

## Reproducibility of BOLD signal change induced by breath holding

Stefano Magon<sup>a,b</sup>, Gianpaolo Basso<sup>b,c,\*</sup>, Paolo Farace<sup>a</sup>, Giuseppe Kenneth Ricciardi<sup>d</sup>,  
Alberto Beltramello<sup>d</sup>, Andrea Sbarbati<sup>a</sup>

<sup>a</sup> Department of Morphological-Biomedical Sciences, Section of Human Anatomy and Histology, University of Verona, Verona, Italy

<sup>b</sup> Center for Mind/Brain Sciences (CIMeC), University of Trento, via delle Regole, 101 - 38100 Mattarello, (TN) Italy

<sup>c</sup> Department of Cognitive Sciences and Education, University of Trento, Rovereto, Italy

<sup>d</sup> Service of Neuroradiology, Borgo Trento Hospital, Verona, Italy

### ARTICLE INFO

#### Article history:

Received 10 June 2008

Revised 26 November 2008

Accepted 23 December 2008

Available online 13 January 2009

#### Keywords:

Breath holding

Cerebrovascular reactivity

fMRI

BOLD reproducibility

### ABSTRACT

Blood oxygen level dependent (BOLD) contrast is influenced by some physiological factors such as blood flow and blood volume that can be a source of variability in fMRI analysis. Previous studies proposed to use the cerebrovascular response data to normalize or calibrate BOLD maps in order to reduce variability of fMRI data both among brain areas in single subject analysis and across subjects. Breath holding is one of the most widely used methods to investigate the vascular reactivity. However, little is known about the robustness and reproducibility of this procedure. In this study we investigated three different breath holding periods. Subjects were asked to hold their breath for 9, 15 or 21 s in three separate runs and the fMRI protocol was repeated after 15 to 20 days. Our data show that the BOLD response to breath holding after inspiration results in a complex shape due to physiological factors that influence the signal variation with a timing that is highly reproducible. Nevertheless, the reproducibility of the magnitude of the cerebrovascular response to CO<sub>2</sub>, expressed as amplitude of BOLD signal and number of responding voxels, strongly depends on duration of breath holding periods. Breath holding period of 9 s results in high variability of the magnitude of the response while longer breath holding durations produce more robust and reproducible BOLD responses.

© 2009 Elsevier Inc. All rights reserved.

### Introduction

Functional Magnetic Resonance imaging (fMRI) based on Blood Oxygen-Level Dependent (BOLD) contrast (Ogawa et al., 1990) has been extensively used to map brain activity both in healthy subjects and patients. The method exploits the modulation of the local cerebral blood flow (CBF) in response to neural activity (Girouard and Iadecola, 2006; Iadecola, 2004). However, this is a complex process and the relationship between the BOLD response and the neural metabolism is only partially understood since BOLD responses are affected by any cerebrovascular hemodynamic changes (Logothetis, 2002; Logothetis et al., 2001). For example, it has been reported that variations of the baseline vascular state due to vasoactive substances (e.g.: caffeine) (Mulderink et al., 2002), pharmacological agents (Brown et al., 2003), age or brain diseases (Ward et al., 2008; Wu and Hallett, 2005) can alter the BOLD response. Even spontaneous variation of breathing pattern (Wise et al., 2004) or brief periods of breath hold (BH) lead to CBF variations and may be a source of confounds or unmodelled signal variability in fMRI studies (Abbott et al., 2005). Furthermore, it has been noted that the BOLD signal response to these factors is not homogeneous among brain areas (Ito et al., 2000; Wise et al., 2004),

resulting in a variability that reduces the statistical power in group fMRI analysis and complicate the interpretation of single subject results (Handwerker et al., 2004). It has been suggested that at least part of this variability can be taken into account when the cerebrovascular reactivity information is modelled within the analysis aimed at identifying the BOLD responses due to neurovascular coupling. Verifying this assumption, Bandettini and Wong (1997) demonstrated that fMRI BOLD maps can be normalized among brain regions using cerebrovascular reactivity to CO<sub>2</sub> inhalation. A similar approach has been proposed also by Cohen et al. (2004) who were able to significantly reduce differences in fMRI BOLD maps obtained at different MR field strengths. Davis et al. (1998) have extended this approach proposing a method based on cerebrovascular reactivity elicited by CO<sub>2</sub> inhalation as a reference to calibrate BOLD fMRI in order to extract quantitative Cerebral Metabolic Rate of Oxygen (CMRO<sub>2</sub>) values.

Breath holding (BH) has been proposed as an alternative method to CO<sub>2</sub> inhalation to generate cerebrovascular reactivity maps (Corfield et al., 2001; Kastrup et al., 1999a,b; Li et al., 1999, 2000; Liu et al., 2002). Under the assumption that CMRO<sub>2</sub> remains invariant, the BH-task leads to a reduction of oxygen and an accumulation of carbon dioxide in the blood stream. The consequent hypercapnia induces cerebral vasodilatation and increased CBF (Markus and Harrison, 1992). Kastrup et al. (2001) showed similarity between BH task and CO<sub>2</sub> inhalation in evaluating the hemodynamic reserve capacity with

\* Corresponding author. Center for Mind/Brain Sciences (CIMeC), University of Trento, via delle Regole, 101 - 38100 Mattarello, (TN) Italy. Fax: +39 0461 88 3066.

E-mail address: [gianpaolo.basso@unitn.it](mailto:gianpaolo.basso@unitn.it) (G. Basso).

BOLD fMRI. However, the BH method is simpler and less invasive than CO<sub>2</sub> inhalation. Thus, it could be applied in almost any clinical or experimental setting. Previous studies showed that significant BOLD responses could be measured during BH with duration as short as 10 s (Liu et al., 2002). Kastrop et al. (1998) studied different breath holding duration (18 s, 30 s, and 40 s) and report BOLD signal changes between 0.8% and 3.5%. Thomason et al. (2005) were able to use a breath hold task even in children, demonstrating significant differences in BOLD cerebrovascular reactivity maps between children and adults. Hence, overall there are many evidences on the efficacy of the breath hold task to generate cerebrovascular reactivity maps. In fact, Thomason et al. (2007) successfully applied the calibration procedure using breath hold to probe individual cerebrovascular reactivity.

In order to use BOLD cerebrovascular reactivity maps to calibrate and/or normalize fMRI data, however, it is necessary to know how much these maps are reliable. One way to assess the reliability of a method is to estimate its reproducibility under the same conditions. To our knowledge, so far no studies have explored the reproducibility of the BOLD BH-induced response. Some studies investigated the reproducibility of BOLD signal changes induced by motor tasks (Havel et al., 2006; Liu et al. 2004), sensory tasks (Peelen and Downing, 2005; Vlieger et al., 2003; Yetkin et al., 1996) and cognitive paradigms (Chee et al., 2003; Fernandez et al., 2003; Harrington et al., 2006; Rau et al., 2007).

The most used way to estimate the reproducibility of BOLD response is the correlation between repeated measurements computed as the Pearson product moment correlation coefficient (Fernandez et al., 2003; Jansen et al., 2006) or the intraclass correlation analysis (ICC) (Wei et al., 2004; Raemaekers et al., 2007). The Pearson correlation coefficient measures the strength of a relation between two variables but it is insensitive to the absolute values of the variation. Thus, it does not provide information about the agreement between two measures (Bland and Altman, 1986). The ICC applied to fMRI studies is used to express how consistent is the activation of each subject relative to the rest of the group over different sessions (Zandbelt et al., 2008). Different types of ICC methods must be applied depending on how the variance of the measurements is estimated (e.g., one-way vs. two-way analysis of variance) (McGraw and Wong, 1996). Each ICC method may produce different results, and different methods have been implemented to investigate the reproducibility in fMRI (Fernandez et al., 2003; Raemaekers et al., 2007; Specht et al., 2003). Furthermore, both ICC and Pearson correlation methods are sensitive to the spread of values in the sample (Bland and Altman, 1990). Since correlation methods give more information about the trend of variation than about the repeatability of the measures, additional methods must be considered to estimate the reproducibility of a measure. The coefficient of variation (CV), expressed as the ratio between the standard deviation and the mean, has been proposed as an index of reproducibility of a measure (Marshall et al., 2004; Tjandra et al., 2005) and in fact it has been used to evaluate the reproducibility of cerebrovascular reactivity maps elicited by CO<sub>2</sub> inhalation; (Leontiev and Buxton, 2007). Its major advantage is that it represents a scale invariant measure of inequality (Allison, 1978).

To investigate the reproducibility of the BOLD signal, different parameters may be taken into account. To date, each study on the reproducibility has focussed only on some aspects of the BOLD response such as the temporal dynamic of the response (Neumann et al., 2003), the signal intensity (Rombouts et al., 1998), the size of activation (Rombouts et al., 1998) and the voxel-by-voxel *t*-value (Fernandez et al., 2003). In the present work we investigate the variability and reproducibility of many different parameters related both to the temporal dynamic and to the magnitude of BOLD response within and across subjects using both the Pearson correlation method and the coefficient of variation. We manipulated the BH duration, in order to identify the most suitable BH paradigm to obtain a reproducible BOLD cerebrovascular reactivity map.

## Methods

### Subjects

Data were collected from 15 healthy subjects who gave their informed consent. Four subjects were excluded from the analysis. One subject was excluded because of excessive head movement, one subject was excluded because his spontaneous breathing during the MR scanning session was irregular (self-report of state of anxiety) and two subjects did not perform the second scanning session. The experimental group was made up of 11 subjects (5 males, 6 females, age: 31.7±8.3; range 20–42).

### MRI acquisition

Images were acquired on a 3.0 T MR head scanner (Siemens, Erlangen, Germany). During each functional run 126 brain volumes were acquired with a T2\*-weighted Gradient Echo (GE) EPI sequence (TR=3000 ms, TE=30 ms, flip angle=90°, 64×64 acquisition matrix, FOV=192×192, number of slices=36 interleaved transversal, slice thickness=3 mm, gap=10%, voxel size=3 mm<sup>3</sup>, readout bandwidth per pixel 1980 kHz). High resolution anatomical images were obtained using a T1-weighted 3D brain volume (TR=2300 ms, TE=3.93 ms, flip angle=90°, 256×256 in-plane resolution, FOV=256, slice thickness=1 mm, number of slices=160 sagittal slices).

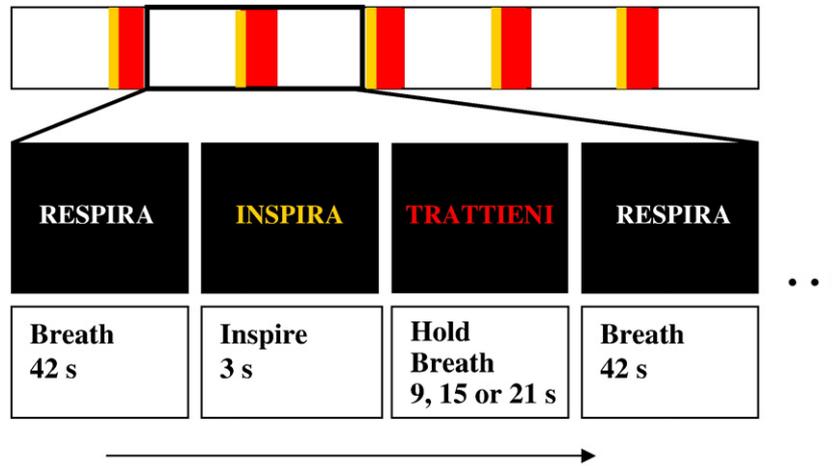
### fMRI protocol

Subjects were asked to hold their breath, after inspiration, for 9, 15 or 21 s (BH9, BH15 and BH21) during three separate runs. Within each run the breath hold duration was the same, either 9 s or 15 s or 21 s. Each run started with two dummy scans lasting 3 s each, followed by 42 s of self paced normal breathing. After this, subjects performed 5 breath hold periods, alternating them with 42 s of self-paced breathing (Fig. 1). Each breath hold period was preceded by an inspiration to be performed within 3 s after the normal breathing period ended. After the last normal breathing period an additional period of spontaneous breathing was added to the BH9 and BH15 runs lasting 60 s and 30 s respectively. This was necessary in order to keep the number of acquired scans, a total of 126 for each run, identical across the three runs. A coloured word was projected over a black screen to assist the subject with timing breath during the run. The subject had to breathe normally while the word “respira” (breathe) was presented on the screen in white. A word “inspira” (inspire) was displayed for 3 s in yellow instructing the subject to perform an inspiration. The subject had to hold their breath as soon as the red word “trattieni” (hold breath) appeared and for its entire display duration. The order of the three runs was balanced across subjects. Subjects repeated the same protocol after 15–20 days.

The task was created using E-Prime (Psychology Software Tools, Inc., Pittsburgh, USA). Presentation of stimuli was synchronized with the MR acquisition using the Integrated Functional Imaging System (MRI Devices Corporation, Gainesville, USA).

Subjects were instructed outside the scanner not to move their head both during breathing and whilst holding their breath. Moreover, they were trained on each breath holding duration to make the inspirations preceding each breath hold as equal as possible both in duration and chest expansion, and to avoid abdominal muscle contraction at the end of the inspiration. An additional short training session was performed after the subject was positioned inside the scanner and head movement was restricted using soft foam pads. Moreover, subjects were reminded of the length of the breath hold duration at the start of each run.

Subject's compliance to the task and ability to hold the breath during the scanning sessions were monitored using an elastic respiratory belt placed around the base of the chest. Subjects were



**Fig. 1.** The fMRI protocol is represented as a bar task that indicates the five task epochs that compose the run. All subjects performed the three breath holding periods (9 s, 15 s and 21 s) in separate runs. For each run, each subject performed five periods of BH alternating them with 42 s of self-paced breathing. A single epoch has been magnified to point out the visual cue that guided the subjects during the run (white: “RESPIRA” breathe, yellow: “INSPIRA” inspire and red: “TRATTIENI” hold breath).

also requested to report about their performance at the end of each scanning session.

#### *fMRI data pre-processing*

Anatomical and functional images were processed using Brain-VoyagerQX 1.9.10 (Brain Innovations, Maastricht, The Netherlands). Anatomical data (T1-3D weighted images) were corrected for  $B_1$  field inhomogeneities and transformed into AC-PC and Talairach standard space (Talairach and Tournoux, 1988). Functional data were pre-processed as followed: 1) the first 6 s (2 scans) of each run were discarded to ensure steady-state of the longitudinal magnetization; 2) slice scan timing correction was performed with sinc-interpolation; 3) 3D motion correction with intra-session alignment was performed (six parameter rigid transformation, 3 rotations and 3 translations). Estimated translation and rotation parameters were inspected to exclude subjects with head movement that exceeded 3 mm or  $2^\circ$ ; 4) spatial smoothing (full width at half-maximum=5) and temporal filtering were applied (high-pass filter 0.006 Hz and linear trend removal) in order to remove baseline drifts. The functional time series data of each subject were first co-registered with the subject's 3-D anatomical dataset, followed by the application of the transformation matrices derived from the spatial transformation steps performed for the 3-D anatomical dataset. These steps resulted in normalized 4-D volume time courses data (Goebel et al., 2006). For each subject the cortical grey matter (GM) was segmented on the T1-3D using the Advanced Segmentation Tools build in BrainVoyager. The core of these tools is an adaptive calculation of the intensity thresholds to localize the boundaries between cortical GM and white matter (WM). As a result, the cortical grey matter can be reliably separated from WM, cerebrospinal fluid and subcortical grey matter structures. In order to achieve the best results, we applied all the required pre-processing steps including signal inhomogeneity correction and enhancement of tissue contrast.

Furthermore, for each subject the portion of the segmented cortical grey matter laying above the AC-PC plane was divided into subregions of interest defined using the proportional grid system described by Talairach and Tournoux (1988), based on which, the entire brain is divided into 864 adjacent sectors. We applied the same proportional grid to the segmented cortical grey matter, halving, however, the resolution of the proportional grid along each of the X, Y and Z axes, apart from the sectors comprised between the AC and PC vertical planes, where the resolution was halved only along the X and Z axes (Fig. 2a). The grey matter contained in each of the resulting 80

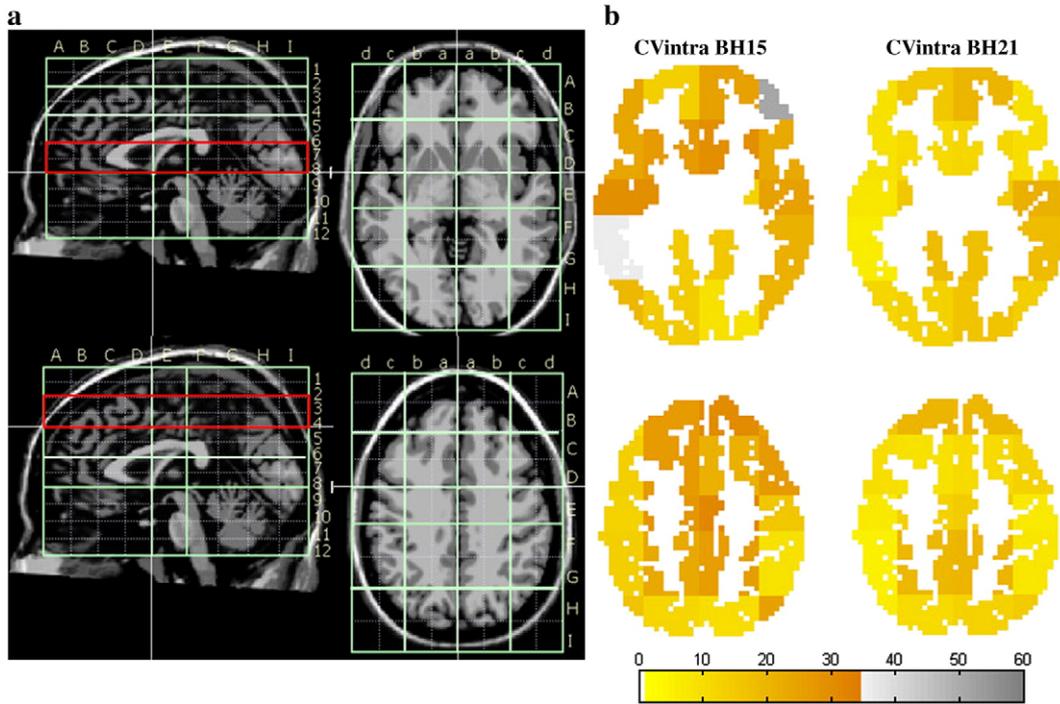
subregions laying above the AC-PC plane was then analyzed separately, discarding all subregions which did not contain any cortical grey matter. This process resulted in 76 subregions since the 4 subregions located at the uppermost vertex of our modified grid system did not contain any grey matter.

#### *Data analysis*

We limited our analyses to the cortical GM that was derived for every single subject based on individual anatomical information and defined it as the region of interest (ROI).

For each run of each subject we averaged the signal timecourse of all voxels included in the ROI. Then, the average value of the 6 s (two scans) preceding each trial was used as a baseline over which we computed the percentage signal change of the timecourse. The timecourse of each run was then segmented into 5 cycles, each starting 6 s (two scans) before the instruction and ending when the following instruction to inspire was presented. The percentage signal change of the 5 cycles was averaged. We selected the most positive and the most negative value (PSC) occurring from the start of the breath hold period. We also computed the Time-to-Peak (TTP) multiplying by 3 s the number of scans occurring from the start of the breath hold period to the positive and the negative PSC. Finally, we computed the integral of the area subtended by the curve around each PSC (Area) setting the limit of the area to be integrated between the points before and after the PSC in the time course where the signal reached the zero or the cycle ended. All calculations were performed using in-house developed routines (MATLAB 7.0, The Mathworks Inc., Natick, USA).

A voxelwise model-driven approach (General Linear Model, GLM) was applied to calculate Beta coefficient and the number of voxels in which signal variation was significant. In order to create the GLM regressors, the block design model of the BH period was convolved with the hemodynamic response function (HRF) commonly adopted for cognitive fMRI studies (Glover, 1999). This HRF is a two-gamma function with a positive time peak at 5.5 s. We expected an additional delay in the MR signal change induced by the BH. This delay depends on many physiological factors related for example to pulmonary perfusion, pulmonary ventilation and  $\text{CO}_2$  sensor reactivity, and cannot be predicted a priori for each subject and could, in theory, also vary between runs. For this reason we performed a prior exploratory analysis for each run using a boxcar function convolved with a single-gamma HRF (Boynton et al., 1996). A cross-correlation analysis shifting this function in time (10 steps, 3 s each) allowed us to



**Fig. 2.** (A) represents the proportional grid system used to divide the brain above AC-PC plane into 80 subregions. (B) shows the  $CV_{intra}$  value of 40 subregions within two axial planes for BH15 and BH21.

identify the time shift (Lag) at which a voxel responded maximally (correlation threshold:  $p < 0.05$ ). We computed the average time shift of all voxels in the cortical grey matter ROI and used this value to shift the two-gamma function modelled in the GLM analysis in time.

Parameter estimates of the GLM (Beta coefficients) were computed for each voxel. The number of voxels with significant signal variation was computed from  $t$ -test statistical parametric maps of signal change during BH compared to normal breathing. The percentage of voxels surviving a threshold of  $p < 0.05$  corrected for multiple comparison (Bonferroni correction) was computed (Volume).

Analysis of variance for repeated measures (1 factor: BH duration; 3 levels: 9 s, 15 s, and 21 s) was implemented using SPSS 13 (SPSS Inc., Chicago, USA). The model was applied separately to all computed parameters (PSC, TTP, Area, Lag, Volume, and Beta). The same model was also applied to compare the signal differences during the three BH durations, separately for each time point of the curves.

We estimated both inter-subjects and intra-subjects variability computing the coefficient of variation (CV). The CV is defined as the standard deviation ( $\sigma$ ) of a measurement normalized to its mean ( $x_{avg}$ ) expressed as a percentage (Eq. 1) and gives information about the variability of the measurements, representing the dispersion of a probability distribution. A CV value of 33% can be used as the upper fiducial limit for acceptable variability in a normal distribution (Johnson and Welch, 1940).

$$CV = 100 \frac{\sigma}{x_{avg}} \quad (1)$$

In order to investigate the reproducibility of MR signal change induced by different BH durations, for each subject we computed the fractional difference between the two measurements on different days, defined as a percentage difference relative to the mean ( $\Delta$ ) (Eq. 2);

$$\Delta = 100 \frac{(x_1 - x_2)}{x_{avg}} \quad (2)$$

where  $x_1$  and  $x_2$  are measurements acquired on day 1 and day 2, respectively, and  $x_{avg}$  is the mean value of these measurements. The

standard deviation of  $\Delta$  ( $\sigma_{\Delta}$ ) contains information about the reproducibility of the measurement (Leontiev and Buxton, 2007). However, we also computed the intra-subjects variability ( $CV_{intra}$ ) defined as:

$$CV_{intra} = 100 \frac{\sigma_{\Delta}}{\sqrt{2}} \quad (3)$$

where the  $\sigma_{\Delta}$  has been adjusted dividing it by square root of 2 (Eq. 3) (Leontiev and Buxton, 2007) because the difference between two independent variables has a probability distribution with a variance that is the sum of the variance of the two separate distribution probabilities. In our data the two independent measures were extracted from the same distribution of probability, thus the variance of resulting sample is equal to two times the variance of the original distribution of probability. Thus, the reproducibility of a measure was assessed with a one-way analysis of variance for repeated measures with three levels (BH9, BH15 and BH21) to determine if  $\Delta$  differed significantly between the three BH durations.

In order to estimate the trial-by-trial reliability, we computed the mean and standard deviation of the peak percentage signal change of each trial. The peak has been defined as the most positive percentage variation occurring in the time interval comprised from the start of the breath hold period of each trial to the start of the next trial. The inter-trial coefficient of variation was obtained dividing the standard deviation by the mean. Thus, for each subject, we obtained 3 inter-trial coefficient of variation per session and, one for each breath holding duration.

To evaluate the inter-subjects variability, a coefficient of variation ( $CV_{inter}$ , Eq. 1) was computed separately for each session and used as an estimator of the variability of the population in performing the breath hold task at different durations.

Furthermore, the  $\Delta$  factor, the  $CV_{intra}$  and the  $CV_{inter}$  of the most positive PSC were computed separately for each of the 76 cortical grey matter subregions. A single sample  $t$ -test was performed to investigate if the average  $CV_{intra}$  of the subregions was above or below 33%. In order to investigate the possible interaction between the breath hold

duration and the subregions, we performed the analysis of variance for repeated measures of  $\Delta$  factor. The BH (15 s and 21 s) was used as within-subjects variable while subregions was used as between-subjects factor. Moreover, a probability density plot was created to inspect the distribution of the  $CV_{\text{inter}}$  among the 76 subregions.

To estimate the reliability of  $t$ -values (computed from the voxelwise analysis) of the two measures, the Pearson correlation coefficient ( $r$ ) was computed for each subject between voxel-by-voxel  $t$ -values pertaining to the first and second sessions.

## Results

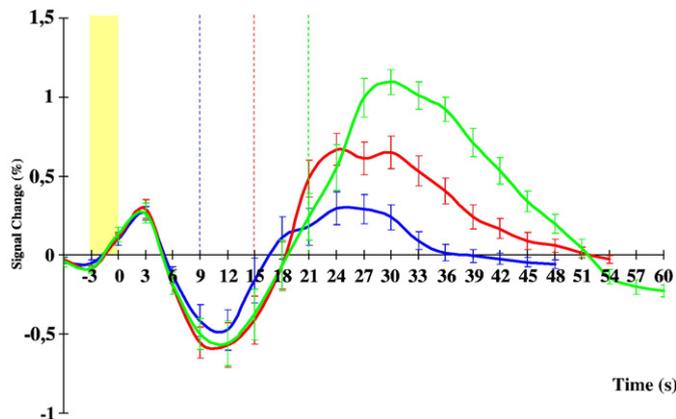
All subjects were able to hold their breath during all runs without discomfort. However, holding the breath for 21 s was reported as more difficult than all other BH durations and the most evident chest movements were noticed during this condition.

### Analysis of breath holding signal changes

The curves of the mean fMRI signal change on the cortical grey matter ROI showed a triphasic structure for all BH durations (Fig. 3). After an initial short positive phase, the signal became negative and started increasing after 10 s, becoming again positive between 18 and 21 s after the start of the breath hold, reaching the maximum peak within a variable delay depending on BH duration. Then, the signal smoothly decreased back to baseline with a variable delay. The three portions of the curve have been analyzed separately.

The analysis of variance for the first positive portion of the curves did not show significant differences between BH durations. The analysis of variance for the negative portion of the curves presented significant differences in PSC ( $F(1, 10)=6.13$ ,  $MSE=0.01$ ,  $p<0.05$ ) and Area ( $F(1, 10)=8.80$ ,  $MSE=1.89$ ,  $p<0.05$ ), whereas no significant differences were observed in TTP ( $F(1, 10)=0.185$ ,  $MSE=2.20$ ,  $p=0.67$ ) (Table 1). The mean comparison showed that the minimum negative peak and the subtended area were significantly smaller for the 9 s BH duration than the 15 s (PSC:  $p<0.05$ ; Area:  $p<0.05$ ) and the 21 s BH duration (PSC:  $p<0.05$ ; Area:  $p<0.05$ ). On the contrary, no significant differences were observed between BH15 and BH21. The BH duration had a strong influence on all parameters of the third, positive portion of the curves (Table 2) with PSC ( $F(1, 9)=111.81$ ,  $MSE=0.023$ ,  $p<0.001$ ), Area ( $F(1, 9)=186.28$ ,  $MSE=7.23$ ,  $p<0.001$ ) and TTP ( $F(1, 9)=11.62$ ,  $MSE=12.55$ ,  $p<0.01$ ) being greater for longer durations.

The point-by-point analysis showed a significant difference ( $p<0.001$ ) between the curves for all timepoints sampled from the TTP of the negative portion of the curves up to the end of the curves.



**Fig. 3.** Average signal change during the three BH durations (BH9=breath holding 9 s; BH15=breath holding 15 s and BH21=breath holding 21 s) across subjects. The yellow bar indicates the inspiration phase. The BH phase started at time zero; the dotted lines indicate the end of the BH period. The error bars indicate the standard error of the group mean.

**Table 1**

Parameters estimated from the negative portion of the curves

	BH9	BH15	BH21
Area*	4.18±4.77	5.88±4.78	5.98±5.5
PSC*	-0.51±0.38	-0.63±0.41	-0.62±0.42
TTP	10.63±2.06	10.09±2.42	10.36±2.46

Mean and standard deviation of computed parameters (Area=integral of the area subtended by curve; PSC=Percent Signal Change; TTP=Time-to-Peak). An asterisk indicates overall comparison significant at  $p<0.05$ .

No overall statistical significant differences were found for all preceding timepoints. Significantly higher values were found for all timepoints sampled after 15 s from the onset of the breath hold for BH15 compared to BH9 (timepoints between 15 and 36 s:  $p<0.001$ ), timepoints between 39 and 45 s:  $p<0.01$ ; timepoints between 48 and 51 s ( $p<0.05$ ). Significantly higher values during BH21 compared to BH9 were found for all timepoints sampled after 24 s from the onset of the breath hold and for the timepoint sampled after 15 s from the onset of the breath hold ( $p<0.001$ ). Significantly higher values during BH21 compared to BH15 were found for all timepoints sampled after 27 s ( $p<0.001$ ). Thus, inconsistent differences were found for those timepoints sampled while the signal of each curve was rising from the minimum negative peak to the maximum. In particular, during this phase no differences were present between BH15 and BH21 while BH9 differed from both other curves. Thus, most of the differences induced by breath hold duration were evident during the third portion of the curves. As a confirmation of the point-by-point analysis, during the third portion of the curves also the delay of the hemodynamic response (Lag:  $F(1, 10)=20.01$ ,  $MSE=1.28$ ,  $p<0.01$ ), the percentage of responding cortical grey matter (Volume:  $F(1, 10)=262.34$ ,  $MSE=114.54$ ,  $p<0.001$ ) and the parameter estimated by the GLM analysis (Beta:  $F(1, 10)=77.52$ ,  $MSE=0.022$ ,  $p<0.001$ ) were higher as the BH duration was longer (Table 2 and Fig. 4).

### Breath holding reproducibility

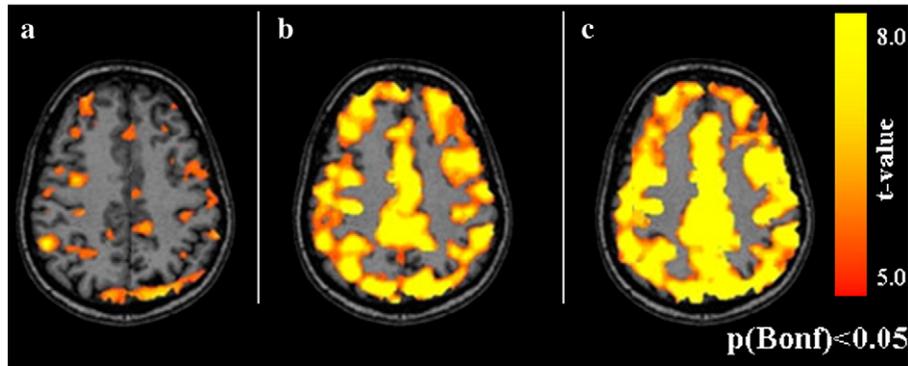
The reproducibility of the MR signal change induced by breath hold depended on duration. Table 3 shows the fractional difference between measurements acquired across different days with the corresponding  $CV_{\text{intra}}$  value. The  $CV_{\text{intra}}$  for TTP and Lag falls under 10%, indicating a very low variability of  $\Delta$  of the timing of the BOLD response across sessions for all BH durations. The  $\Delta$  of the magnitude of the BOLD response had a low variability for the BH15 and BH21 that were associated with a  $CV_{\text{intra}}$  value below the fiducial level of 33% for Area, PSC, Volume and Beta. Notably, the BH9 was associated with a  $CV_{\text{intra}}$  value well above the fiducial level for the Area and Volume parameters, indicating a significant variability of the magnitude of the BOLD response across sessions for the shortest BH duration. The

**Table 2**

Parameters estimated from the third positive portion of the curves and voxelwise analysis

	BH9	BH15	BH21
Area***	5.64±3.57	13.05±5.82	21.33±6.18
PSC***	0.43±0.24	0.80±0.27	1.17±0.25
TTP**	25.95±3.81	31.09±4.3	34.36±3.32
Volume***	4.59±4.34	33.54±22.64	61.09±18.54
Beta***	0.46±0.196	0.77±0.191	1.03±0.186
Lag**	16.14±2.54	17.06±2.53	18.30±2.07

Mean and standard deviation of parameters computed from the analysis of the average resulting curves (Area=integral of the area subtended by curve; PSC=Percent Signal Change; TTP=Time-to-Peak) and parameters computed from voxelwise analysis (Volume=percentage of cortical grey matter voxels in which variation was significant; Beta=General Linear Model estimated coefficient; and Lag=time shift at which a voxel responded maximally). The number of asterisks denotes overall comparison significance: \* $p<0.05$ ; \*\* $p<0.01$ ; \*\*\* $p<0.001$ .



**Fig. 4.** Single subject example of number of voxels that responds significantly (Volume) in the three breath holding durations, (a) 9 s (b) 15 s and (c) 21 s.

repeated measures analysis of variance, implemented using the  $\Delta$  as an independent measure and the breath hold duration as within-subjects factor with three levels, confirmed that breath hold duration had a significant effect on the magnitude of the BOLD response (PSC ( $F(1,9)=11.70$ ,  $MSE=910.29$ ,  $p<0.01$ ) Area ( $F(1,9)=12.15$ ,  $MSE=1298.73$ ,  $p<0.01$ ), Beta ( $F(1,10)=5.09$ ,  $MSE=426.25$ ,  $p<0.05$ ) and Volume ( $F(1,10)=14.62$ ,  $MSE=1967.54$ ,  $p<0.01$ )) while no significant effect was found for the TTP and the Lag. In particular, the means comparison of each BH duration against each other confirmed that the  $\Delta$  value for the Volume parameter was significantly higher for BH9 compared to BH15 ( $p<0.05$ ) and BH21 ( $p<0.01$ ) while no significant difference was found between BH15 and BH21. BH9 was associated with a significantly higher Delta value also for PSC ( $p<0.01$ ), Area ( $p<0.01$ ) and Beta ( $p<0.05$ ) compared to BH21.

Fig. 5 shows the  $CV_{inter}$  values for both sessions. The  $CV_{inter}$  value for TTP and Lag was within an acceptable range of variability for all durations, with BH21 and BH15 showing the lowest variability of the timing of the response. For what concerns PSC and Beta, the  $CV_{inter}$  of BH9 was associated with the highest variability while the BH21 was associated with the lowest variability. The  $CV_{inter}$  of BH15 was close to 33% in both sessions. These results suggest that the BH21 is associated with the lowest dispersion of the magnitude of the BOLD response.

An analysis of variance for repeated measures revealed that the inter-trial coefficient of variation was not significantly different across sessions ( $F(1,10)=0.55$ ,  $MSE=479.18$ ,  $p=0.47$ ) while it was significantly affected by the breath hold duration ( $F(1,10)=6.86$ ,  $MSE=1079.03$ ,  $p=0.02$ ). Thus we collapsed the inter-trial coefficient of variations across the 2 sessions and performed a second analysis of variance to explore how the breath hold duration affected the inter-trial variability. The mean inter-trial coefficient of variation was significantly different between all durations (BH9 vs. BH15:  $p<0.05$ ; BH9 vs.

BH21:  $p<0.01$ ; BH15 vs. BH21:  $p<0.01$ ) becoming lower as the BH duration increased (Fig. 6). However, it must be noted that the dispersion of the CV inter-trial for BH9 was high (mean: 38.08%; range: 8.73%–207.08%) compared to the response during BH15 (mean: 19.59%; range: 1.18%–40.98%) and BH21 (mean: 12.14%; range: 2.52%–37.52%), with some subjects showing a very variable BOLD percent signal change across trials. This indicates that the BOLD signal response across trials during the BH9 was overall much more variable across subjects.

We limited the analysis of the 76 cortical grey matter subregions to the  $\Delta$  factor, the  $CV_{intra}$  and  $CV_{inter}$  of the most positive PSC since this parameter is the most widely used for normalization or calibration procedures. Also, we restricted the subregions analysis to BH15 and BH21 since the analysis of the BH9 was already associated with a very high variability of the magnitude of the response across trials and sessions when the entire cortical grey matter ROI was considered.

The analysis of variance of  $\Delta$  factor shows a significant effect of breath hold duration ( $F(1,750)=75.42$ ,  $MSE=806.97$ ,  $p<0.001$ ) and a significant effect of subregions factor ( $F(75,750)=75.42$ ,  $MSE=1264.65$ ,  $p<0.05$ ). On the contrary, the interaction between BH duration and subregions did not show a significant effect ( $F(75,750)=0.751$ ,  $MSE=806.97$ ,  $p=0.94$ ). The average and standard deviation of  $CV_{intra}$  of the 76 subregions was  $23\pm 9$  and  $15\pm 8$  for BH15 and BH21 respectively. The single sample  $t$ -test analysis showed that the  $CV_{intra}$  was significantly lower than 33% for both BH duration (for BH15  $t=-17.7$ ;  $df=75$ ; two-tailed  $p<0.001$  and for BH21  $t=-8.8$ ;  $df=75$ ; two-tailed  $p<0.001$ ). Moreover, the cumulative probability distribution of the  $CV_{intra}$  value of all subregions (Fig. 7) demonstrates that 88% and 97% of subregions for BH15 and BH21 respectively were associated with a value below 33%. Fig. 2b displays a coloured map of the  $CV_{intra}$  for 40 subregions within two axial planes. Overall, these results suggest a comparable and good reproducibility of the magnitude of the response for BH15 and BH21 even at a subregion level of analysis.

The density distribution of the  $CV_{inter}$  of the subregions (Fig. 8) shows a higher dispersion of values across subregions for BH15 compared to BH21 within both sessions, as 57 and 62 regions showing a higher  $CV_{inter}$  for BH15 compared to BH21 within the first and the second sessions respectively.

Due to the method that we applied, each subregion contained a different number of voxels ranging between 4 and 856. However, a correlation analysis between the number of voxels within each subregion and the corresponding CV values resulted in a very low value ( $r=-0.27$ ) which excludes that the number of voxels may be a major bias in the analysis.

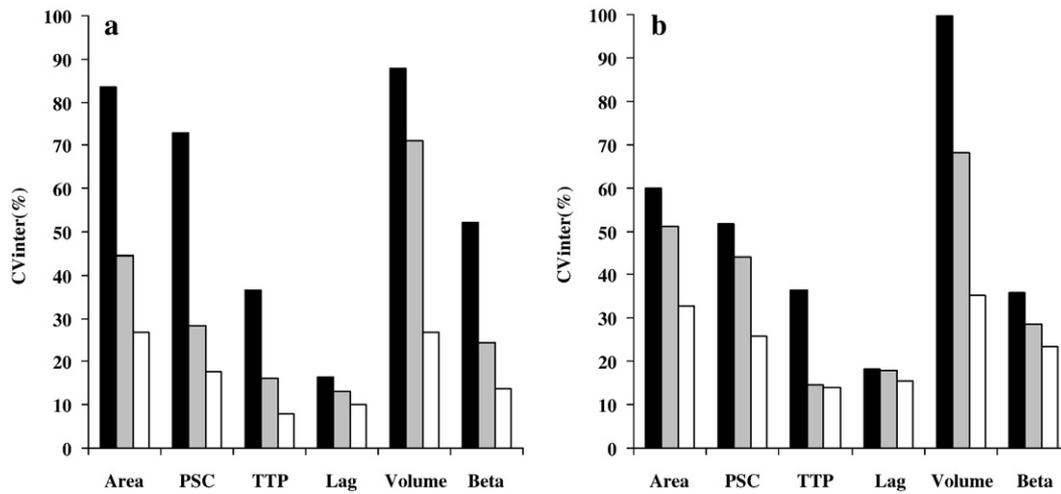
The Pearson test between  $t$ -values of all voxels of cortical grey matter ROI relative to the first and second sessions showed an increase of the correlation coefficient as BH duration increased. Group mean value showed that the correlation coefficient for BH9 was 0.44 (ranging from 0.15 to 0.54; mean  $p<0.001$ ), for BH15 was 0.55

**Table 3**

Variability between sessions: Delta factors and intra-subject coefficient of variation

	A			B		
	Delta			$CV_{intra}$		
	BH9	BH15	BH21	BH9	BH15	BH21
Area**	62.71 ± 55.38	34.52 ± 41.69	16.47 ± 15.30	39	29	11
PSC**	50.96 ± 41.22	31.39 ± 30.72	11.75 ± 8.51	29	22	6
TTP	8.22 ± 8.31	10.43 ± 8.10	9.24 ± 7.62	6	6	5
Volume**	94.92 ± 64.40	42.10 ± 32.44	22.58 ± 17.40	45	23	12
Beta*	30.91 ± 32.14	17.48 ± 7.16	11.04 ± 10.17	23	5	7
Lag	12.41 ± 8.49	9.02 ± 8.97	11.43 ± 8.18	6	6	6

Column A reports the mean and standard deviation of the  $\Delta$  factor for each parameter. The number of asterisks denotes overall comparison significance for Delta: \* $p<0.05$ ; \*\* $p<0.01$ ; \*\*\* $p<0.001$ . Column B reports the coefficient of variation between sessions ( $CV_{intra}$ ) for each parameter. (Area=integral of the area subtended by curve; PSC=Percent Signal Change; TTP=Time-to-Peak; Volume=percentage of cortical grey matter voxels in which variation was significant; Beta=General Linear Model estimated coefficient; and Lag=time shift at which a voxel responded maximally).



**Fig. 5.** Inter-subjects variability for each parameter for each session. Black=breath holding 9 s; Grey=breath holding 15 s and White=breath holding 21 s. (a) first session and (b) second session.

(ranging from 0.30 to 0.70; mean  $p < 0.001$ ) and for BH21 was 0.70 (ranging from 0.60 to 0.83; mean  $p < 0.001$ ). Fig. 9 shows representative scatter plots for one subject and for all three BH tasks.

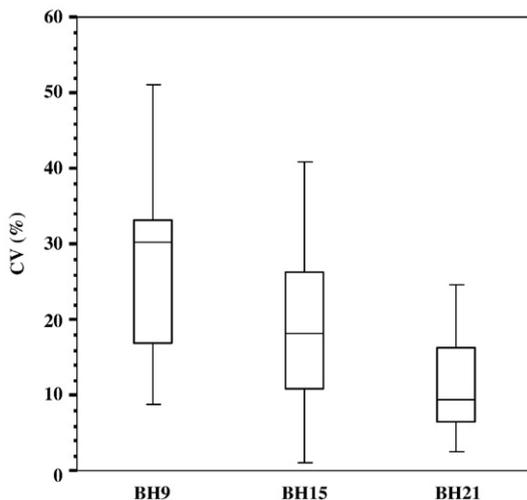
## Discussion

The observed MR response induced by BH showed a triphasic shape. As reported in previous studies, the initial biphasic shape of the curve depends on the combination of many factors happening concomitantly to the respiration act, including the reduction of pulmonary vascular tension and variation in heart rate (Nakada et al., 2001; Thomason et al., 2005; Birn et al., 2008) and autonomic modulation of cerebral myogenic autoregulatory responses (Zhang et al., 2004). Moreover, Zhang et al. (2004) observed that, after a Valsalva manoeuvre the initial biphasic shape may be followed by an overshoot in CBF. In order to control for this possible effect, we instructed the subjects to avoid abdominal muscle contraction at the end of the inspiration. The hypothesis that the initial positive/negative variation is linked to the physiological modification induced by the respiratory act is supported by the observation that this initial variation is not observed when the breath hold is performed after expiration (Kastrup et al., 1998).

The rise of CO<sub>2</sub> concentration in the bloodstream after a breath hold or CO<sub>2</sub> inhalation is not instantaneous (Corfield et al., 2001). In

fact, previous reports have shown a time shift between CO<sub>2</sub> inhalation and the BOLD response (Van der Zande et al., 2005). Our data are in agreement with these reports showing a lag between the start of the breath hold and the expected linear CBF increase induced by the CO<sub>2</sub> accumulation in the bloodstream. This effect is evident in the third portion of the curve where the intensity of the signal increases as BH duration increases as it would be expected from previous studies that have shown a direct relation between the level of CO<sub>2</sub> inhalation and the BOLD signal changes (Posse et al., 2001; Rostrup et al., 2000) and are in keeping with previous reports that investigated the signal variation due to breath holding (Abbott et al., 2005; Liu et al., 2002). However, due to the initial signal modulation by the respiratory act, we cannot determine exactly when the CO<sub>2</sub> accumulation triggered the brain auto-regulatory response. In particular, we cannot rule out the possibility that the early negative undershoot masked the start of this hemodynamic response. It has been demonstrated that, even after a single inspiration act, the negative portion of the resulting BOLD curve is slow, peaking around 16 s after the starting of the inspiration act and recovering to baseline without overshooting after almost 40 s (Birn et al., 2008). Hence, a temporal overlap between the rising phase of the BOLD signal induced by the effect of CO<sub>2</sub> and the resolving negative MR signal phase due to the preceding inspiration act must be expected. This interaction will be most evident for short period of BH that will induce small increase of the CO<sub>2</sub>. In particular, in our study only for BH9 spontaneous respiration activity restarted when the CO<sub>2</sub> effect started to be evident and the negative portion of the curve was approaching its maximum peak. This could lead to an additional interaction with MR signal fluctuation due to spontaneous breathing (Windischberger et al., 2002; Wise et al., 2004) and may explain why we found a significant difference in the initial rising phase from the negative peak to the positive peak between the BH9 compared to BH15 and BH21. Despite these uncertainties, it can be assumed that for all durations of breath hold used in our study, the most representative BOLD cerebrovascular response to CO<sub>2</sub> is characterized by the third portion of the curves. Thus we decided to consider only the third positive portion of the curve to compute the cerebrovascular reactivity maps and perform the reproducibility analysis.

Leontiev and Buxton (2007) evaluated the reproducibility of individual subjects cerebrovascular reactivity BOLD maps using CO<sub>2</sub> inhalation. However, to our knowledge this is the first study to explore this issue using breath holding. The parameters most frequently used to characterize the BOLD signal response are the delay between the stimulus and the observed signal, the amplitude of the response and its full width at half-maximum. Given the complexity of the shape of the signal induced by breath holding, we decided to estimate the



**Fig. 6.** Inter-trial coefficient of variation for each BH duration.

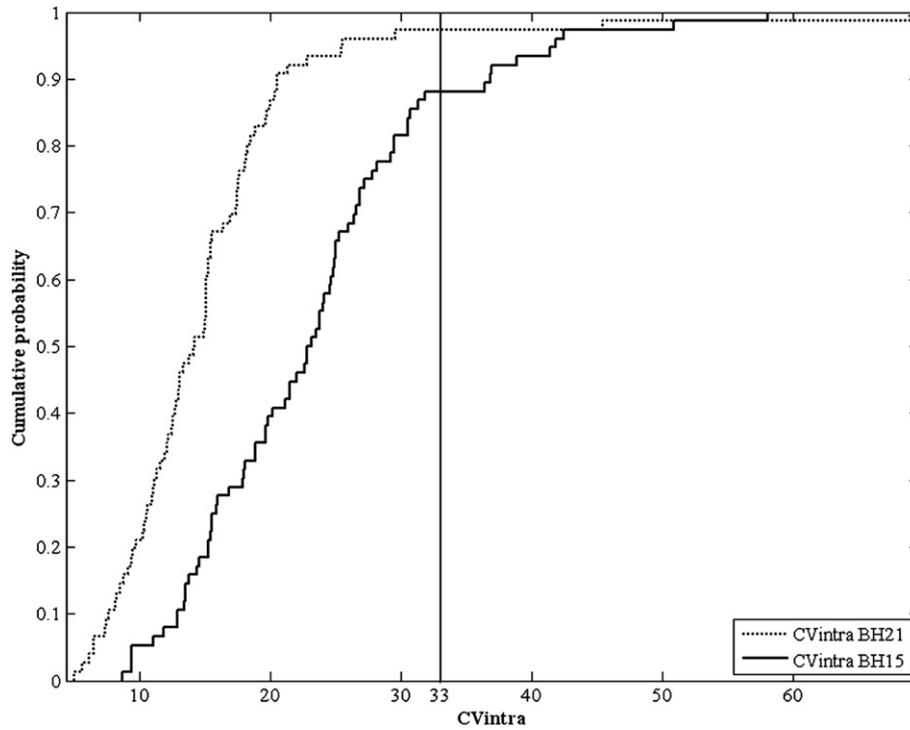


Fig. 7. The graph indicates the cumulative probability function of subregions  $CV_{intra}$  for BH15 and BH21.

magnitude of the response computing also the integral of the area subtended by the curve because this parameter does not require any assumption about the shape of the response itself. The reproducibility of MR signal change induced by BH tasks was then assessed computing for each subject the difference of specific parameters of the BOLD response between sessions ( $\Delta$ ) and evaluating the dispersion of the resulting distribution ( $CV_{intra}$ ). A CV value 33% was taken as the maximum acceptable level of variability so that a parameter was considered more reproducible at one BH duration compared to

another when a significantly lower  $\Delta$  was found and the corresponding  $CV_{intra}$  was lower than 33%.

Interestingly, the timing (Lag and TTP) of the BOLD response displayed a  $CV_{intra}$  well below 33% with no significant differences of the  $\Delta$  across breath hold durations. On the contrary, for what concerns the magnitude of the response (Area, Volume, PSC and Beta), only BH15 and BH21 satisfied the last condition for all estimated parameters while BH9 was associated with a CV value above 33% in two of them, namely Area and Volume. Thus, our data suggest that at

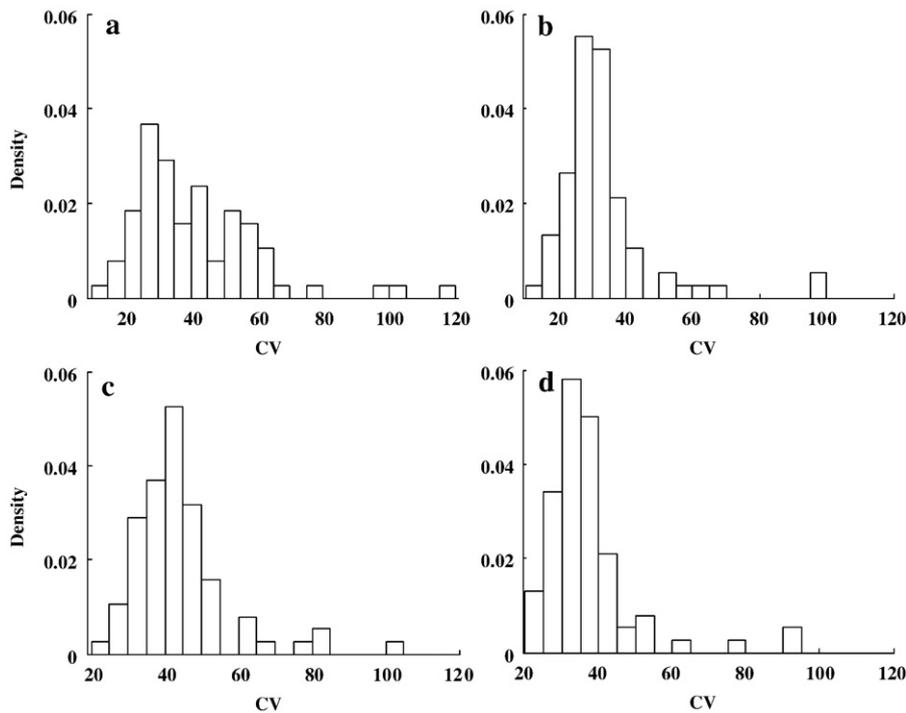


Fig. 8. The figure represents the density distribution of subregions  $CV_{inter}$  (a) BH15 first session; (b) BH21 first session; (c) BH15 second session; (d) BH21 second session.

BH9 the BOLD response is highly variable across sessions. The inter-trial variability of the maximum percentage signal change within each subject at BH9 was significantly higher compared to BH15 and BH21 (Fig. 6). Thomason and Glover (2008) showed that the magnitude of the BOLD signal induced by an inspiration act is strongly influenced by the depth of the inspiration. Although we monitored the chest expansion of the subjects, we did not fully control that the chest expansion (inspiration depth) was equivalent across different BH duration tasks. It is possible that the depths of the inspiration acts for the BH9 task were more variable across subjects resulting in the observed significantly high variability of the BOLD response across trials. Thus, only BH15 and BH21 seem to provide reproducible cerebrovascular reactivity maps. Moreover, between these two breath hold durations, the magnitude of the response appears to be significantly more reproducible at BH21 than BH15.

One of the reasons for which it is useful to acquire cerebrovascular reactivity maps is the possibility to use them to calibrate and/or normalize statistical maps obtained with cognitive tasks. In our study we explored the reproducibility of cerebrovascular reactivity maps under the same conditions to assess if these maps can be reliably used for this aim. The calibration/normalization procedures, however, are usually done within regions or even on a voxel-by-voxel basis. Thus we performed the reproducibility analyses also on a smaller spatial scale taking into account only the parameter mostly used for calibration, that is, the PSC. We focussed this analysis only on the BH15 and BH21 which were associated with a good reproducibility for the cortical grey matter taken as a whole.

The subregions reproducibility analyses confirmed the results obtained from the whole brain cortical grey matter. Notably, the  $CV_{intra}$  value was significantly below 33% even when the analysis was performed region-by-region indicating an acceptable variability. Nonetheless, the analysis of variance of  $\Delta$  factor shows that not all subregions display the same absolute difference in PSC between sessions, but it is important to note that this fact is independent from the breath holding duration.

Thus, overall the data suggest an acceptable reproducibility of the cerebrovascular maps for BH15 and BH21 both when the maps are considered as a whole and when they are analyzed region-by-region.

As expected, in our study the variability of each parameter between subjects pool was higher than the variability between sessions. These results are in agreement with previous studies (Leontiev and Buxton, 2007; McGonigle et al., 2000; Tjandra et al., 2005). This fact has suggested that most of the variability in study groups may be due to physiological differences among subjects in cerebrovascular response and that single subject vasoreactivity

information can be used to reduce the group variability in fMRI studies (Bandettini and Wong, 1997). Thus, if a cerebrovascular reactivity map is acquired for calibration and/or normalization, in order to achieve an optimal result, not only these vasoreactivity BOLD maps must be stable and reproducible, but they have to characterize as much as possible the individual cerebrovascular response catching the variability among the population. For this reason the best protocol should reflect as much as possible the variability of the magnitude of the response across subjects, that is the main parameter used to calibrate the BOLD response. However, the actual variability of the BOLD responses due to physiological differences among subjects is actually unknown. Given this uncertainty it is not possible to set the optimal level of  $CV_{inter}$  that reflects the actual variability of the population. Nevertheless we assume that the lower the  $CV_{inter}$  the higher the probability that the measured response is closer to either a floor or a ceiling effect at the population level. Thus, between the two breath holding paradigms that led to a stable response across trials and a good reproducibility across sessions, we would suggest using the one with the highest  $CV_{inter}$  for PSC. Among the two breath hold paradigms providing reproducible cerebrovascular reactivity maps, BH21 is associated with a lower dispersion of the magnitude of the response across subjects compared to BH15. The subregions analyses confirmed this result. In particular it is possible to observe (Fig. 8) that BH15 presents high dispersion of  $CV_{inter}$  values indicating a different magnitude of the response across subjects and across subregions compared to BH21.

A disadvantage of the BH21 paradigm is that a number of factors could complicate the interpretation of cerebrovascular reactivity maps obtained using long breath hold period. In fact, all subjects reported that holding the breath for 21 s was more difficult than all other BH durations. This self-report was supported by the observation of the most evident chest movements during BH21 compared to the other durations.

In conclusion, our data show that the BOLD response to breath holding after inspiration results in a complex shape due to physiological factors that influence the signal variation with a timing that is highly reproducible. The cerebrovascular response to  $CO_2$  is mostly reflected by the third positive portion of the curve. Nonetheless, the reproducibility of the magnitude of the cerebrovascular response to  $CO_2$ , expressed as amplitude of BOLD signal change and number of responding voxels, is strongly affected by the duration of BH periods. Our results suggest that a 9 s BH paradigm results in higher variability of the magnitude of the BOLD response across trials, across subjects and across sessions. The less variable response across sessions is associated with the paradigm where 21 s breath hold

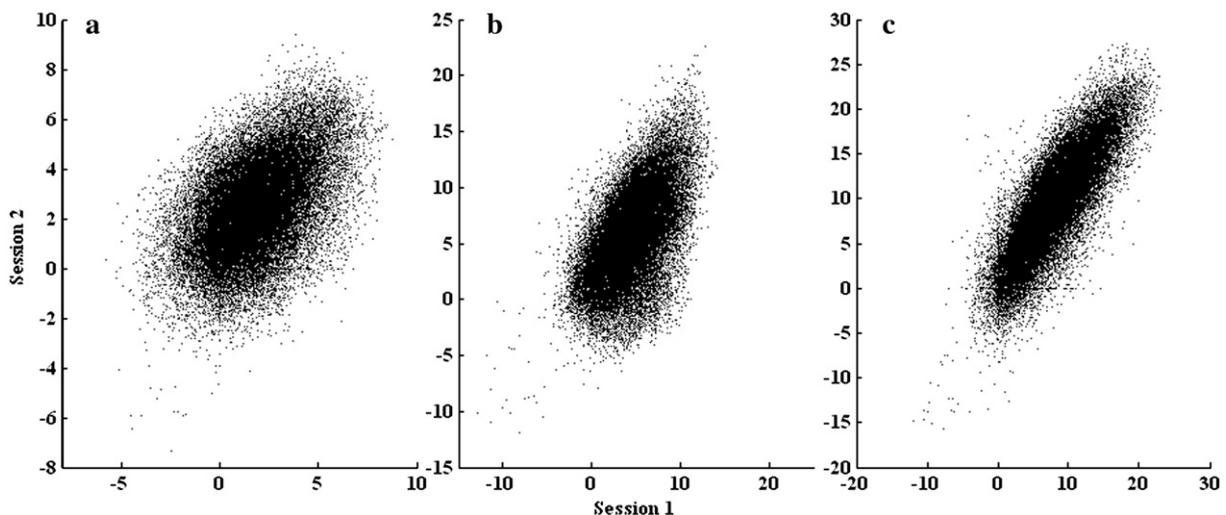


Fig. 9. Representative single subject example of scatter-plot of  $t$ -value for two sessions for three breath holding durations, (a) 9 s (b) 15 s and (c) 21 s.

periods are alternated with normal breathing. Nevertheless, the BH 21 paradigm is also associated with the lowest differences across subjects. Thus, if cerebrovascular maps are generated to calibrate/normalize fMRI BOLD maps, the alternation of breath holding periods of 15 s with normal breathing is associated with acceptable reproducibility across sessions and stability across trials of the BOLD response and it seems to be the best paradigm to catch the variability of the response of the population.

## Acknowledgments

This work was partially supported by Merrill Lynch International. The authors are thankful to Dr. Simon Robinson and to three anonymous reviewers for very constructive comments.

## References

- Abbott, D.F., Opdam, H.I., Briellmann, R.S., Jackson, G.D., 2005. Brief breath holding may confound functional magnetic resonance imaging studies. *Hum. Brain Mapp.* 24, 284–290.
- Allison, P.D., 1978. Measures of inequality. *Am. Sociol. Rev.* 43, 865–880.
- Bandettini, P.A., Wong, E.C., 1997. A hypercapnia-based normalization method for improved spatial localization of human brain activation with fMRI. *NMR Biomed.* 10, 197–203.
- 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.
- Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1, 307–310.
- Bland, J.M., Altman, D.G., 1990. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput. Biol. Med.* 20, 337–340.
- Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J., 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* 16, 4207–4221.
- Brown, G.G., Eyler Zorrilla, L.T., Georgy, B., Kindermann, S.S., Wong, E.C., Buxton, R.B., 2003. BOLD and perfusion response to finger-thumb apposition after acetazolamide administration: differential relationship to global perfusion. *J. Cereb. Blood Flow Metab.* 23, 829–837.
- Chee, M.W., Lee, H.L., Soon, C.S., Westphal, C., Venkatraman, V., 2003. Reproducibility of the word frequency effect: comparison of signal change and voxel counting. *NeuroImage* 18, 468–482.
- Cohen, E.R., Rostrup, E., Sidaros, K., Lund, T.E., Paulson, O.B., Ugurbil, K., Kim, S.G., 2004. Hypercapnic normalization of BOLD fMRI: comparison across field strengths and pulse sequences. *NeuroImage* 23, 613–624.
- Corfield, D.R., Murphy, K., Josephs, O., Adams, L., Turner, R., 2001. Does hypercapnia-induced cerebral vasodilation modulate the hemodynamic response to neural activation? *NeuroImage* 13, 1207–1211.
- Davis, T.L., Kwong, K.K., Weisskoff, R.M., Rosen, B.R., 1998. Calibrated functional MRI: mapping the dynamics of oxidative metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 95, 1834–1839.
- Fernandez, G., Specht, K., Weis, S., Tendolkar, I., Reuber, M., Fell, J., Klaver, P., Ruhlmann, J., Reul, J., Elger, C.E., 2003. Intra-subject reproducibility of presurgical language lateralization and mapping using fMRI. *Neurology* 60, 969–975.
- Girouard, H., Iadecola, C., 2006. Neurovascular coupling in the normal brain and in hypertension, stroke, and Alzheimer disease. *J. Appl. Physiol.* 100, 328–335.
- Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage* 9, 416–429.
- Goebel, R., Esposito, F., Formisano, E., 2006. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum. Brain Mapp.* 27, 392–401.
- Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage* 21, 1639–1651.
- Harrington, G.S., Tomaszewski Farias, S., Buonocore, M.H., Yonelinas, A.P., 2006. The intersubject and intrasubject reproducibility of fMRI activation during three encoding tasks: implications for clinical applications. *Neuroradiology* 48, 495–505.
- Havel, P., Braun, B., Rau, S., Tonn, J.C., Fesl, G., Bruckmann, H., Ilmberger, J., 2006. Reproducibility of activation in four motor paradigms. An fMRI study. *J. Neurol.* 253, 471–476.
- Iadecola, C., 2004. Neurovascular regulation in the normal brain and in Alzheimer's disease. *Nat. Rev. Neurosci.* 5, 347–360.
- Ito, H., Yokoyama, I., Iida, H., Kinoshita, T., Hatazawa, J., Shimosegawa, E., Okudera, T., Kanno, I., 2000. Regional differences in cerebral vascular response to PaCO<sub>2</sub> changes in humans measured by positron emission tomography. *J. Cereb. Blood Flow Metab.* 20, 1264–1270.
- Jansen, A., Menke, R., Sommer, J., Forster, A.F., Bruchmann, S., Hempleman, J., Weber, B., Knecht, S., 2006. The assessment of hemispheric lateralization in functional MRI-robustness and reproducibility. *NeuroImage* 33, 204–217.
- Johnson, N.L., Welch, B.L., 1940. Applications of the non-central *t*-distribution. *Biometrika* 31, 362–389.
- Kastrup, A., Li, T.Q., Takahashi, A., Glover, G.H., Moseley, M.E., 1998. Functional magnetic resonance imaging of regional cerebral blood oxygenation changes during breath holding. *Stroke* 29, 2641–2645.
- Kastrup, A., Kruger, G., Glover, G.H., Neumann-Haefelin, T., Moseley, M.E., 1999a. Regional variability of cerebral blood oxygenation response to hypercapnia. *NeuroImage* 10, 675–681.
- Kastrup, A., Li, T.Q., Glover, G.H., Moseley, M.E., 1999b. Cerebral blood flow-related signal changes during breath-holding. *AJNR Am. J. Neuroradiol.* 20, 1233–1238.
- Kastrup, A., Kruger, G., Neumann-Haefelin, T., Moseley, M.E., 2001. Assessment of cerebrovascular reactivity with functional magnetic resonance imaging: comparison of CO<sub>2</sub> and breath holding. *Magn. Reson. Imaging* 19, 13–20.
- Leontiev, O., Buxton, R.B., 2007. Reproducibility of BOLD, perfusion, and CMRO<sub>2</sub> measurements with calibrated-BOLD fMRI. *NeuroImage* 35, 175–184.
- Li, T.Q., Kastrup, A., Takahashi, A.M., Moseley, M.E., 1999. Functional MRI of human brain during breath holding by BOLD and FAIR techniques. *NeuroImage* 9, 243–249.
- Li, T.Q., Kastrup, A., Moseley, M.E., Glover, G.H., 2000. Changes in baseline cerebral blood flow in humans do not influence regional cerebral blood flow response to photic stimulation. *Magn. Reson. Imaging* 12, 757–762.
- Liu, H.L., Huang, J.C., Wu, C.T., Hsu, Y.Y., 2002. Detectability of blood oxygenation level-dependent signal changes during short breath hold duration. *Magn. Reson. Imaging* 20, 643–648.
- Liu, J.Z., Zhang, L., Brown, R.W., Yue, G.H., 2004. Reproducibility of fMRI at 1.5 T in a strictly controlled motor task. *Magn. Reson. Med.* 52, 751–760.
- Logothetis, N.K., 2002. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 357, 1003–1037.
- 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.
- Marshall, I., Simonotto, E., Deary, I.J., MacLulich, A., Ebmeier, K.P., Rose, E.J., Wardlaw, J.M., Goddard, N., Chappell, F.M., 2004. Repeatability of motor and working-memory tasks in healthy older volunteers: assessment at functional MR imaging. *Radiology* 233, 868–877.
- Markus, H.S., Harrison, M.J., 1992. Estimation of cerebrovascular reactivity using transcranial Doppler, including the use of breath-holding as the vasodilatory stimulus. *Stroke* 23, 668–673.
- McConigle, D.J., Howseman, A.M., Athwal, B.S., Friston, K.J., Frackowiak, R.S., Holmes, A.P., 2000. Variability in fMRI: an examination of intersession differences. *NeuroImage* 11, 708–734.
- McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* 1, 30–46.
- Mulderink, T.A., Gitelman, D.R., Mesulam, M.M., Parrish, T.B., 2002. On the use of caffeine as a contrast booster for BOLD fMRI studies. *NeuroImage* 15, 37–44.
- 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.
- Neumann, J., Lohmann, G., Zysset, S., von Cramon, D.Y., 2003. Within-subject variability of BOLD response dynamics. *NeuroImage* 19, 784–796.
- Ogawa, S., Lee, T.M., Kay, A.R., Tank, D.W., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. U. S. A.* 87, 9868–9872.
- Peelen, M.V., Downing, P.E., 2005. Within-subject reproducibility of category-specific visual activation with functional MRI. *Hum. Brain Mapp.* 25, 402–408.
- Posse, S., Kemna, L.J., Elghahwagi, B., Wiese, S., Kiselev, V.G., 2001. Effect of graded hypo- and hypercapnia on fMRI contrast in visual cortex: quantification of T<sup>\*</sup>(\*) changes by multiecho EPI. *Magn. Reson. Med.* 46, 264–271.
- Raemaekers, M., Vink, M., Zandbelt, B., van Wezel, R.J., Kahn, R.S., Ramsey, N.F., 2007. Test-retest reliability of fMRI activation during prosaccades and antisaccades. *NeuroImage* 36, 532–542.
- Rau, S., Fesl, G., Bruhns, P., Havel, P., Braun, B., Tonn, J.C., Ilmberger, J., 2007. Reproducibility of activations in Broca area with two language tasks: a functional MR imaging study. *AJNR Am. J. Neuroradiol.* 28, 1346–1353.
- Rombouts, S.A., Barkhof, F., Hoogenraad, F.G., Sprenger, M., Scheltens, P., 1998. Within-subject reproducibility of visual activation patterns with functional magnetic resonance imaging using multislice echo planar imaging. *Magn. Reson. Imaging* 16, 105–113.
- Rostrup, E., Law, I., Blinkenberg, M., Larsson, H.B., Born, A.P., Holm, S., Paulson, O.B., 2000. Regional differences in the CBF and BOLD responses to hypercapnia: a combined PET and fMRI study. *NeuroImage* 11, 87–97.
- Specht, K., Willmes, K., Shah, N.J., Jancke, L., 2003. Assessment of reliability in functional imaging studies. *J. Magn. Reson. Imaging* 17, 463–471.
- Talairach, J., Tournoux, P., 1988. *Co-Planar Stereotaxic Atlas of the Human Brain*. G. Thieme, Stuttgart.
- Thomason, M.E., Glover, G.H., 2008. Controlled inspiration depth reduces variance in breath-holding-induced BOLD signal. *NeuroImage* 39, 206–214.
- Thomason, M.E., Burrows, B.E., Gabrieli, J.D., Glover, G.H., 2005. Breath holding reveals differences in fMRI BOLD signal in children and adults. *NeuroImage* 25, 824–837.
- Thomason, M.E., Foland, L.C., Glover, G.H., 2007. Calibration of BOLD fMRI using breath holding reduces group variance during a cognitive task. *Hum. Brain Mapp.* 28, 59–68.
- Tjandra, T., Brooks, J.C., Figueiredo, P., Wise, R., Matthews, P.M., Tracey, I., 2005. Quantitative assessment of the reproducibility of functional activation measured with BOLD and MR perfusion imaging: implications for clinical trial design. *NeuroImage* 27, 393–401.

- Van der Zande, F.H., Hofman, P.A., Backes, W.H., 2005. Mapping hypercapnia-induced cerebrovascular reactivity using BOLD MRI. *Neuroradiology* 47, 114–120.
- Vlioger, E.J., Lavini, C., Majoie, C.B., den Heeten, G.J., 2003. Reproducibility of functional MR imaging results using two different MR systems. *AJNR Am. J. Neuroradiol.* 24, 652–657.
- Ward, N.S., Swayne, O.B., Newton, J.M., 2008. Age-dependent changes in the neural correlates of force modulation: An fMRI study. *Neurobiol Aging* 29, 1424–1426.
- Wei, X., Yoo, S.S., Dickey, C.C., Zou, K.H., Guttmann, C.R., Panych, L.P., 2004. Functional MRI of auditory verbal working memory: long-term reproducibility analysis. *NeuroImage* 21, 1000–1008.
- Windischberger, C., Langenberger, H., Sycha, T., Tschernko, E.M., Fuchsjaeger-Mayerl, G., Schmetterer, L., Moser, E., 2002. On the origin of respiratory artifacts in BOLD-EPI of the human brain. *Magn. Reson. Imaging* 20, 575–582.
- Wise, R.G., Ide, K., Poulin, M.J., Tracey, I., 2004. Resting fluctuations in arterial carbon dioxide induce significant low frequency variations in BOLD signal. *NeuroImage* 21, 1652–1664.
- Wu, T., Hallett, M., 2005. A functional MRI study of automatic movements in patients with Parkinson's disease. *Brain* 128, 2250–2259.
- Yetkin, F.Z., McAuliffe, T.L., Cox, R., Haughton, V.M., 1996. Test–retest precision of functional MR in sensory and motor task activation. *AJNR Am. J. Neuroradiol.* 17, 95–98.
- Zandbelt, B.B., Gladwin, T.E., Raemaekers, M., van Buuren, M., Neggers, S.F., Kahn, R.S., Ramsey, N.F., Vink, M., 2008. Within-subject variation in BOLD-fMRI signal changes across repeated measurements: quantification and implications for sample size. *NeuroImage* 42, 196–206.
- Zhang, R., Crandall, C.G., Levine, B.D., 2004. Cerebral hemodynamics during the Valsalva maneuver: insights from ganglionic blockade. *Stroke* 35, 843–847.