.01	0.01	0.04
ECG: HR+CRF	0.03	0.04	-0.01	-0.03	0.02	0.02	-0.04	0.31	-0.04	0.01	0.02	-0.03	-0.03	-0.03	0.03	0.03	0.05	0.01	0.01	0.01	0.01	0.01	0.01	0.01
RESP: RV+RRF	-0.02	0.02	-0.01	0.00	0.01	0.01	0.15	0.17	0.00	0.00	-0.03	0.07	0.02	-0.18	0.00	0.09	-0.19	-0.03	0.07	0.02	-0.18	0.00	0.09	-0.19
PO-high sine (1)	-0.04	0.01	-0.04	0.00	-0.03	0.04	0.01	0.15	0.01	0.15	0.01	-0.03	0.02	-0.04	-0.02	0.05	-0.02	-0.03	0.02	-0.04	-0.02	0.05	-0.02	0.00
PO-high cosine (1)	0.02	0.04	0.01	0.02	0.00	0.01	0.85	0.01	0.05	-0.01	-0.06	0.07	-0.08	0.06	0.06	-0.02	0.00	-0.01	-0.06	0.07	-0.08	0.06	-0.02	0.00
PO-high sine (2)	-0.01	0.01	0.02	-0.02	-0.03	-0.01	0.01	-0.02	-0.01	-0.03	0.03	0.01	0.04	0.01	0.04	0.01	0.04	0.06	0.01	-0.03	0.03	0.01	0.04	0.06
PO-high cosine (2)	0.00	-0.01	0.01	0.00	0.28	-0.01	-0.01	0.01	0.01	-0.04	0.05	0.01	0.01	-0.04	0.01	0.03	-0.03	0.01	0.01	-0.04	0.05	0.01	0.03	-0.03
PO-low sine	-0.03	0.01	0.01	-0.09	-0.03	0.06	-0.33	0.00	0.11	-0.01	0.01	-0.01	0.01	-0.01	0.01	0.01	0.01	0.01	0.01	-0.01	-0.01	0.01	0.01	-0.04
PO-low cosine	-0.05	0.01	0.06	0.00	0.02	-0.04	-0.08	0.05	-0.01	0.04	0.01	0.04	0.01	0.04	0.01	0.04	0.01	0.04	0.01	-0.04	0.05	-0.01	0.04	-0.01
PO: HR+CRF	0.15	-0.01	0.02	-0.03	-0.04	-0.09	0.04	0.04	0.04	0.01	-0.14	-0.09	0.04	0.04	0.04	0.01	-0.06	-0.03	-0.04	-0.09	0.04	0.04	0.01	-0.06
PO: RV+RRF	-0.01	0.02	0.00	0.01	-0.14	-0.09	-0.03	0.07	-0.12	0.00	0.01	-0.14	-0.09	-0.03	0.07	-0.12	0.00	0.01	-0.14	-0.09	-0.03	0.07	-0.12	0.00
Raw PO	0.01	0.02	0.04	-0.08	0.08	-0.03	0.05	-0.02	-0.02	0.01	0.01	-0.03	-0.04	-0.09	0.04	0.04	0.01	-0.06	-0.03	-0.04	-0.09	0.04	0.01	-0.06
Random noise	0.00	0.01	-0.04	0.02	-0.03	0.05	0.03	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Task	-0.02	0.02	0.02	0.01	-0.83	-0.44	0.00																	
X	0.07	-0.57	0.00	0.10	0.13																			
Y	-0.20	-0.17	-0.04	-0.03																				
Z	-0.04	-0.07	-0.02																					
Pitch	0.41	0.01																						
Roll	-0.03																							

The phasic model was slightly better than the down-sampled PO model ($t(9) = 1.84, p = 0.05$) and variation noise model ($t(9) = 1.53, p = 0.08$), but neither of these passed the significance threshold.

Since the BOLD signal is temporally correlated, i.e., it has an autoregressive structure (Biswal et al., 1995), it is possible that using simple white noise could artificially increase the chances of observing significant effects. Therefore we re-ran the bootstrap simulations using an auto-correlated noise model. This yielded nearly identical results as the white noise model: down-sampled PO $z = -0.93 \pm 0.40$ (standard error), variation model $z = -0.42 \pm 0.33$, and phasic model $z = -1.56 \pm 0.38$. Similar to our previous analysis, we did not observe a significant difference in goodness-of-fit between the down-sampled PO and either the variation noise model ($t(9) = 0.99, p = 0.71$) or the phasic model ($t(9) = 1.69, p = 0.06$); however, the phasic model still performed slightly better than the variation model ($t(9) = 2.09, p = 0.03$). The similarities in outcomes suggest that white noise models work equally as well as temporally correlated, non-white noise for these types of bootstrap model selection tests.

These results confirm that the dynamic fluctuations of the PO signal itself can serve as another model of physiological noise in the BOLD response. Unfortunately our results cannot determine whether the PO model can be used to replace or simply supplement the phasic or the variation physiological noise models. However, knowing when to include a given noise model as a nuisance regressor necessitates some understanding of its collinearity with other noise models and components of the GLM. Table 4¹ shows the average cross correlation between all regression vectors used in this study, including movement values for head translation (x, y, and z) and rotation (pitch, roll, and yaw). Correlations between regressors generated with data from ECG/pneumatic-belt and PO component signals are highlighted for reference. While somewhat small, these are some of the strongest correlations observed. We also found negative correlations between the sine components of the respiration phase (both from the pneumatic

belt and low frequency PO) and the x-component of head movements. This is consistent with the assumption that a portion of the respiration artifact is due to subtle changes in head position. Most importantly, however, none of the physiological regressors demonstrated a significant correlation with the eye movement task itself. This table can be used as a guide for future models that optimally combine different noise types into the GLM.

To summarize, our analysis presents an alternative form of physiological noise modeling for fMRI analysis where the fluctuation patterns of the PO signal itself can be used to improve the estimates of task-relevant BOLD responses. Utilizing a permutation test approach, we were able to confirm that this change is greater than what is expected by chance. The overall effect of including the simple down-sampled PO amplitude regressor was smaller than the improvement observed using a more standard phase or variation model; however, after accounting for differences in the number of model parameters, we found that it performed comparably as well as the variation and phasic noise models.

Time locking of the PO signal

As stated above, PO measures the fluctuations in blood oxygenation at local vasculature near the recording site. Differences in arterial distance between the index finger and brain vasculature will mean that the signal recorded at the finger may not be time-locked to the pattern observed in the brain (see Nilsson et al., 2007). To estimate this lag difference, we looked at how shifting the PO signal forward and backward in time, before down-sampling, affects the estimate of task-related BOLD response (i.e., T-contrast) in each ROI.

Fig. 10A shows the results of this temporal lag analysis for a single ROI, averaged across all subjects. Shifting the PO signal forward and backward in time resulted in a systematic change in estimated t-values. As expected, there is significant variability across subjects, resulting in large error bars; however, this is not surprising given that individual differences in arterial distances would change the relative optimal shift value for different brain regions. Averaging the

¹ Standard deviations were excluded from Table 3 for the sake of clarity.

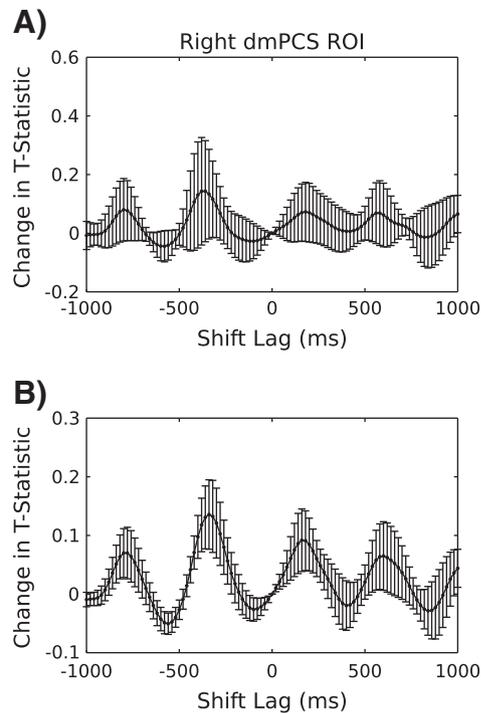


Fig. 10. A) Changes in estimated task statistic in a single ROI, when the PO vector is shifted in time before downsampling. Error bars reflect standard deviation. B) Mean and standard error of shift affects on estimated task statistics across all 11 ROIs. Peak changes in t-values occurred when the PO vector was shifted 340 ms back in time.

mean subject response across all 11 ROIs, we see a consistent pattern emerge where shifting the PO signal back 340 ms produces the greatest improvement in task-related t-values (Fig. 10B).

There are many potential sources of this phase misalignment. Sensor placement site (e.g., finger, ear-lobe, etc.), signal recording onset relative to scanning onset, and individual variability in vascular systems will determine the precise optimal shift required to best account for physiological noise. Nonetheless, this analysis suggests that physiological fluctuations in the brain generally happen *earlier* than those observed in distal limbs. In cases where this ambiguity in phase alignment is significant, a more optimal regression model would include a basis set of shifted signals at different lags in order to account for many possible phase lags between the PO signal and BOLD fluctuations.

Discussion

Our results demonstrate that the PO signal captures multiple sources of noise in the BOLD signal. Low frequency components of the PO signal reflect slow changes in oxygenation and small head movements related to breathing. Higher frequency PO components reflect mechanical pulsations related to contractions of the heart muscle and fast local changes in oxygenated blood as a result of these blood pressure changes. Each component can be used to identify different types of physiological noise in the BOLD response. This indicates that the PO signal itself captures the combined dynamics of respiration and cardiac processes and might be useful as an alternative model of physiological noise. Indeed, after accounting for differences in the number of model terms, a down-sampled version of the PO signal improves the estimation of task-relevant BOLD responses comparably as well as other noise models that run the risk of over-parameterization (McCullagh and Nelder, 1989).

In our analysis, the down-sampled PO model most closely correlated with the phasic noise model than with the variation

nuisance regressors. This was a little surprising since the RETROICOR model generally accounts secondary, non-BOLD perturbations such as magnetic field modulations induced by respiration and T1 inflow/pulsatile motion from cardiac rhythms (Glover et al., 2000), whereas the variation and PO models should both reflect modulations of the BOLD signal itself. This difference may be due to the fact that the variation model captures slow variations in the BOLD signal that happen several seconds after the mechanical events that induce them (Birn et al., 2008; Chang et al., 2009). The simple PO model in its current form does not capture these slower components. However, it should be noted that the *variation* in pulse height of the raw PO signal has recently been shown to account for physiological noise in the BOLD signal without necessitating a secondary convolution term (Van Houdt et al., 2010). It is the goal of future work to identify any potential impulse response functions that better map PO fluctuations to BOLD modulations.

Before using the down-sampled PO model to account for physiological noise, either as an independent regressor or in conjunction with other physiological noise models, some caution should be taken with regard to the nature of the signal being recorded. Some photoplethysmograph systems high-pass filter and rescale/threshold PO data during the recording process. This filtering can attenuate or even eliminate the low-frequency respiration components of the PO signal, as well as the structure of pulsatility variability. Analysis of the spectral power of the PO signal should reveal the characteristic respiration and heart-beat peaks shown in Fig. 5. Otherwise independent measures of respiration should be taken in order to capture both the high and low-frequency components of physiological noise. Also, many systems do not precisely time-lock physiological recording to scan onset. This variable lag between physiological signal and scan acquisition may significantly impact the effectiveness of using a physiological nuisance regressor. As mentioned above (Time locking of the PO signal section), the introduction of a basis set of shifted regressor values could be one option for compensating for these effects. However, such models may also run the risk of over-parameterization. Thus, with large datasets, likelihood ratio tests (Stuart and Ord, 1987) can be used to perform optimal model selection. Otherwise, bootstrap permutation approaches, similar to what we used here, should be considered to determine whether the inclusion of any set of nuisance regressors significantly improves the overall model fit of the data.

Finally these problems in physiological signal integrity can be exacerbated by certain preprocessing steps commonly used in fMRI. For example, many researchers apply slice-time correction to functional data. This interpolates voxel values to more closely emulate the same point in time. The time-scale of this temporal correction can corrupt the phasic properties of underlying physiological signals in the BOLD response and reduce the power of any nuisance regressor. Also, many analysis packages high-pass filter data (either the functional data itself or the design matrix). In much the same way that high-pass filtering reduces the sensitivity of the PO signal to respiration artifacts, this filtering of the functional time-series can impair the ability to capture the lower frequency respiration noise components. Therefore, careful consideration of signal recording and preprocessing steps should be taken when applying physiological noise correction in GLM analysis.

Acknowledgments

The authors would like to thank the reviewers, P. Sabes, K. Erickson, and A. Weinstein for their helpful comments. J. Diedrichsen and J. Schlerf assisted with discussion, code development, preliminary analysis, and pilot studies. This project was funded by a Sandler Postdoctoral Translational Neuroscience fellowship and a Swartz Foundation fellowship to TV.

References

- Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998a. The inferential impact of global signal covariates in functional neuroimaging analyses. *Neuroimage* 306, 302–306.
- Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998b. The variability of human, BOLD hemodynamic responses. *Neuroimage* 369, 360–369.
- Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., Friston, K., 1998. Identifying global anatomical differences: deformation-based morphometry. *Hum. Brain Mapp.* 357, 348–357.
- 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.
- Birn, R.M., Smith, M.A., Jones, T.B., Bandettini, P.A., 2008. The respiration response function: the temporal dynamics of fMRI signal fluctuations related to changes in respiration. *Neuroimage* 40, 644–654.
- Biswal, B., Yetkin, F.Z., Haughton, V.M., Hyde, J.S., 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541.
- Brockwell, P., Davis, R., 2002. *Introduction to Time Series and Forecasting*, 2nd Edition. Springer Hall, New York.
- Brooks, J.C., Beckmann, C.F., Miller, K.L., Wise, R.G., Porro, C.A., Tracey, I., Jenkinson, M., 2008. Physiological noise modelling for spinal functional magnetic resonance imaging studies. *Imaging* 39, 680–692.
- Bullitt, E., Zeng, D., Mortamet, B., Ghosh, A., Aylward, S.R., Lin, W., Marks, B.L., Smith, K., 2010. The effects of healthy aging on intracerebral blood vessels visualized by magnetic resonance angiography. *Neurobiol. Aging* 31, 290–300.
- Chang, C., Glover, G.H., 2009a. Effects of model-based physiological noise correction on default mode network anti-correlations and correlations. *Neuroimage* 47, 1448–1459.
- Chang, C., Glover, G.H., 2009b. Relationship between respiration, end-tidal CO₂, and BOLD signals in resting-state fMRI. *Neuroimage* 47, 1381–1393.
- Chang, C., Cunningham, J.P., Glover, G.H., 2009. Influence of heart rate on the BOLD signal: the cardiac response function. *Neuroimage* 44, 857–869.
- Crone, E.A., Somsen, R.J., Zanolie, K., 2006. A heart rate analysis of developmental change in feedback processing and rule shifting from childhood to early adulthood. *J. Exp. Child Psychol.* 95, 99–116.
- Deckers, R.H., Gelderen, P.V., Ries, M., Barret, O., Duyn, J.H., Ikonomidou, V.N., Fukunaga, M., Glover, G.H., Zwart, J.A., 2006. An adaptive filter for suppression of cardiac and respiratory noise in MRI time series data. *Analysis* 33, 1072–1081.
- Fischer, G.W., Benni, P.B., Satyapriya, A., Afonso, A., Luozzo, G.D., Griep, R.B., Reich, D.L., Surgery, C., Sinai, M., Gustave, O., Place, L.L., 2010. Mathematical model for describing cerebral oxygen desaturation in patients undergoing deep hypothermic circulatory arrest. *Br. J. Anaesth.* 104, 59–66.
- Geweke, J.F., Singleton, K.J., 1980. Interpreting the likelihood ratio statistic in factor models when sample size is small. *J. Am. Stat. Soc.* 75, 133–137.
- 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, 162–167.
- Hara, M., Tsuchiya, K., Nanke, T., Mori, N., Miyake, F., Sato, C., Murayama, M., 1999. Development of a new analytical method for the electrocardiogram using short-time first Fourier transforms. *Jpn Circ. J.* 63, 941–944.
- Harvey, A., Pattinson, K., Brooks, J., Mayhew, S., Jenkinson, M., Wise, R.G., 2008. Brainstem functional magnetic resonance imaging: disentangling signal from physiological noise. *J. Magn. Reson. Imaging* 28, 1337–1344.
- Hu, X., Le, T.H., Parrish, T., Erhard, P., 1995. Retrospective estimation and correction of physiological fluctuation in functional MRI. *Magn. Reson. Med.* 34, 201–212.
- Huettel, S.A., Singerman, J.D., McCarthy, G., 2001. The effects of aging upon the hemodynamic response measured by functional MRI. *Neuroimage* 13, 161–175.
- Jennings, J.R., Molen, V.D., W.M., 2002. Cardiac timing and the central regulation of action. *Psychol. Res.* 66, 337–349.
- Kiebel, S., Holmes, A., 2004. The general linear model. In: F. R.S.J., F. K.J., F. C.D., D. R.J., P. C. J., Z. S. A., J., P. W. (Eds.), *Human Brain Function*. Elsevier, San Diego, pp. 749–779.
- Kriegeskort, N., Simmons, W.K., Bellgowan, P.S., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* 12, 535–540.
- Lehmann, E., 2004. On likelihood ratio tests. In: Rojo, J. (Ed.), *IMS Lecture Notes – Monograph Series: Optimality – The Second Lehman Symposium*. Institute of Mathematical Sciences Press, Ohio, pp. 1–8.
- Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157.
- Lund, T.E., Madsen, K.H., Sidaros, K., Luo, W., Nichols, T.E., 2006. Non-white noise in fMRI: does modelling have an impact? *Neuroimage* 29, 54–66.
- Manly, B., 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd. Chapman and Hall, New York.
- Mannheimer, P., 2007. The light–tissue interaction of pulse oximetry. *Anesth. Analg.* 105, S10–S17.
- McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*, 2nd. Chapman and Hall, New York.
- Nilsson, L., Goscinski, T., Kalman, S., Lindberg, L., Johansson, A., 2007. Combined photoplethysmographic monitoring of respiration rate and pulse: a comparison between different measurement sites in spontaneously breathing subjects. *Acta Anaesthesiol. Scand.* 1250–1257.
- Preacher, K., Hayes, A.F., 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Meth.* 40 (3), 879–891.
- Shmueli, K., Van Gelderen, P., de Zwart, J.A., Horowitz, S.G., Fukunaga, M., Jansma, J.M., Duyn, J.H., 2007. Low-frequency fluctuations in the cardiac rate as a source of variance in the resting-state fMRI BOLD signal. *Neuroimage* 38, 306–320.
- Stuart, A., Ord, J., 1987. *Kendall's Advanced Theory of Statistics*, 5th. Oxford University, New York.
- Thesen, S., Heid, O., Mueller, E., Schad, L.R., 2000. Prospective acquisition correction for head motion with image-based tracking for real-time fMRI. *Magn. Reson. Med.* 44, 457–465.
- Triantafyllou, C., Hoge, R.D., Krueger, G., Wiggins, C.J., Potthast, A., Wiggins, G.C., Wald, L.L., 2005. Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. *Neuroimage* 26, 243–250.
- Van Houdt, P.J., Ossenblok, P.P., Leijten, F.S., Velis, D.N., Stam, C.J., Munck, J.C., 2010. Correction for pulse height variability reduces physiological noise in functional MRI when studying spontaneous brain activity. *Hum. Brain Mapp.* 325, 311–325.
- Zhao X., Lynch J.G., Chen Q., 2010. Reconsidering Baron and Kenny: myths and truths about mediation analysis. *J. Consum. Res.* 37, 197–206.