◆ Human Brain Mapping 28:59–68(2007) ◆

Calibration of BOLD fMRI Using Breath Holding Reduces Group Variance During a Cognitive Task

Moriah E. Thomason,[1]* Lara C. Foland,[2] and Gary H. Glover[1,2]

[1]*Neurosciences Program, Stanford University School of Medicine, Stanford, California*
[2]*Department of Radiology, Stanford University School of Medicine, Stanford, California*

Abstract: The proportionality of blood oxygen level-dependent (BOLD) response during a cognitive task and that from a hypercapnic challenge was investigated in cortical structures involved in working memory (WM). Breath holding (BH) following inspiration was used to induce a BOLD response characteristic of regional vasomotor reactivity but devoid of metabolic changes. BOLD effects measured during BH were used to normalize individual subject activations during WM, which effectively reduced the confounding influence of individual- and region-specific differences in hemodynamic responsivity common to both tasks. In a study of seven subjects, the BH calibration reduced intersubject variability in WM effect amplitude by 24.8% ($P < 0.03$). Reduced intersubject variability resulted in a 23.7% increase in group WM activation voxel extent significant at $P < 0.001$, with further increases at more stringent thresholds. Because the BH task does not require CO_2 inhalation or other invasive manipulations and is broadly applicable across cortical regions, the proposed approach is simple to implement and may be beneficial for use not only in quantitative group fMRI analyses, but also for multicenter and longitudinal studies. *Hum Brain Mapp 28:59–68, 2007.*

Key words: breath holding; BOLD fMRI; calibration; normalization; hemodynamics; vasoreactivity

INTRODUCTION

Functional magnetic resonance imaging (fMRI) is a powerful tool for probing neural function in health and disease. Many studies seek to draw quantitative inferences about various cognitive functions from measurements of the amplitude, spatial extent, and temporal characteristics of the blood oxygen level-dependent (BOLD) response [Harms and Melcher, 2003; Ogawa et al., 1990; Saad et al., 2001]. However, the BOLD contrast is not a direct measure of neuronal metabolism but rather results from changes in regional blood flow and attendant alterations in oxygenation [Ogawa et al., 1990]. In particular, focal increases in neuronal activity are accompanied by increased cerebral metabolic rate of oxygen ($CMRO_2$), as oxygen facilitates conversion of glucose into adenosine triphosphate (ATP), the brain's primary source of energy. This in turn causes a net increase in local cerebral blood flow (CBF) and blood volume (CBV). Although some details of the process are still not fully understood, CBF is increased by dilation of arterial sphincters in response to a combination of chemical messengers including increased [CO_2], [H^+], and [NO] and decreased [O_2] [Roland, 1993]. Thus, increased neuronal firing causes a cascade that culminates in a vasomotor reaction wherein blood flow is upregulated, i.e., metabolic increases cause a

Contract grant sponsor: National Institutes of Health (NIH); Contract grant numbers: P41-RR09784, F31-MH71996, M01-RR000827.

L.C. Foland is now at the Laboratory of Neuroimaging, University of California, Los Angeles.

*Correspondence to: Moriah E. Thomason, Neurosciences Program, Stanford University School of Medicine, Lucas MRI/S Building, MC 5488, 1201 Welch Road, Stanford, CA 94305-5488.
E-mail: moriah@stanford.edu

Received for publication 27 September 2005; Accepted 13 January 2006

DOI: 10.1002/hbm.20241
Published online 2 May 2006 in Wiley InterScience (www.interscience.wiley.com).

local, task-activation-initiated hemodynamic response that is characteristic not only of the neural activity but also the vascular reactivity. Because of this, it would be prudent to remove variations in vasoreactivity that affect BOLD signal in order to achieve a more accurate measure of the underlying neuronal activity.

Global hypercapnia is one means of probing characteristic differences in hemodynamic response patterns and has been suggested for normalizing BOLD response between different subjects, different brain regions, and various features of the scanning environment (acquisition parameters, field strength, etc.) [Bandettini and Wong, 1997; Cohen et al., 2004; Thomason et al., 2005]. The present study used BOLD effect measured in response to a minimally cognitive, hypercapnic challenge, i.e., a breath-holding (BH) task, to apply correction to a working memory (WM) cognitive task. It was expected that BH would be useful for normalization of BOLD signal amplitude because BH-induced BOLD response is proportional to neural responses measured by BOLD, as shown in the next section. The correction was performed within subjects on a voxel-by-voxel basis in order to reduce variance across subjects, while retaining regional specificity of vasomotor reactivity.

THEORY

A model based on tight coupling between blood volume and blood flow was proposed to explain the relationship between BOLD signal during task activation, S_{act}, and measurable hemodynamic quantities [Davis et al., 1998]:

$$S_{act} = S_0[f_{act}^{\alpha-\beta} m^{\beta} - 1], \quad (1)$$

where S_0 is a constant that depends on vasomotor reactivity and other local characteristics, $f_{act} = CBF_{act}/CBF_0$ is the fractional increase in flow relative to baseline, $m = CMRO_{2act}/CMRO_{20}$ is the fractional increase in metabolic rate of oxygen, and α and β are constants. The coupling between blood volume and blood flow may be characterized by $CBV \propto CBF^{\alpha}$, where $\alpha = 0.4$ has been observed [Davis et al., 1998; Hoge et al., 1999]. β varies between 1 and 2 depending on the degree of susceptibility vs. diffusion weighting and relative contributions from intra- and extravascular blood compartments, which in turn depend on field strength. For 3T a reasonable compromise value is 1.0, while 1.5 is more characteristic at 1.5T [Buxton et al., 1998]. The relationship between CBF and $CMRO_2$ is controversial, but all experimental studies demonstrate a disproportionately larger increase in CBF than is necessary to support the increased oxygen consumption during task activation. For simplicity, we assume in accordance with previous findings [Fox and Raichle, 1984; Hoge et al., 1999; Kastrup et al., 2002] that changes in CBF are proportional to changes in $CMRO_2$, as:

$$f_{act} - 1 = (m - 1)n, \quad (2)$$

where the proportionality n has been observed to vary from ~2 [Hoge et al., 1999] to ~5 [Fox and Raichle, 1984].

In contrast with task activation whereby local metabolic increases cause upregulation of CBF, during a BH maneuver changes in local CBF are initiated by events *outside* the brain. As the chest expands during sustained inspiratory breath holding, the heart rate transiently increases to supply the extra blood volume but quickly drops below its baseline rate because of reduced intrathoracic cardiovascular resistance [Nakada et al., 2001; Thomason et al., 2005; West, 1985]. This results in a reduction in global blood supply to the brain and a concomitant diminution in CBF. However, in the brain basal metabolism continues to consume oxygen, and thus energy stores and O_2 become depleted while CO_2 concentrations increase. As in task activation, these messengers trigger a vasomotor reaction dependent on the basal $CMRO_2$, and cause an increase in CBF consistent with maintenance of the baseline state. Thus, even though there is no change in cerebral metabolism during BH, regulatory processes respond in a fashion similar to that during task activation to maintain hemodynamic stasis in the capillary bed in accord with the regionally variable vascular function. As a result, BH can be used as a means to characterize vasoreactivity devoid of cognition [Kastrup et al., 1998, 1999; Thomason et al., 2005], and the response amplitude in each region will reflect that region's vascular reaction to the challenge.

With the assumption that $CMRO_2$ does not change during BH, from Eq. 1 the BOLD signal equation for BH becomes simply:

$$S_{BH} = S_0[f_{BH}^{\alpha-\beta} - 1]. \quad (3)$$

Note that for a given brain region, the multiplicative constant S_0 is common between the two conditions of task activation and BH. This suggests that hemodynamic and other variations between brain regions and individuals encompassed in S_0 can be removed by dividing S_{act} by S_{BH} to obtain normalized values:

$$S_{norm} = [f_{act}^{\alpha-\beta} m^{\beta} - 1]/[f_{BH}^{\alpha-\beta} - 1]. \quad (4)$$

Note that Eq. 4 depends on the flow changes during the activation and BH tasks as well as the $CMRO_2$ changes during activation. However, as long as the flow changes during BH and task activation are similar in a given region of cortex, the normalized values should have reduced variability when comparing different brain regions or in comparisons between individuals or groups of individuals because S_0 variations are removed.

In order to examine this approach, BOLD signals for task and BH activation defined by Eqs. 1 and 3 were calculated for various parameters. Figure 1 shows modeled BOLD signal changes during task activation and BH, and BOLD signal for task activation vs. BH while CBF is varied, for several values of n. Studies have shown the value of $F = f_{BH}/f_{act}$ can be expected to vary between ~1.15 for a sensory task with relatively large CBF changes and ~1.6 for a more subtle cognitive task [Hoge et al., 1999; Kastrup et al., 2002]. As

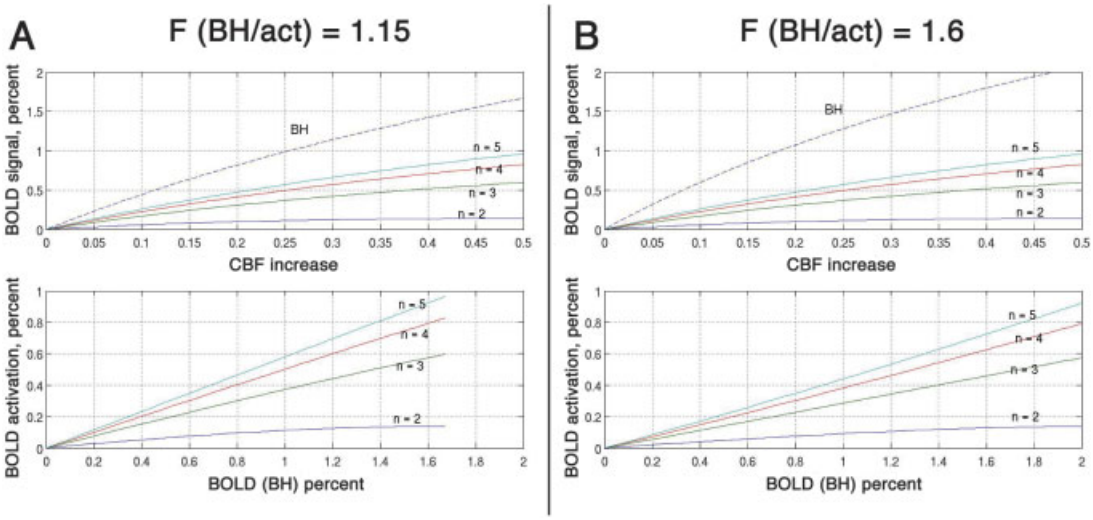

Figure 1.

Top: BOLD response as CBF is varied for BH and task activation, with metabolism-CBF coupling parameter n varied and $F = 1.15$ (**A**) and 1.6 (**B**) as shown. The signal changes are larger for BH for a given CBF because $CMRO_2$ changes during activation downregulate flow-induced modulations in deoxyhemoglobin. Bottom: corresponding BOLD signal during task activation vs. BOLD for BH as n is varied. Note that a nearly linear relationship is obtained that is characteristic of activation-induced metabolic rate, and that variations in F have relatively small effect.

may be seen in Figure 1, the slope of the activation response vs. BH response (S_{norm}) depends strongly on n but less so on F, indicating that the normalization may be expected to provide a measurement indicative of local metabolism and having reduced dependence on other aspects of the local hemodynamics.

Hence, we may hypothesize that by reducing vasoreactivity characteristics unconnected to neuronal metabolism, the proposed normalization process should result in reduced intergroup and intersubject differences when drawing inferences about cognition from BOLD measurements. In addition, it has been shown that although BH BOLD response is relatively uniform across the brain, it does vary globally from ~2.8%–3.5% [Thomason et al., 2005]. Thus, this normalization may also be beneficial when utilizing measurements of the hemodynamic response function (HRF) in sensory regions for characterizing subject or group differences in a nonsensory cognitive task [Aguirre et al., 1998] by providing an interregion transfer coefficient.

According to the model in Eq. 1, variations in S_{act} observed across a population of subjects will reflect individual variations in both metabolic response to the task, m, as well as variance in vasoreactivity, described by S_0. Thus, some of the group variance in the normalized response $S_{norm} = S_{act}/S_{BH}$ will be reduced to the extent that S_{act} and S_{BH} are correlated within subjects (as they should be according to the model). Let r be the correlation coefficient between the two measurements for the population. Then, the standard deviation of the normalized measurements σ_{norm} will be:

$$\sigma_{norm} = [1 - r^2]^{1/2}\sigma_{act}, \quad (5)$$

where σ_{act} is the standard deviation (SD) for the measured activation. Equation 5 suggests that a testable consequence of our theory is that the group variance should be reduced by normalization, and the reduction may be predicted by the extent to which the measurements of activation signal and BH signal are correlated.

SUBJECTS AND METHODS

Subjects

Data were collected from 7 healthy, right-handed, native English speakers (3 male, 4 female; mean age, 22.5; range, 20–26 years) after giving informed consent as approved by the Stanford Institutional Review Board.

Experimental Paradigm

Breath holding

Subjects performed seven repetitions of alternating periods of breath holding and self-paced breathing in 18-s blocks. During the task, subjects held their breath after inspiration. Both *end-expiration* and *end-inspiration* protocols invoke a hypercapnic response and a vasomotor compensatory reaction similar to that during task-induced $CMRO_2$ upregulation [Kastrup et al., 1998]; however, the inspiratory-BH task has several advantages. While holding one's breath at end-expiration, there is a compelling urge to breath toward the end of the apneic period, which can be uncomfortable for some subjects and lead to task-correlated motion. In addition, this discomfort may lead to cortical responses in

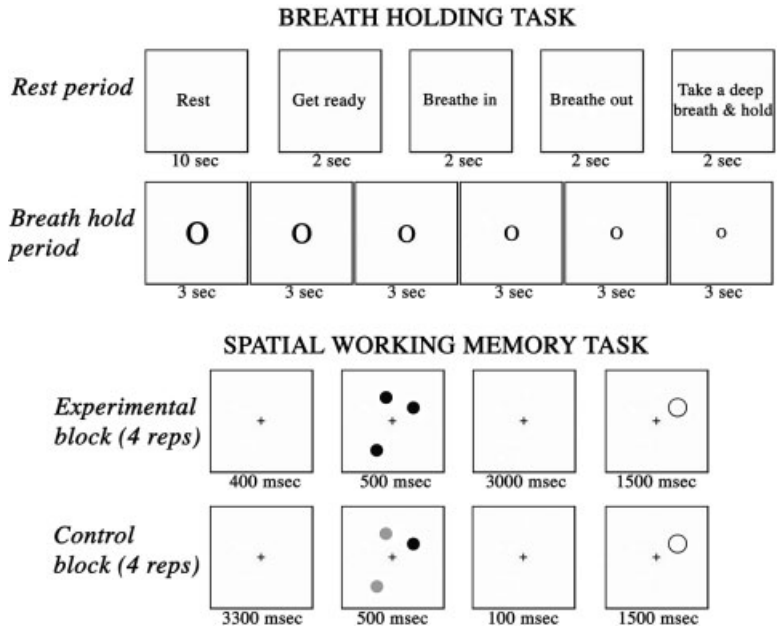

Figure 2.

Top: Breath holding task stimuli and timing. Experimental epoch cues subject with a nonverbal, shrinking ring that demarcates the passage of time. Control period involves a preparatory stage that concludes with subject inspiration. Bottom: Working memory task stimuli and timing. Experimental condition contains a 3-s maintenance phase that is not present in the control condition. Both the control and the experimental trials lasted a total of 5.4 s and each was repeated four times within a block for a total block duration of 21.6 s. No delay occurred between blocks. There were 18 total blocks, half control and half experimental, for a total of 72 trials.

the very regions that BH might be most valuable for calibration, i.e., prefrontal cortex, thereby introducing a confound. Furthermore, it was important that the paradigm be simple to implement across a broad spectrum of populations such as children and older adults. Subject compliance to task timing and ability to hold breath were measured by using a pneumatic belt placed around the abdomen to monitor breathing as well as by subject report. All subjects were able to successfully hold their breath during all trials. Trial timing was cued by visual stimulus that included a "rest" and "get ready" phase during self-paced breathing and a nonverbal stimulus (a shrinking ring) during breath holding (Fig. 2, top).

Working memory

Subjects performed a spatial memory task in which three visual locations indicated by three target dots were held in mind over a 3-s delay. Dots were randomly arrayed across four invisible concentric circles centered around a fixation cross. After the delay period a location probe, consisting of a single outline circle, appeared for 1500 ms, either encircling the location of one of the previous target dots or not (Fig. 2, bottom). Subjects pressed one of two buttons: one indicating location-match of probe ring to a target dot and the other indicating a nonmatch of probe ring and target dot location.

MRI Acquisition

Magnetic resonance imaging was performed on a 3.0 T GE (Milwaukee, WI) whole-body scanner with a custom quadrature birdcage headcoil. Head movement was minimized using a bite bar and foam padding. Twenty-three oblique axial slices were taken parallel to the anterior/posterior commissure (AC-PC) with 4-mm slice thickness, 1-mm interslice skip. High-resolution T2-weighted fast spin echo structural images (TR = 3000 ms, TE = 68 ms, ETL = 12, FOV = 24 cm, matrix 192 × 256) were acquired for anatomical reference. A T2*-sensitive gradient echo spiral in/out pulse sequence [Glover and Lai, 1998; Glover and Law, 2001] was used for functional imaging (TR = 1500 ms, TE = 30 ms, flip angle = 70°, FOV = 24 cm, matrix 64 × 64). An automated high-order shimming procedure based on spiral acquisitions was used to reduce B0 heterogeneity [Kim et al., 2002]. Spiral in/out methods have been shown to increase signal-to-noise ratio (SNR) and BOLD contrast to noise ratio in uniform brain regions as well as to reduce signal loss in regions compromised by susceptibility-induced field gradients generated near air–tissue interfaces such as PFC [Glover and Law, 2001]. Compared to traditional spiral imaging techniques, spiral in/out methods result in less signal dropout and greater task-related activation in PFC regions [Preston et al., 2004]. A high-resolution T1 volume scan (124 slices, 1.2 mm thickness) was collected for every subject

using an IR-prep 3D FSPGR sequence for T1 contrast (TR = 8.9 ms, TE = 1.8 ms, TI = 300 ms, flip angle 15°, FOV = 24 cm, slice thickness 1.2 mm, matrix 256 × 192 × 128).

Data Analysis

Preprocessing and model generation

fMRI data for both WM and BH tasks were preprocessed using SPM (Wellcome Department of Cognitive Neurology, London) and custom MATLAB routines (MathWorks, Natick, MA). Preprocessing included correction for motion and linear signal drift. Subsequent processing followed slightly different paths for region of interest (ROI) analysis and for group activation comparison. Prior to calibration and ROI analysis, images were not spatially normalized, but were spatially smoothed with an 8-mm full-width at half-maximum (FWHM) 3D Gaussian filter. fMRI-measured activity during the tasks epochs (BH and WM, separately) was compared to baseline activity during the baseline conditions (normal breathing and recognition with no-delay, respectively). Regressors for the corresponding condition blocks were modeled as a boxcar function convolved with the canonical HRF. Statistical analysis at the single-subject level treated each voxel according to SPM's General Linear Model (GLM).

A model-driven approach (the GLM) was used to analyze the BH BOLD response as an alternative to quantification of peak-to-peak amplitudes, because given the short block lengths used here the BOLD response becomes highly sinusoidal in nature as shown in our prior work [see fig. 4 of Thomason et al., 2005], and fitting a simple model is much more robust than quantifying extremal differences in a noisy waveform. While a sinusoidal model could have been used instead of a (negative) boxcar convolved with an HRF, experiments demonstrated that the canonical HRF is a good model for the biphasic BH response. Moreover, the boxcar approach was felt to be a better analog to that employed for WM. Finally, it should be noted that short block designs are not particularly sensitive to the details of the HRF employed in the model.

To test the effectiveness of the modeling approach, we calculated the voxel-wise correlations between the GLM and the BH data timeseries, as determined from the BH activation map T-scores, with the expectation that strong correlation would be evidence of a good model. We also plotted timeseries data from individual ROIs in WM regions, and compared the plots with covariates fit to the data using the GLM and Pearson's correlation analysis.

For second-order (group) analysis, images for each subject were spatially normalized to a common reference gray matter template brain. Calibrated and uncalibrated, normalized images were compared in separate group analyses of variance (ANOVAs) [Friston, 2005; Friston et al., 1995].

Calibration

The WM activation results were calibrated for each subject using their BH scan in accordance with Eq. 4. Let $S_{meas}(i,j)$ be the task signal intensity (effect size) in the i^{th} voxel for the j^{th} subject. Then the corresponding calibrated intensity S_{calib} is given by:

$$S_{calib}(i,j) = S_{meas}(i,j)\left(1 - w(i,j) + w(i,j)\frac{\overline{S}_{BH}}{S_{BH}(i,j)}\right), \quad (6)$$

where

$$\overline{S}_{BH} = \frac{1}{N_{sub}}\sum_{j=1}^{N_{sub}}\left(\sum_{i=1}^{N_{vox}} w(i,j)S_{BH}(i,j) \Big/ \sum_{i=1}^{N_{vox}} w(i,j)\right), \quad (7)$$

N_{sub} is the number of subjects and/or scans in the group, N_{vox} is the number of voxels in the scan and w is a binary mask that is defined by the strength of the BH BOLD signal as well as the T-score for the WM task activation:

$$w(i,j) = \begin{pmatrix} 1, (S_{BH}(i,j) > S_{BH}^{thres}) \cap (T_{WM}(i,j) > T_{WM}^{thres}) \\ 0, otherwise \end{pmatrix}, \quad (8)$$

where T_{WM} is the T-score obtained in the WM task and S_{BH}^{thres} and T_{WM}^{thres} are thresholds for the BH signal and T score of the WM task, respectively. Thus, from Eq. 8 no correction is applied in a given voxel if either the BH BOLD signal or the WM T score are less than their respective thresholds. The thresholds were chosen as 0.5% and 3.5, respectively, after experiments showed that $0.25 \leq S_{BH}^{thres} \leq 1.0$ and $3.0 \leq T_{WM}^{thres} \leq 4.0$ had little effect on the final WM variability results other than the number of voxels included in the analysis.

In this fashion, voxels are corrected independently based on the amplitude of the BH signal. A voxel with BH signal equal to the global average BH signal (Eq. 7) will not be changed, but others will have amplitude either increased or decreased based on the inverse of the BH signal. While the formulation above applies to the BOLD amplitude itself, the calibration can be applied to either the SPM contrast maps or T score maps, since the T score is linearly related to the effect size. Equation 6 was implemented in a MATLAB program that reads the SPM contrast images from the BH scan to obtain S_{BH}, and makes corrections to either the SPM T maps or contrast maps (S_{meas}) for the task images.

Quantification

Average T-values and contrast weighted parameter estimates were measured for voxels that exhibited the greatest response during the WM task. Within each subject's WM network, masks were created by centering spheres with 8-mm radius on subject-specific peak areas. Within those masks the 10 voxels with the highest T-score comprised area ROIs. Correlation between BOLD effect during BH and BOLD effects observed for WM in five primary areas was examined: right and left premotor cortex, right and left

superior parietal cortex, and cingulate cortex. All data were collected on the same day, within the same scan session, and data were not spatially normalized for this analysis. Thus, we effectively assessed BOLD response in regions defined by a cognitive process, in this case working memory, on an individual subject basis, and compared that response to the BOLD effects achieved during BH for the same subject in the same region and scan session. These measurements were used to develop a calibrated WM response. We examined the degree of intersubject variability of effect size in ROIs before and after calibration using an F-test.

In a separate analysis, activation volumes (or extent) during WM were quantified for individual subjects by counting voxels with T-values that exceeded a threshold of 6.0 (corrected $P < 0.0001$). Because activation volumes were found to differ widely across subjects, the effect of calibration was to reduce variability in individual subjects' activation volumes to a modest extent, but the reduction in variability did not reach significance when compared with the variability for uncalibrated activation volumes. Therefore, we also computed a normalized activation volume difference $\Delta V_f(j)$ for the j^{th} individual, defined as $\Delta V_f(j) = (V(j) - \overline{V})/V(j)$, where $\overline{V}$ is the group average over the individual volumes $V(j)$. Thus, ΔV_f measures the difference in activation volume between an individual and the group average, normalized by the individual's activation volume. The purpose of this is to render the impact of calibration more apparent by diminishing the influence of large individual subject volume differences on empirical results. Note that ΔV_f can be both positive and negative. Considerate of this and using this metric, we tested variability across the group before and after calibration using an F-test. In addition to this test of variability, group WM activation volumes were quantified before and after calibration by counting voxels with T-values that exceeded thresholds of 3.5, 4.0, 5.0, and 6.0 (uncorrected P values of 0.008 to 0.0004).

RESULTS

Figure 3 shows an example of a typical measured BH timeseries and the corresponding modeled timeseries. The Pearson's correlation coefficient is 0.88, demonstrating that the GLM provides an excellent representation of the average BH response. In addition, activation maps generated from the BH task resulted in correlation coefficients of >0.60 in individual voxels within the WM regions of interest, again showing high precision in representing the BH BOLD amplitude. This is not surprising, since only voxels with strong BH response were included in the calibration by the threshold procedure. Thus, the strong correlation observed between the data and model both within ROIs and in individual voxels demonstrates the effectiveness of the GLM approach we took to quantifying timeseries-average BH signal amplitude.

Across individual brains, voxels were tested for correlations between BH and WM BOLD response, as demonstrated in Figure 4A for a typical subject. As expected, strong correlations were observed. Furthermore, correlations between WM and BH BOLD amplitudes in ROIs defined by WM resulted in a correlation coefficient of 0.503 ($P < 0.01$) for 35 ROIs (five areas across seven subjects) (Fig. 4B, Table I). Regional differences in WM/BH correlations for areas of interest in the WM network were observed with the following values: right and left parietal cortex (n = 14), $r = 0.532$, $P \leq 0.05$; right and left premotor cortex (n = 14), $r = 0.780$, $P \leq 0.001$; cingulate cortex (n = 7), $r = 0.794$, $P < 0.05$.

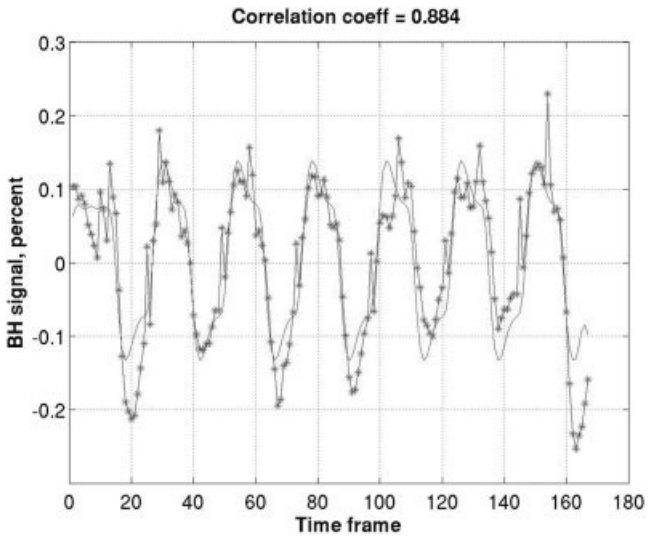

Figure 3.
Measured BH timeseries for an ROI in Brodmann's Areas 4 and 6 for one subject (* where sampled), together with the GLM timeseries. The correlation is 0.88, suggesting that the modeling approach provides accurate quantification of average BH signal amplitude.

Reduction in variance occurred for both BOLD signal *amplitude* measured in WM task-related regions and the *extent* of activation observed in WM statistical parametric maps. Figure 5 demonstrates more homogeneous patterns of amplitude and extent of activation across subjects after calibration. Because the correction was done on a voxel-by-voxel basis in order to maintain regional information about vasoreactivity, the effect was nonuniform suppression or enhancement of activation for particular areas that was subject-specific.

Reduced intersubject variability was confirmed by measurements of effect amplitude in ROIs and whole-brain activation volumes. Tables I and II show the individual BOLD signal change in ROIs and activation volumes, respectively, before and after calibration. For ROI data presented in Table I, the average BOLD response before and after calibration was (0.85 ± 0.55)% and (0.77 ± 0.37)%, respectively, resulting in a 24.8% fractional reduction in intersubject BOLD response variability (i.e., 0.48 ÷ 0.638 = 0.752 reduction in SD/ave). This reduction in variability was significant at $P < 0.03$, although the average effect size itself did not differ significantly. The SD was reduced from 0.55% to 0.37%, i.e., a ratio of 0.68. Given an overall BH/WM correlation of 0.5

(above), this reduction exceeds, but is in reasonable agreement with, the prediction from Eq. 5 of 0.87. Similarly (Table II), WM activation voxel counts before and after calibration showed reduced variability (718 ± 608) and (800 ± 516), respectively. After the individual voxel measurements were normalized for the group average voxel counts as discussed above, the normalized volume difference variability was reduced significantly: 5.35 vs. 1.77, $P < 0.02$.

Second-order group analysis revealed significantly greater extent and magnitude of activation after calibration was performed. This is consistent with the reduced variance observed at the individual subject level. These WM data sets that differed only in whether or not the BH calibration was applied prior to modeling. At $P < 0.001$, a 23.7% greater activation volume was observed with calibration. The net effect is described as a ratio of calibrated to noncalibrated values shown in Table III. The ratio between calibrated and noncalibrated volumes becomes more substantial at higher thresholds. Figure 6 shows group activation maps displayed at the lowest of the T thresholds (3.5) and overlain on a reference anatomical image. Differences between group results with and without calibration were apparent at this threshold, but became increasingly evident at even higher thresholds.

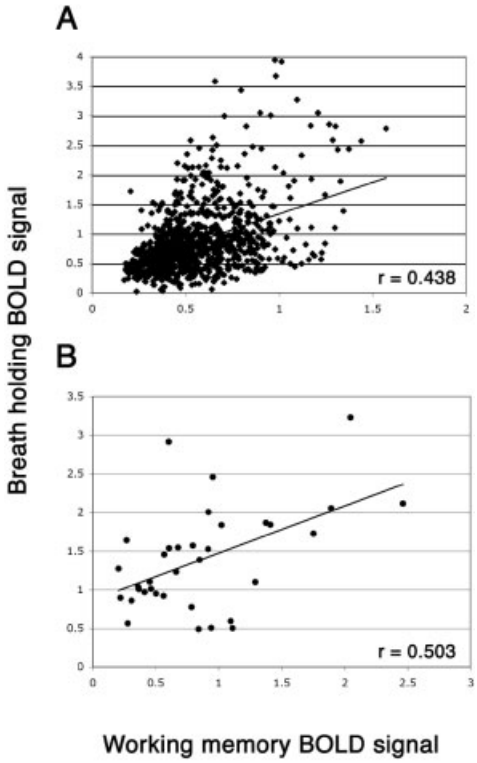

Figure 4.
Scatterplots showing significant correlations between BOLD effects during breath holding (BH) as compared to BOLD effects during working memory (WM) for: (**A**) Subject 3, all voxels (975) with T > 6.0, and (**B**) five regions of interest (ROIs) across seven subjects. ROIs were subject-specific, and defined by WM task activation. BOLD effects for each task were extracted from the same ROIs and shown to scale with one another.

TABLE I. Individual effect sizes for five ROIs in each of 7 subjects before and after calibration

Subject	WM effect size	BH effect size	WM effect size after calibration
1	0.41	0.97	0.59
	0.57	1.46	0.54
	0.31	0.86	0.50
	0.20	1.27	0.20
	0.28	0.56	0.28
2	2.46	2.12	1.61
	1.89	2.06	1.28
	1.41	1.84	1.06
	2.05	3.23	0.88
	1.02	1.84	0.77
3	1.29	1.10	1.62
	1.75	1.73	1.41
	0.85	1.39	0.85
	0.95	2.46	0.54
	0.80	1.57	0.70
4	1.11	0.50	1.11
	0.84	0.49	0.84
	0.45	1.11	0.57
	0.46	1.01	0.63
	0.22	0.90	0.34
5	0.79	0.78	1.40
	1.10	0.59	1.10
	0.94	0.51	0.94
	0.56	0.92	0.85
	0.50	0.95	0.74
6	0.68	1.55	0.61
	0.92	2.01	0.63
	0.36	1.04	0.48
	0.36	1.01	0.50
	0.60	2.92	0.29
7	1.38	1.87	1.02
	0.91	1.53	0.83
	0.61	1.54	0.55
	0.66	1.23	0.74
	0.27	1.64	0.23
Average	0.856	1.387	0.777
SD	0.546	0.662	0.373
SD/average	0.638	0.478	0.480

WM (working memory) and BH (breath holding) effects are correlated at 0.503 ($P < 0.01$). The variability associated with the average effect amplitude across ROIs was 0.64 (SD/average) before calibration, and 0.48 after calibration, for a ratio of 0.75 and a net reduction of 24.8% (significant at $P < 0.03$).

DISCUSSION

As hypothesized, BH BOLD signal appears to be an effective calibration metric. We observed a significant correlation between BOLD response to BH and BOLD response to a cognitive task (WM) that was subject- and region-specific. The calibration procedure presented here relies on the effectiveness of BH BOLD response for characterizing local vasoreactivity without significant cognitive modulation. The strong interrelatedness of WM and BH BOLD response was

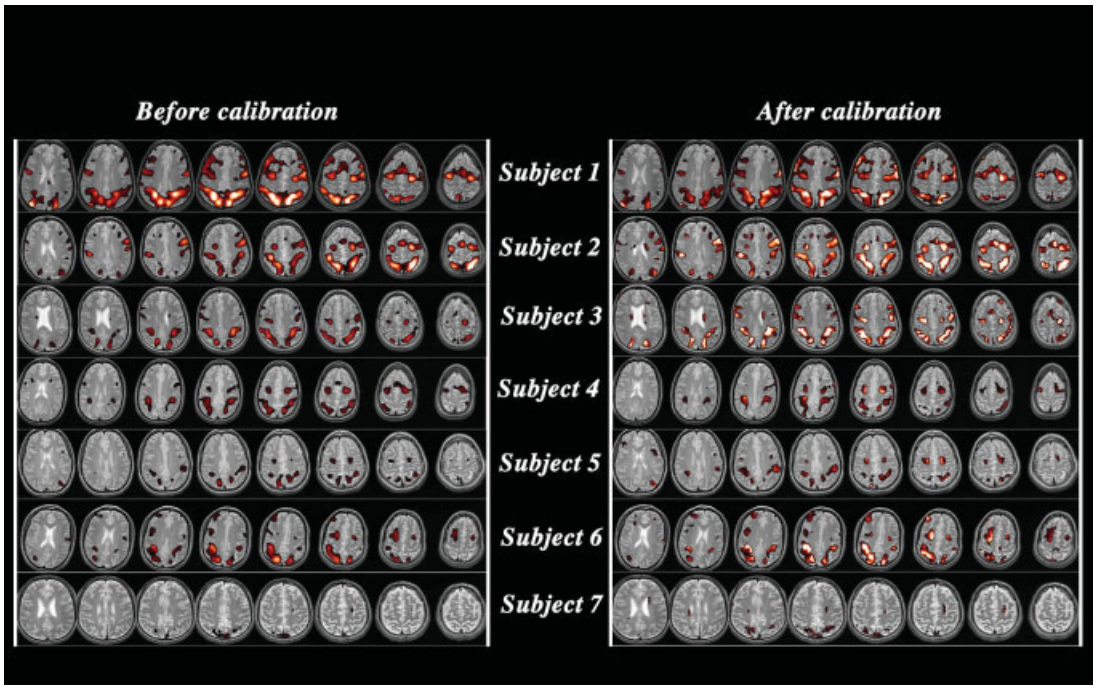

Figure 5.
Individual activation maps for all seven subjects before calibration (left) and after voxel-wise calibration (right). Greater uniformity of activation across subjects is evident in the panels on the right. All activation maps are displayed at a Bonferonni-corrected threshold of $P < 0.004$ ($5 < t < 20$).

an essential confirmation of this theory (Eq. 5). Application of BH-based calibration served to amplify or to suppress regional responses to the WM task in individual subjects. The calibration reduced overall variability in response magnitude and extent across subjects, and the reduction came by eliminating the underlying features of hemodynamic response variations common to both tasks. Intersubject variations in WM BOLD effect were reduced by 24.8%, which was significant at $P < 0.03$ (Table I), and variations in activation extent from group mean were reduced significantly when normalized ($P = 0.016$) (Table II). Thus, by reducing the confound of purely vascular intersubject and intrasubject differences it may be expected that the calibrated data should have a remaining intersubject variance that more accurately reflects neuronal differences between subjects, which in turn should provide more confidence in making inferences about subject or group cognition.

Intrasubject as well as intersubject variations in activation before and after calibration are demonstrated in Table I and Figure 5. Inspection shows the calibration served to make

TABLE II. Subject WM activation volumes summarized in voxel counts at a constant threshold

Subject	WM volume	Calibrated WM volume	Normalized volume difference, ΔV_f	Calibrated normalized volume difference
1	1901	1640	0.62	0.51
2	1032	1192	0.30	0.33
3	637	929	−0.13	0.14
4	458	447	−0.57	−0.79
5	678	848	−0.06	0.06
6	274	398	−1.62	−1.01
7	47	144	−14.28	−4.55
Average	718	800	−2.25	−0.76
SD	608	516	5.35	1.77
SD/average	0.847	0.645		

Calibration significantly decreased overall activation volume differences ($P = 0.016$).
WM, working memory.

TABLE III. Group analysis: whole brain voxel volume data, voxel counts at each of four significance levels

t threshold	3.5	4.0	5.0	6.0
P value	0.008	0.004	0.001	0.0004
No calibration	4126	2641	742	240
Calibrated	4370	2856	918	360
Ratio (calib/noncalib)	1.059	1.081	1.237	1.500

alterations in the WM response amplitude that varied across different brain regions in the WM network within individuals, presumably because of the vasoreactive component common to the two tasks. Qualitatively, the calibrated activation maps appear to be more homogeneous (Fig. 5), and intrasubject differences that remain may more accurately reflect the neuronal differences that occur in the various substrates of WM. The SD values in Table I are the average over all ROIs, which therefore represents both intersubject and intrasubject variability. Because of our small sample size, we were unable to resolve the relative contributions of inter- vs. intrasubject reduction in variability. Future studies with greater statistical power will have the opportunity to approach this issue. In the meantime it is important to note that the 24.8% decrease in variability could be larger or smaller in different WM brain regions.

The WM activation extent (volume) was found to vary widely across subjects before and after calibration, although calibration reduced the differences significantly. The most reduction in intersubject variation of BOLD effect that the BH calibration could have been expected to provide was calculated to be 0.87 based on Eq. 5 given the BH/WM correlation shown in Figure 4B, although the actual reduction was 24.8%. Thus, the large variations in individual activation observed both before and after calibration (Fig. 5) most likely remain dominated by true task performance differences, and the calibrated maps likely represent a more accurate picture of individual subjects' neural behavior in WM. Second-order group analysis of the calibrated data demonstrated significantly greater activation extent in comparison with activation from the original data (Table III, Fig. 6). This was not surprising, given the reduced intersubject variance and our hypothesis that utilizing neurovascular information derived from BH BOLD effect on cognitive task data would provide better study control by partially constraining the confound of neurovascular heterogeneity. Moreover, the significant improvement we report was based on a small group of subjects, and it is expected that the benefit would be even greater with greater measurement power.

While the calibration process significantly reduces intersubject variance in BOLD effect during the cognitive task, and thereby increases group activation, this in itself does not prove that the results provide a more quantitative depiction of neural processes involved in the cognitive task. However, because relatively strong correlation was observed between the BH task and the same subjects' WM task across the group (Fig. 4), and the BH response occurs with no intentional change in $CMRO_2$, we suggest that the constant term S_0 in Eq. 1 contributes undesired variance across subjects that is independent of metabolism. As long as the ratio of flow changes in the two tasks (F in Fig. 1) does not vary greatly across subjects, the calibration should therefore indeed result in response that more closely follows the underlying task-induced metabolism changes. Since the simulations in Figure 1 show that the results are relatively insensitive to F across a broad range of values, the conclusion is felt to be warranted.

A potential concern with the use of a BH task is the reproducibility of performing the maneuver [Nakada et al., 2001; Thomason et al., 2005]. However, our results demonstrate that even if some variance in the BH behavior may occur, the resultant effects were small enough to provide reduced intersubject task activation variations and improved group effects during our study. In the future, attention to means of controlling for such variations may be beneficial.

In conclusion, the present study provides a theory and method for application of a vasoreactivity calibration and a demonstration that its use may provide BOLD response results that reflect reduced influence of hemodynamic differences between brain regions and between subjects. This is the beginning of an improved approach in fMRI studies that may help reduce individual confounds that drive nonrel-

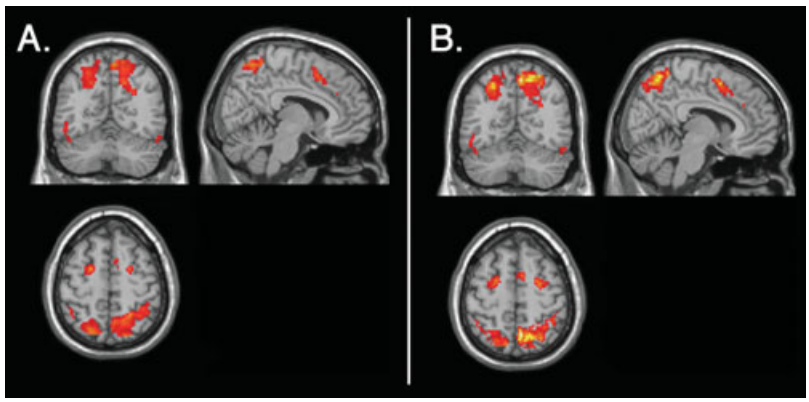

Figure 6.
Group activation maps before (**A**) and after (**B**) calibration. **A:** Coronal, sagittal, and axial sections for group noncalibrated data. **B:** Greater extent and magnitude of activation for the same dataset after correction using BH at the individual subject level, before being submitted to group analysis. Results are displayed at $3.5 < t < 11$ for all maps.

evant differences in BOLD response. Researchers have long cautioned that interpretations of fMRI results are subject to confounding factors as simple as subjects' caffeine intake prior to scanning or lack of sleep the night before. Others have suggested that the presence of phenotypic alterations in brain vasculature in psychiatric populations [Cohen et al., 1995; Curtis et al., 1999] may also influence BOLD response. Utilizing the method proposed here, hemodynamic features of the activation response unrelated to the task of interest may be suppressed, so that subtle response features of interest to cognitive neuroscientists using fMRI may be enhanced. Since the BH task may be used to characterize responsiveness across all cortical areas in a regionally specific way and does not require invasive manipulations, this approach is broadly applicable and simple to implement. Therefore, it may be highly beneficial for use not only in group analyses but in multicenter and longitudinal fMRI studies. Indeed, this approach was introduced by the authors as one of several calibration means that are currently being evaluated in the fBIRN multicenter fMRI study of schizophrenia (http://www.nbirn.net/TestBeds/Function/index.htm).

ACKNOWLEDGMENTS

We thank Tessa B. Johung, Brittany E. Burrows, and Judit Pungor for research assistance, and the anonymous reviewers for valuable comments.

REFERENCES

Aguirre GK, Zarahn E, D'Esposito M (1998): The variability of human, BOLD hemodynamic responses. Neuroimage 8:360–369.

Bandettini PA, Wong EC (1997): A hypercapnia-based normalization method for improved spatial localization of human brain activation with fMRI. NMR Biomed 10:197–203.

Buxton RB, Wong EC, Frank LR (1998): Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn Reson Med 39:855–864.

Cohen BM, Yurgelun-Todd D, English CD, Renshaw PF (1995): Abnormalities of regional distribution of cerebral vasculature in schizophrenia detected by dynamic susceptibility contrast MRI. Am J Psychiatry 152:1801–1803.

Cohen ER, Rostrup E, Sidaros K, Lund TE, Paulson OB, Ugurbil K, Kim SG (2004): Hypercapnic normalization of BOLD fMRI: comparison across field strengths and pulse sequences. Neuroimage 23:613–624.

Curtis CE, Iacono WG, Beiser M (1999): Relationship between nailfold plexus visibility and clinical, neuropsychological, and brain structural measures in schizophrenia. Biol Psychiatry 46:102–109.

Davis TL, Kwong KK, Weisskoff RM, Rosen BR (1998): Calibrated functional MRI: mapping the dynamics of oxidative metabolism. Proc Natl Acad Sci U S A 95:1834–1839.

Fox PT, Raichle ME (1984): Stimulus rate dependence of regional cerebral blood flow in human striate cortex, demonstrated by positron emission tomography. J Neurophysiol 51:1109–1120.

Friston KJ (2005): Models of brain function in neuroimaging. Annu Rev Psychol 56:57–87.

Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SC, Frackowiak RS, Turner R (1995): Analysis of fMRI time-series revisited. Neuroimage 2:45–53.

Glover GH, Lai S (1998): Self-navigated spiral fMRI: interleaved versus single-shot. Magn Reson Med 39:361–368.

Glover GH, Law CS (2001): Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. Magn Reson Med 46: 515–522.

Harms MP, Melcher JR (2003): Detection and quantification of a wide range of fMRI temporal responses using a physiologically-motivated basis set. Hum Brain Mapp 20:168–183.

Hoge RD, Atkinson J, Gill B, Crelier GR, Marrett S, Pike GB (1999): Linear coupling between cerebral blood flow and oxygen consumption in activated human cortex. Proc Natl Acad Sci U S A 96:9403–9408.

Kastrup A, Li TQ, Takahashi A, Glover GH, Moseley ME (1998): Functional magnetic resonance imaging of regional cerebral blood oxygenation changes during breath holding. Stroke 29: 2641–2645.

Kastrup A, Krèuger G, Glover GH, Moseley ME (1999): Assessment of cerebral oxidative metabolism with breath holding and fMRI. Magn Reson Med 42:608–611.

Kastrup A, Kruger G, Neumann-Haefelin T, Glover GH, Moseley ME (2002): Changes of cerebral blood flow, oxygenation, and oxidative metabolism during graded motor activation. Neuroimage 15:74–82.

Kim DH, Adalsteinsson E, Glover G, Spielman D (2002): Regularized higher-order in vivo shimming. Magn Reson Med 48:715–722.

Nakada K, Yoshida D, Fukumoto M, Yoshida S (2001): Chronological analysis of physiological T2* signal change in the cerebrum during breath holding. J Magn Reson Imaging 13:344–351.

Ogawa S, Lee TM, Kay AR, Tank DW (1990): Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci U S A 87:9868–9872.

Preston AR, Thomason ME, Ochsner KN, Cooper JC, Glover GH (2004): Comparison of spiral-in/out and spiral-out BOLD fMRI at 1.5 and 3 T. Neuroimage 21:291–301.

Roland PE (1993): Brain activation. New York: John Wiley & Sons.

Saad ZS, Ropella KM, Cox RW, DeYoe EA (2001): Analysis and use of FMRI response delays. Hum Brain Mapp 13:74–93.

Thomason ME, Burrows BE, Gabrieli JD, Glover GH (2005): Breath holding reveals differences in fMRI BOLD signal in children and adults. Neuroimage 25:824–837.

West JB (1985): Physiological basis of medicine. Baltimore: Williams and Wilkins.