



# Impact of task-related changes in heart rate on estimation of hemodynamic response and model fit

Sarah F. Hillenbrand<sup>a,\*</sup>, Richard B. Ivry<sup>a,b,c</sup>, John E. Schlerf<sup>a,c</sup>

<sup>a</sup> Helen Wills Neuroscience Institute, University of California, Berkeley, USA

<sup>b</sup> Department of Psychology, University of California, Berkeley, USA

<sup>c</sup> 5201 Tolman Hall, University of California, Berkeley, Berkeley, CA 94720, USA

## ARTICLE INFO

### Article history:

Received 9 June 2015

Accepted 22 February 2016

Available online 2 March 2016

### Keywords:

fMRI

Physiological noise

Respiration

Movement

Hemodynamic response function

## ABSTRACT

The blood oxygen level dependent (BOLD) signal, as measured using functional magnetic resonance imaging (fMRI), is widely used as a proxy for changes in neural activity in the brain. Physiological variables such as heart rate (HR) and respiratory variation (RV) affect the BOLD signal in a way that may interfere with the estimation and detection of true task-related neural activity. This interference is of particular concern when these variables themselves show task-related modulations. We first establish that a simple movement task reliably induces a change in HR but not RV. In group data, the effect of HR on the BOLD response was larger and more widespread throughout the brain than were the effects of RV or phase regressors. The inclusion of HR regressors, but not RV or phase regressors, had a small but reliable effect on the estimated hemodynamic response function (HRF) in M1 and the cerebellum. We next asked whether the inclusion of a nested set of physiological regressors combining phase, RV, and HR significantly improved the model fit in individual participants' data sets. There was a significant improvement from HR correction in M1 for the greatest number of participants, followed by RV and phase correction. These improvements were more modest in the cerebellum. These results indicate that accounting for task-related modulation of physiological variables can improve the detection and estimation of true neural effects of interest.

© 2016 Elsevier Inc. All rights reserved.

## Introduction

Functional magnetic resonance imaging (fMRI) is widely used to examine responses of the human brain to a variety of tasks and stimuli. One disadvantage of the method is that it only provides an indirect measurement of neural activity by measuring changes in the blood oxygenation level dependent (or BOLD) signal (Ogawa and Lee, 1990). These changes occur on a much slower time scale than changes in the activity of local neural populations (Logothetis, 2003). A complex relationship exists between neural activity and the changes in blood flow, volume, and oxygenation that form the basis of the BOLD signal (Buxton and Frank, 1997; for review, see Logothetis & Wandell, 2004). Fortunately, the BOLD response is essentially linear and time-invariant (Boynton et al., 1996; Dale and Buckner, 1997; Friston et al., 1994); therefore, the brain response to many events can be efficiently extracted with events separated by only a few seconds (Dale, 1999).

In order to determine whether a particular brain area is active in such a task, researchers typically adopt a regression approach, using a

general linear model (GLM) and regression analysis to identify brain regions in which the BOLD response matches a set of predictions (Friston et al., 1995; see Poline and Brett, 2012 for a broad evaluation of the costs and benefits of this approach). Since the coupling between the neural and BOLD responses has a similar shape across a wide variety of conditions, a canonical hemodynamic response function (HRF) is frequently employed in fMRI analyses (Friston et al., 1998). While this approach can greatly simplify fMRI analysis, it does come at a cost, given that the HRF has been shown to differ across individuals, brain regions, and events (Handwerker et al., 2004). Generating a predicted brain response using the canonical HRF can therefore result in a poorer fit in comparison to individualized, region-specific, or task-specific HRFs, potentially leading to a mischaracterization of brain activity (Hernandez et al., 2002; Handwerker et al., 2004).

The fit of any GLM can be diminished by failing to account for factors that are correlated with each other, a problem that is especially pronounced in event-related studies of BOLD signal that are more susceptible to noise. Two important, measurable, and often ignored physiological covariates are heartbeats and respiration (Glover et al., 2000). The beating of the heart causes pulsations in blood vessels and cerebrospinal fluid (CSF), creating artifacts near large blood vessels, around ventricles, and even in deep sulci (Dagli et al., 1999). Additional artifacts are introduced by respiration, as the rise and fall of the chest

\* Corresponding author at: 5201 Tolman Hall, University of California, Berkeley, Berkeley, CA 94720, USA.

E-mail addresses: [sarahhildenbrand@gmail.com](mailto:sarahhildenbrand@gmail.com) (S.F. Hillenbrand), [ivry@berkeley.edu](mailto:ivry@berkeley.edu) (R.B. Ivry), [schlerf@gmail.com](mailto:schlerf@gmail.com) (J.E. Schlerf).

cavity during breathing causes both head motion and magnetic field disturbances (Glover et al., 2000; Raj et al., 2001). It is therefore desirable to measure heartbeats and breathing to account for their influence on the BOLD signal. The RETROICOR method developed by Glover et al. (2000) provides one such approach, charting the phase of cardiac and respiration processes relative to image acquisition. Variance attributable to the phase of these processes may be removed in preprocessing or accounted for by including nuisance regressors in a GLM.

One limitation of the RETROICOR correction is that it does not take into account how changes in the rate of physiological processes affect the BOLD signal. Changes in heart rate (HR), as well as respiration variance (RV, a measure which accounts for changes in amplitude and rate), can cause fluctuations in the BOLD signal (Birn et al., 2006, 2008; Shmueli et al., 2007; Chang et al., 2009; Chang and Glover, 2009). Taking these factors into account during resting state scans has been shown to alter the spatial spread of connectivity maps (van Buuren et al., 2009; and for review, see Birn, 2012). While resting state studies often measure correlations associated with the “default mode network” (Raichle et al., 2001; Greicius et al., 2003), one study found that these correlations were in fact robust to correction for physiological noise, but activations in a task-positive network were reduced (van Buuren et al., 2009).

Importantly, changes in HR and RV are frequently task-related, associated with variations in arousal (Tursky et al., 1969), movement preparation (Damen and Brunia, 1987), response inhibition (for review see Jennings and van der Molen, 2002), feedback processing (Crone et al., 2003, 2005), cognitive interference and planning (van 't Ent et al., 2014), and pharmacological state (Khalili-Mahani et al., 2013). Indeed, changes in HR, RV, and other autonomic indicants, such as skin conductance and pupil dilation, are common dependent variables in the study of a range of cognitive processes and their associated brain responses (for review, see Critchley, 2009). Furthermore, autonomic variables are themselves regulated by efferent signals from the brain, making the direction of influence between brain and body difficult to discern (see Iacovella and Hasson, 2011).

We previously demonstrated the importance of considering task-related changes in physiological processes in a study designed to identify brain regions responsive to movement errors (Schlerf et al., 2012). When physiological regressors were not included in the GLM analysis, reaching errors led to a broadly distributed decrease in the BOLD response in the cerebellum. However, there was also a reliable reduction in HR following movement errors. When the model included HR, the cerebellar deactivations were no longer evident. Instead, an increase in the BOLD signal was observed on error trials, restricted to the arm area of the anterior cerebellum. Thus, the expected error signal in the cerebellum was only evident after task-dependent changes in HR were included in the model of the BOLD response. Nevertheless, to our knowledge, no other study has investigated the impact of physiological noise correction on task responses in the cerebellum.

In the current study, we systematically investigated the potential consequences of task-related fluctuations in HR and RV on the HRF, as well on the model fit of the BOLD response, using a progressive series of analyses. Rather than focusing on movement errors, we examined responses to a simpler behavior: arm movements produced in the absence of visual feedback. We chose to measure responses to this simple type of event for two reasons: First, it allowed us to situate the error-specific changes observed by Schlerf et al. (2012) in the more general context of movement-related changes. Second, the use of a simple motor behavior decreases the likelihood that neural activity related to cognitive processing (e.g., error processing) is driving the physiological changes. As such, this would increase our confidence that the impact of HR and RV on the BOLD signal do not reflect the efferent regulation of autonomic processes (e.g., Kobayashi et al., 2007; Iacovella and Hasson, 2011).

We first demonstrated that HR is consistently affected by arm movement. In contrast, changes in RV are more variable. We then examined

the effect of these variables on the BOLD signal throughout the brain. We next asked how the inclusion of physiological regressors in the GLM influenced the shape of the estimated arm movement-related HRF in two motor regions: primary motor cortex (M1) and the cerebellum. Finally, we quantified the added explanatory power of different sets of physiological regressors, either in isolation or in combination.

## Material and methods

### Participants

Eleven healthy, right-handed participants were tested (7 female, mean age 24.1 years). The participants provided written, informed consent under a protocol approved by the University of California, Berkeley Institutional Review Board.

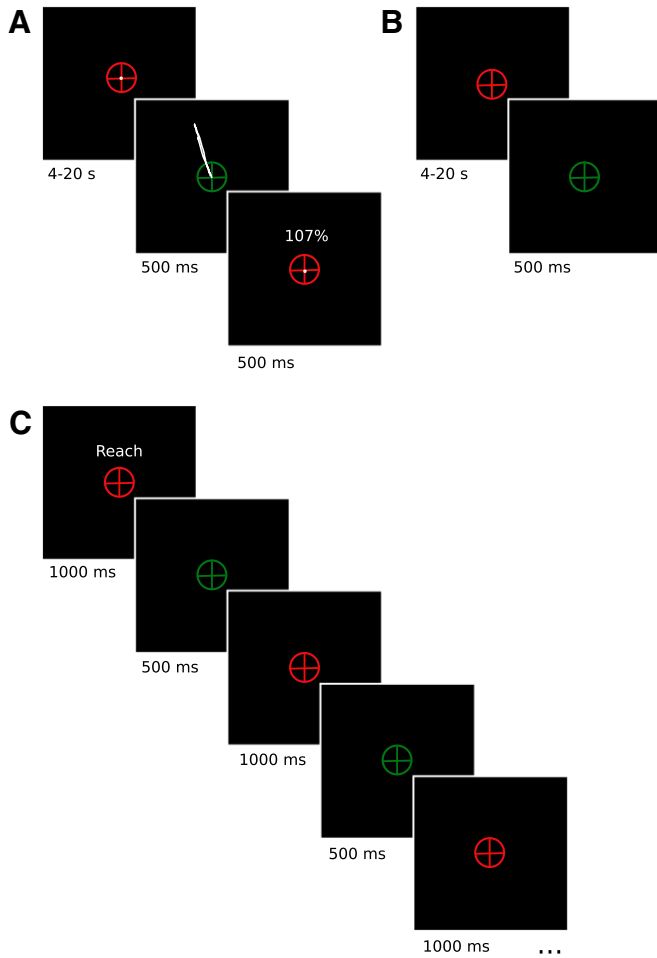
### Task

Prior to scanning, participants were fitted with a custom bite bar. During the scanning session, the bite bar was mounted to the head coil to minimize head movement. Stimuli were backprojected onto a screen mounted inside the bore of the magnet and viewed via a mirror mounted to the head coil. From a supine position, the participants held a robotic manipulandum (<http://www.fmrirobot.org>) in their right hand. The manipulandum was positioned over the participant's abdomen and could be freely moved in a plane parallel to the scanner bed.

Participants were trained to make short (8 cm) out-and-back reaching movements along the axis of the body toward their head, chiefly by flexion about the elbow. They were instructed to terminate each return movement such that in between trials, the hand rested comfortably near the navel. Participants were instructed to move when a central fixation crosshair changed color from red to green. For all runs, the green crosshair was presented for 500 ms, regardless of inter-trial interval. Participants were told to initiate the movement as soon as they saw the color change. To minimize corrective movements and processing load, there was no visual feedback of hand position during scanning. At the termination of each return movement (when the hand coordinates were no more than 1 cm apart for a minimum of 500 ms), the start position of the hand for the next trial was automatically adjusted to the center of fixation.

All participants completed a training session in a mock scanner 1–7 days prior to the scanning session. This session served to familiarize the participants with the bite bar, manipulandum, and scanning environment, and to train them in the movement task. The training session consisted of four runs and was designed to train participants to make movements in the scanner of approximately uniform amplitude without relying on feedback. The training runs provided feedback that became progressively less informative as the training continued. In the first run, the participants received online feedback of the cursor position and feedback about reach amplitude at the end of each movement. Reach amplitude feedback was given in numeric form, shown above the fixation crosshair as a percentage of the desired 8 cm amplitude for 500 ms immediately following completion of the return movement (Fig. 1A). For the next run, they were only given reach amplitude feedback (no online cursor feedback), and for the final two runs, no feedback was provided, as in the actual scan session (Fig. 1B). At various points in the training session, the experimenter provided verbal coaching concerning movement initiation, speed, and amplitude.

The scanning session consisted of an anatomical scan and three functional scans: one localizer run and two task runs. The localizer run lasted 6 min and 40 s and consisted of 12 12-second blocks, with rest periods lasting 21.3 s in between each block (Fig. 1C). There were two types of blocks: reach and auditory (6 of each block type). Reach blocks were indicated by the appearance of the word “Reach” on the screen. Participants then produced eight out-and-back movements, initiating each movement when they saw the fixation crosshair turn green. Over each



**Fig. 1.** Task. Participants were instructed to move when a red crosshair turned green. Participants were first trained with a visible cursor and feedback about reach amplitude (A). This feedback was eliminated during the late training runs. During task runs, participants did not see a cursor or feedback, and movement onset cues appeared on a pseudorandom schedule with the inter-movement interval lasting between 4 and 20 s (B). During reach blocks of the localizer run, participants were cued at the onset of the block with the appearance of the word 'Reach'. In each block, eight green movement onset cues appeared, once every 1500 ms (C, only the first two of eight movement cues are shown).

12-second block, the crosshair turned green every 1500 ms. Auditory blocks were indicated by the appearance of the word “Listen” on the screen, and participants heard eight popping sounds, one every 1500 ms. These trials were included for a separate task and will not be discussed further. The reach and auditory blocks alternated in a pseudorandom manner, with the order counterbalanced across participants.

The two task runs (as well as a practice task run completed during the anatomical scan) consisted of 30 randomly timed reaches, with inter-movement intervals ranging between 4 and 22 s, for a run duration of 5 min (Fig. 1B). To ensure our ability to obtain parameter estimates at twelve lags (24 s) following each movement, an optimization procedure was used to create the order of the inter-movement intervals: 1000 random sequences of intervals were generated, and the six most efficient were selected, based on the contrast Move vs. Rest. From these six, sequences were selected at random for each participant's training, practice, and task runs.

Efficiency was calculated following the guidelines of Dale, 1999, according to the equation:

$$\text{Efficiency} = 1/\text{trace}\left(\left(X^T X\right)^{-1}\right) \quad (1)$$

where the design matrix  $X = [X_1 X_2 \dots X_{12}]$  is a horizontal concatenation of delta (stick) functions that represent the stimulus timing at each lag. While efficiency is typically computed for a particular contrast, in the case of HRF extraction that contrast is simply the identity matrix.

#### Imaging parameters

Data were collected on a 3 T MAGNETOM Trio scanner (Siemens Healthcare, Erlangen, Germany) at the Henry H. Wheeler, Jr. Brain Imaging Center at the University of California, Berkeley. A 12-channel receive-only radiofrequency head coil was used. One high-resolution T1-weighted MPRAGE anatomical scan (TR, 1900 ms; TE, 2.52 ms;  $1 \times 1 \times 1$  mm voxels; acquisition matrix  $176 \times 256 \times 256$ ; field of view  $17.6 \times 25 \times 25$  cm) was acquired for each participant. Echo-planar imaging (EPI), specifically, a custom variant of ep2d\_bold on VB15 software, was used to collect functional imaging data (gradient-echo EPI sequence; TR, 2000 ms; TE, 26 ms; 36 sagittal slices acquired linearly from left to right;  $3.3 \times 3.13 \times 3.13$  mm voxels; flip angle,  $90^\circ$ ; acquisition matrix  $64 \times 64$ ; field of view  $20 \times 20$  cm; 150 volumes per task run, 200 volumes per localizer run) using parallel imaging reconstruction (GRAPPA) with an in-plane acceleration factor of 2.

For the majority of participants, these parameters did not result in whole-brain coverage; small portions of the lateral aspects of the brain were excluded. Slices were positioned with a bias toward capturing the left precentral gyrus, the more lateral of our two regions of interest. All regions of interest were well medial to the slice positioning boundaries.

#### Ruling out MRI artifacts

To ensure high quality functional scans, pilot scans of a phantom were run using task runs from the present study. Previous work has suggested that movement of a similar robotic manipulandum was not sufficient to cause any fMRI artifacts (Diedrichsen et al., 2005); we sought to confirm this result using our manipulandum and scanner. We also sought to assess the impact of movement in the scanner bore on data quality when parallel imaging reconstruction techniques (GRAPPA) are used.

These scans were collected with various configurations: 1) with and without an experimenter moving her arm in the scanner bore at each movement cue onset, 2) with and without the presence of the robotic manipulandum, and 3) with and without GRAPPA (total of eight pilot scans). We then generated statistical maps of the temporal signal-to-noise ratio (tSNR). Inspection of the tSNR within the scans of the phantoms revealed that the use of GRAPPA halved tSNR, an expected cost of acceleration. In contrast, movement in the scanner bore reduced tSNR by only about 10% and the presence of the robotic manipulandum had no effect (See Supplementary Material).

Of particular concern was the avoidance of a “striping” artifact that can result when scanner movement interacts with interleaved slice acquisitions (See Supplementary Material). Visual inspection of the phantom scans and the data from the present study confirmed that the use of a linear slice acquisition scheme successfully prevented the appearance of this artifact.

#### fMRI data preprocessing

Imaging data were preprocessed and analyzed using SPM5 (<http://www.fil.ion.ucl.ac.uk/spm/>) and custom routines written in Matlab (<http://www.mathworks.com/products/matlab/>). The first five volumes of each run were discarded to allow for T1 equilibration effects. Images were slice-time corrected using sinc interpolation, then realigned to the mean image of each run to correct for head movement. All images were then coregistered to an anatomical scan and smoothed with a Gaussian kernel (FWHM = 8 mm).

## Physiological monitoring

During scanning, physiological signals were recorded using a BIOPAC physiological monitoring system (<http://www.biopac.com>). Heartbeat was measured using a photoplethysmograph placed on the participant's left index finger. Participants were instructed to rest their left hand comfortably at their side to avoid movement-related artifacts in the signal. Respiration was measured with a pneumatic pressure sensor placed several centimeters below the sternum and held in place by an elastic strap. Once a clear signal was established, participants were instructed to breathe normally. Cords from both sensors were checked once the participant was placed in the scanner to ensure that the arm movements did not alter the quality of the physiological signals. Analog TTL signals generated by the scanner were recorded to time lock these physiological measurements with the EPI images. All data were recorded at 125 Hz, with separate recordings initiated for each EPI sequence. Recordings began at least 10 s before the start of scanning and lasted at least 20 s past the termination of the imaging sequence.

## Assessment of task-related changes in physiological variables and creation of physiological regressors

The effect of reaching movements on our two physiological variables of interest, heart rate (HR) and respiratory variation (RV), was assessed at the group level and individually. Using HR and RV regressors, we evaluated task-related changes by time-locking the estimate of these variables to each movement cue onset. This procedure was performed separately for each run, with the data normalized by subtracting out the mean value for that run.

Physiological regressors were created using the Physiological Log Extraction for Modeling (PhLEM) v1.0 toolbox for SPM5 (Verstynen and Deshpande, 2011). The PhLEM package marks respiration and heartbeat events using an automatic peak detection algorithm. On a few occasions, the spacing between the detected peaks indicated that the algorithm failed to detect a peak. In these cases (0.03% of all events), the data were visually inspected and peaks were manually added at the appropriate deflection in the waveform.

Regressors to predict the effects of respiratory and cardiac phase were made using the RETROICOR method (Glover et al., 2000) as implemented in PhLEM (Verstynen and Deshpande, 2011). Using a Fourier expansion, the first two harmonics of the heartbeat and respiration events were computed. Both sine and cosine waveforms were included, yielding a total of eight phase regressors.

The HR time series was computed following the methods outlined in Chang et al. (2009). For each 6 s window centered on a given 2 s TR in the fMRI time series, the mean inter-beat interval was computed. This interval was then inverted and multiplied by 60 to convert the data into a measure of beats per minute. This time series was then shifted by 0–11 TRs to produce 12 heart rate regressors at lags of 0–22 s. These will be referred to as HR regressors.

The respiration time series was computed using a measure of respiratory variation (RV) (Chang et al., 2009), a simpler and more robust alternative to respiration volume per unit time (Birn et al., 2008). Following previous work, we opted to examine task-related changes in RV rather than peak-to-peak respiration rate (Chang et al., 2009). The RV regressor captures changes in both depth and frequency of breaths, yielding a measurement that is associated with tidal volume. Changes in tidal volume are hypothesized to trigger a feedback loop involving vasodilation, blood flow changes, and ultimately, compensatory changes in the rate and depth of the breaths themselves (Birn et al., 2006; Chang and Glover, 2009). Similar to the procedure for HR, RV was obtained by computing the standard deviation of the respiration waveform within a 6 s sliding window centered on each 2 s TR. This time series was then shifted by 0–11 TRs to produce 12 respiration volume, RV regressors.

## Effects of physiological variables on BOLD response and movement-related activations

Separate GLMs were run for each of the two task runs for each participant. The basic, uncorrected GLM modeled movements as delta (stick) functions at movement cue onset and at 11 lags of 2 s each, using a Finite Impulse Response (FIR) expansion model. Additional GLMs included different sets of physiological regressors. The basic physiological regressors were 12 lags of heart rate (HR), 12 lags of respiratory variation (RV), or 8 phase regressors. The set of phase regressors always included both cardiac and respiratory phases, implemented using the RETROICOR method (Glover et al., 2000).

To characterize the effects of RV, HR, and phase on whole-brain activity, group-level statistical maps of the effects of each of these sets of regressors were calculated. Three separate GLM analyses from the two task runs were performed, each sharing movement onset regressors, while differing in the set of nuisance regressors (one GLM each with RV, HR, or Phase). To assess the spatial distribution of variance explained by each of these sets of nuisance regressors, beyond that explained by the shared movement regressors, whole-brain F-statistic maps (including all 12 lags of the physiological nuisance regressors) were created for each set of nuisance regressors for each individual. These are shown in Fig. 4A–C.

Group-level statistical maps of the Move vs. Rest contrast were calculated both with and without each physiological correction, for a total of six maps. To assess the spatial distribution of variance explained by the movement regressors, whole-brain F-statistic maps for the contrast Move vs. Rest (including all lags in the movement onset regressor) were created from each GLM. Only the effect of movement as estimated using the uncorrected GLM is shown in Fig. 4D.

These maps were first calculated at the individual level and then warped to the MNI152 atlas template (Montreal Neurological Institute, Montreal, QC, Canada) and resampled to the atlas's  $1 \times 1 \times 1$  mm resolution for display purposes only. These and all subsequent visualizations of fMRI data were created using AFNI (<http://afni.nimh.nih.gov>). Note, however, that all analyses were conducted in participants' native space.

## ROI localization

We focused on two a priori anatomical regions of interest (ROIs) to examine activations during arm movements. The precentral gyrus ROI was selected to include contralateral primary motor cortex, and the right anterior lobe of the cerebellum (lobules I–V) ROI was selected to include the ipsilateral cerebellar representation of the hand (Table 1).

A GLM analysis was performed on the localizer data from each participant. Movement blocks and the irrelevant sound blocks were modeled by convolving their onset timing with the canonical

**Table 1**

Regions of interest.

T-statistic values at the local maximum at center of ROI in M1 and cerebellum (T-peak), along with center coordinate of each ROI. ROIs were formed by including the 83 nearest contiguous voxels above the threshold ( $p < .05$ , family-wise error corrected). Voxels not included in clusters surrounding local maxima using this threshold were excluded from ROIs.

| Participant | M1 T-peak | M1 coordinate | Ce T-peak | Ce coordinate |
|-------------|-----------|---------------|-----------|---------------|
| 1           | 10.49     | −30 −28 68    | 11.68     | 7 −56 −13     |
| 2           | 14.54     | −21 −45 58    | 10.54     | 16 −43 −27    |
| 3           | 13.7      | −23 −32 68    | 8.39      | 25 −61 −19    |
| 4           | 10.38     | −22 −30 69    | 9.16      | 28 −42 −28    |
| 5           | 17.13     | −18 −34 64    | 15.45     | 16 −32 −26    |
| 6           | 9.76      | −12 −19 68    | 11.94     | 19 −25 −28    |
| 7           | 12.85     | −27 −29 69    | 16.09     | 2 −42 −10     |
| 8           | 12.76     | −28 −9 57     | 20.76     | 27 −20 −26    |
| 9           | 14.42     | −45 −20 58    | 16.15     | 4 −30 −11     |
| 10          | 8.69      | −12 −19 68    | 9.86      | 28 −13 −19    |
| 11          | 9.28      | −28 −28 59    | 9.78      | 22 −27 −28    |



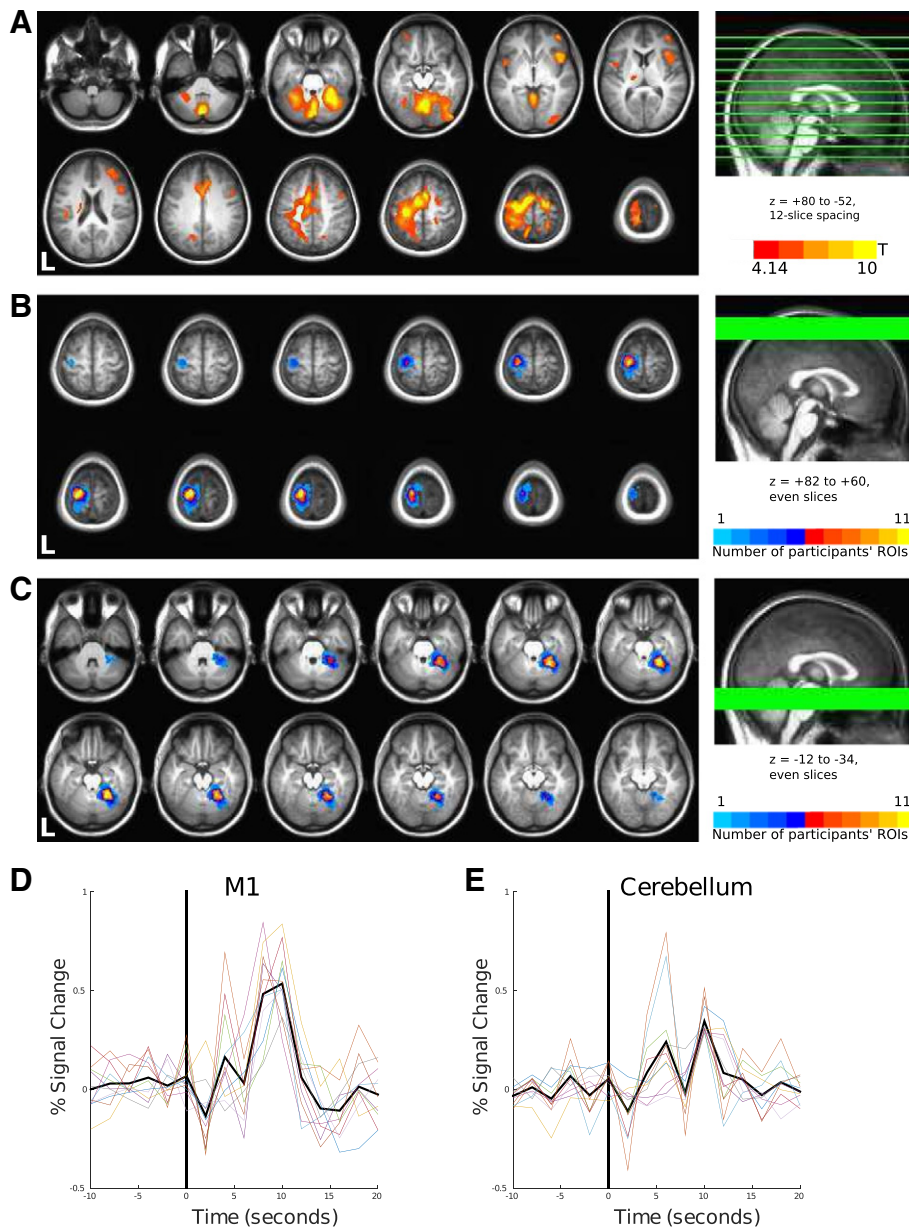
hemodynamic response function (HRF) in SPM. No physiological regressors were included in this analysis. Selecting ROIs based on their fit to a canonical HRF may introduce a bias in the estimates of the deconvolved hemodynamic responses, as this will preferentially identify voxels in which the response already resembles the canonical HRF. However, the block design of the localizer task reduces this bias, as the model fit in such a design is only weakly affected by the choice of response function (Handwerker et al., 2004, 2012).

Whole-brain  $t$ -statistic maps for the contrast Move vs. Rest were created. Local maxima in our a priori regions of interest were selected as peaks around which to center the individually specified ROIs. From these peaks, the ROIs were expanded outward to include all of the contiguous voxels within the  $t$ -statistic map thresholded at  $p < .05$ , family-wise error corrected.

The number of voxels meeting this statistical criterion within each ROI was computed for each individual. Across participants, the smallest

number of voxels meeting this criterion was 83. To keep the size of the ROIs constant, the cerebellar and M1 ROIs were each limited to the 83 closest contiguous voxels. While M1 ROIs were centered on local maxima in the left (contralateral) precentral gyrus, the ROI could extend beyond the precentral gyrus. No local maximum was identified within the precentral gyrus for one participant. However, we were able to identify a peak in the postcentral gyrus, based on visual inspection of the anatomical scan to which functional images were coregistered, with the surrounding set of 83 contiguous voxels extending into the precentral gyrus. For two additional participants, a small number ( $<10$ ) of voxels in the ROI extended into the postcentral gyrus.

Cerebellar ROIs were centered on local maxima in the right ipsilateral anterior lobe (lobules I–V). To make the cerebellar ROIs, the  $t$ -statistic map was first masked by the cerebellar segmentation produced by the Spatially Unbiased Infra-tentorial (SUIT) toolbox in SPM5 (Diedrichsen et al., 2009). These masks were hand-edited by overlaying them on



**Fig. 2.** Regions of interest and movement-evoked BOLD responses. Group-level  $t$ -statistic maps for the contrast Move vs. Rest in the localizer run, thresholded at  $p < .001$ , uncorrected (A). Overlap map of individual 83-voxel ROIs in M1 (B) and cerebellum (C), spatially normalized for visualization only and overlaid on group-averaged anatomical slices in MNI space. Color bar at right indicates proportion of overlap of the ROIs of individual participants. Midsagittal view of slices is shown on the right for each panel. Mean evoked change in the BOLD signal, as measured during task runs, is plotted for individuals (thin lines, different color for each participant) as well as the group (thick lines), for M1 (D) and for the cerebellum (E).

**Table 2**

Comparison of deconvolved HRFs from the Uncorrected model with each of seven Corrected models.

The interaction term quantifies time-dependent differences in the HRF between models. Significant interaction effects were found for all models that included HR (bold).

| Model                  | M1            |               | Cerebellum    |               |
|------------------------|---------------|---------------|---------------|---------------|
|                        | Interaction F | Interaction P | Interaction F | Interaction P |
| <b>RV + HR + phase</b> | 4.64          | <b>0.0001</b> | 3.98          | <b>0.0001</b> |
| <b>HR + phase</b>      | 5.72          | <b>0.0001</b> | 6.43          | <b>0.0001</b> |
| RV + phase             | 0.93          | 0.51          | 0.73          | 0.71          |
| <b>HR + RV</b>         | 3.98          | <b>0.0001</b> | 4.09          | <b>0.0001</b> |
| <b>HR only</b>         | 5.02          | <b>0.0001</b> | 6.96          | <b>0.0001</b> |
| RV only                | 0.71          | 0.73          | 0.41          | 0.95          |
| Phase only             | 1.86          | 0.053         | 1.15          | 0.33          |

the anatomical scan for each participant and removing voxels outside the cerebellum (e.g., from parts of occipital cortex that lie just dorsal to the anterior lobe of the cerebellum).

For visualization purposes only, a group-level t-statistic map for the contrast Move vs. Rest was created, and the ROIs were warped to the MNI (M1) or SUIT (cerebellum) templates and overlaid on a group-averaged anatomical scan in MNI space to create an overlap map (Fig. 2). For comparison of task-related changes in the BOLD time series to task-related changes in HR and RV, the raw BOLD signal in these regions was first normalized by converting to units of percent signal change (subtracting, from each timepoint at each voxel, the mean value of each voxel within each run, and then dividing by the mean value). The mean time series was then calculated for the M1 and cerebellar ROIs for each individual, and these data were time-locked to each movement cue onset to visualize task-related changes in the BOLD signal.

#### HRF estimation

As described in [Assessment of task-related changes in physiological variables and creation of physiological regressors](#) Section 2.7, a separate GLM was run for each of the two task runs for each participant. The basic, uncorrected GLM modeled movement cue onsets and 11 lags of 2 s each, using a Finite Impulse Response (FIR) expansion model. Within the M1 and cerebellar ROIs, parameter estimates were obtained and averaged across voxels. The mean parameter estimates for the contrast Move vs. Rest within each ROI, when plotted over the 11 lags, provide a good estimate of the hemodynamic response to movement within the ROI.

Additional GLMs included combinations of sets of physiological regressors: 12 lags of HR, 12 lags of RV, or 8 phase regressors. The set of phase regressors always included both cardiac and respiratory phases, as commonly implemented using the RETROICOR method (Glover et al., 2000). This approach is better suited for comparing rate-sensitive corrections (RV, HR) to rate-insensitive ones (Phase), and precludes comparing the phase-based influences of respiratory and cardiac processes to each other directly. We estimated HRFs from models including HR alone, RV alone, phase alone, all pairwise combination of these three regressors (HR + RV, HR + phase, and RV + phase), and a model that included all three.

In summary, the HRF was estimated using eight separate GLM analyses: the Uncorrected model (not including physiological covariates) and the seven Corrected models (Table 2). The parameter estimates comprising the deconvolved HRF for each of the Corrected models

were compared to those of the Uncorrected model using a 2-way repeated-measures ANOVA, with factors Time (12 TRs beginning with movement cue onset) and Model (2 models: Uncorrected vs. Corrected).

#### Percent variance explained and model comparisons

The percent variance explained by each of the eight models was calculated and compared to that of the other models using three tiers of nested model pairs. The first level involved a comparison of each of the models containing a single set of regressors (HR at 12 lags, RV at 12 lags, or the 8 phase regressors) to the Uncorrected model. The second level compared models with two sets of regressors to the models containing the individual sets of regressors (e.g., HR plus phase was compared to HR only and, in a separate comparison, to phase only). The third level compared the model that included all three regressors to each of the three paired-regressor models.

Analyzing nested models allowed us to determine what effect each additional regressor or set of regressors had on the overall percent variance explained, as well as to determine whether the variance explained by a given set of regressors could be redundantly explained by another set of regressors. This is particularly important, as it was previously unknown whether RV and HR have redundant effects on the BOLD response, especially in a task-dependent context. It is possible that both variables will provide a comparable improvement in fit over the uncorrected model. However, Schlerf et al. (2012) found that only HR provided a useful correction, perhaps related to the observation that movement errors caused a significant change in HR, but not in RV.

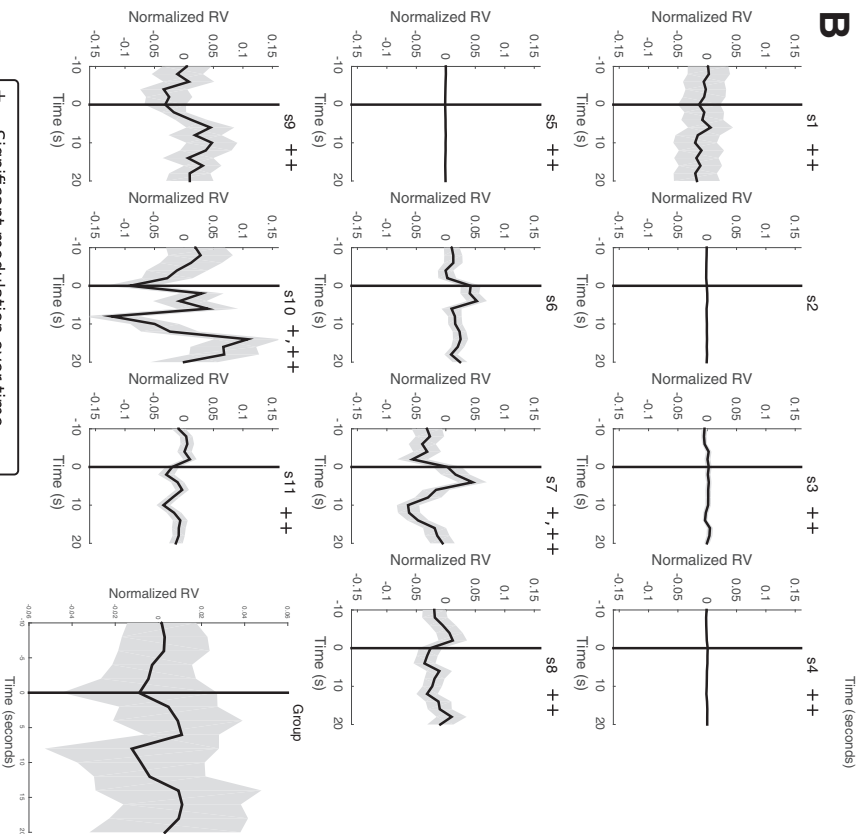
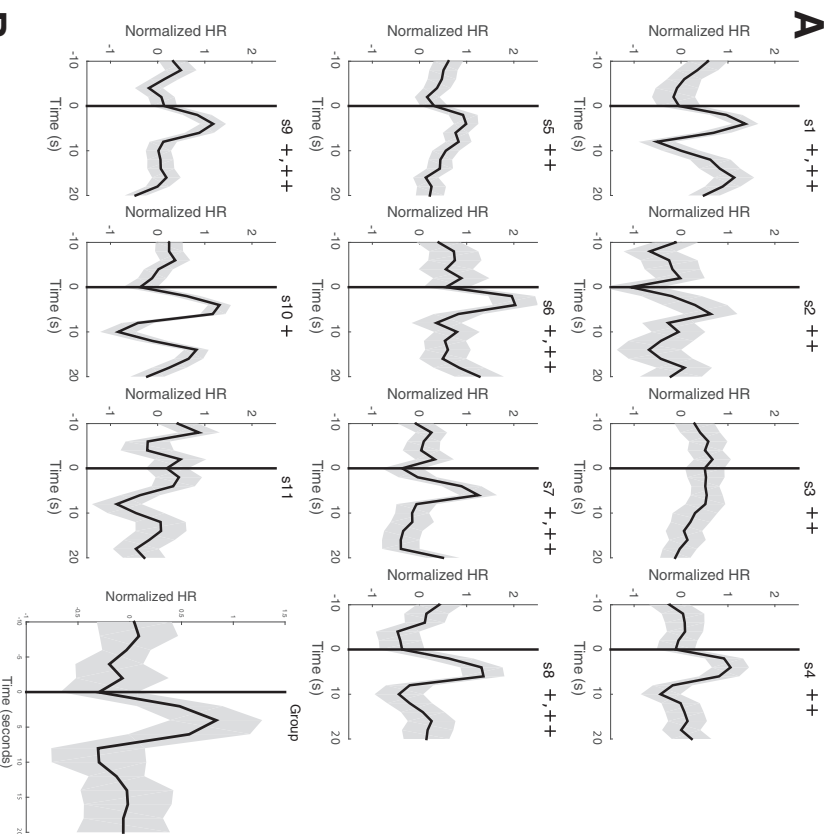
Each additional tier of models includes more parameters, and as such will increase the explained variance. To determine whether the additional parameters explained significantly more variance, pairwise F-tests were used, taking into account the changes in degrees of freedom. These were calculated according to the equation for the comparison of nested models, where a *reduced* model contains some subset of the regressors used in the *full* model:

$$F = [(RSS_{reduced} - RSS_{full}) / (P_{full} - P_{reduced})] / [RSS_{full} / (N - P_{full} - R)] \quad (2)$$

where RSS is the Residual Sum of Squares,  $P$  the number of regressors (or parameters),  $N$  the number of data points in each fMRI time series (run),  $R$  the number of runs, and where the *reduced* model is the model that includes fewer physiological parameters than the *full* model, at each level of comparison.

We note that transfer functions have been proposed to describe the relationship between HR (Chang et al., 2009) or RV (Birn et al., 2008) and the BOLD signal. These functions may be convolved with the HR and RV time series to produce a single regressor describing the predicted changes in each variable. For our study, it would have been possible to use a Finite Impulse Response expansion model for the deconvolution of HRFs, and then use this transfer function for all other analyses. However, given the variation in HRF shape across individuals, tasks, and brain regions (Handwerker et al., 2004), we were interested in whether the transfer function might also differ across brain regions. In particular, the cerebellum has not been featured in the calculation of these functions, and, in fact, is often cut off by the slice selection procedure. As such, the current study provides a first effort to measure cerebellar-specific influences of RV and HR. Thus we chose to estimate responses from a set of lagged regressors. This agnostic approach imposes no constraint on the shape of the transfer function between physiological

**Fig. 3.** Change in physiological measures evoked by reaches. Individual and group mean evoked change in heart rate (HR, panel A) and respiratory variation (RV, panel B), normalized by subtracting the mean values across each run. Shaded areas represent standard error of the mean across movement trials in individual plots, and standard error of the mean across participants in group plot. Vertical lines indicate movement cue onset, time = 0 s. Participants whose movement-evoked changes in HR or RV were significant over either 5- or 11-TR windows, and whose model fits were significantly improved by the inclusion of regressors based on these measures, are noted (see legend).



+ Significant modulation over time  
 ++ Significant improvement in model fit  
 from inclusion of regressors in GLM

variables and the BOLD signal. This allows for the flexible allocation of explainable variance to any of the lagged regressors, rather than designating certain lags to bias this allocation.

## Results

### Head motion

Use of sagittal slices ensured that the majority of head movement would be in-plane, mitigating its influence. Head motion during the task was minimal, likely because the participants were restrained by a bite bar and completed a training session within the scanner environment. The maximum excursion of the head in any direction was calculated for each of the localizer and task runs. The median maximum excursion was 0.3 mm (mean = 0.4 mm, sd = 0.2 mm) for the localizer run and 0.2 mm (mean = 0.2 mm, sd = 0.8 mm) for the task runs. Head position drifted more than 1 mm in only two individuals, both during the initial localizer run (1.3 mm and 1.4 mm).

### Evoked change in heart rate and respiration

Across participants, each reach induced an increase in heart rate of about 1%, peaking at 4 s after movement onset and returning to baseline at approximately 8 s (Fig. 3A). This result is in agreement with previous reports of the influence of movement on HR (Damen and Brunia, 1987; Jennings et al., 1991, 1992; Schlerf et al., 2012). A negligible mean increase of about 0.0015% was observed for RV, with two broader peaks, one around 6 s after the movement cue and the other at around 16 s (Fig. 3B).

To statistically evaluate these changes, one-way repeated-measures ANOVAs were conducted on the two dependent variables (HR, RV). Four time windows were used in separate ANOVAs to identify short- and long-term changes evoked by movements, as well as changes leading up to the movement and longer-scale fluctuations. The first window was limited to the five samples after the movement, spanning 0–10 s. The second and third windows each included eleven samples: one extended from 20 s prior to movement cue onset and included the sample coincident with the movement cue onset (time = 0 s), while the other extended from the movement cue onset to 20 s following the movement cue onset. The fourth window included 21 samples, starting 20 s prior to movement cue onset and extending to 20 s after the movement cue onset.

At the group level, the effect of time on HR was significant for all four time windows, although the effect was considerably weaker in the window preceding the movement cue onset, and would not survive a correction for multiple comparisons (5 sample:  $F(4,10) = 17.12$ ,  $p < .00001$ ; 11 sample preceding:  $F(10,10) = 2.05$ ,  $p = .04$ ; 11 sample following:  $F(10,10) = 12.06$ ,  $p < .0001$ ; 21 sample:  $F(20,10) = 7.85$ ,  $p < .00001$ ). In an analysis of each individual (Fig. 3), six of the 11 participants' data sets contained significant effects ( $p < .05$ ) of time on HR over the 5-sample window, four over the 11-sample window following movements, and two over the 21-sample window. No participant's data contained a significant effect of time on HR over the 11-sample window preceding movements.

At the group level, the effect of time was not significant for RV over any time window (5 sample:  $F(4,10) = 1.68$ ,  $p = .17$ , 11 sample preceding:  $F(10,10) = 0.88$ ,  $p = .56$ ; 11 sample following:  $F(10,10) = 1.22$ ;  $p = .29$ , 21 sample:  $F(20,10) = 1.24$ ;  $p = .23$ ). In the individual analysis, one participant's data contained a significant effect of time ( $p < .05$ ) on RV over the 5-sample window, two participants' data were significant over the 11-sample window following movements, and one was significant over the 21-sample window. No participant's data contained a significant effect of time on RV over the 11-sample window preceding movements.

### Whole-brain effects of HR, RV, phase, and movement regressors

Fig. 4 displays the group-level whole-brain F-statistic maps, thresholded at  $p < .001$ , uncorrected (see [Assessment of task-related changes in physiological variables and creation of physiological regressors](#) section). The effect of HR was widespread throughout the brain, including both M1 and the cerebellum. The effect was located primarily in gray matter and extending to some subcortical structures. In contrast, very few (only five total) voxels reached the same threshold for the effect of RV. The effect of phase was moderate, primarily evident in periventricular spaces and other loci near transitions between tissue types. This pattern is expected as phase correction is intended to correct for bulk motion of the head and chest cavity, as well as pulsatile (cardiac) motion (Glover et al., 2000).

A parallel analysis was used to test for the effect of the movement onset regressors (for any of the 12 lags). This produced an F-statistic map revealing widespread effects of arm movements (Fig. 4D). Note that this effect is not restricted to motor structures; rather, this unconstrained probe is likely sensitive to a range of effects correlated with many aspects of movement with differing time courses. However, the regions of greatest effect in Fig. 4D correspond to regions activated by movement in the localizer run.

Movement- and physiological variable-related activations appear to be greatly diminished in the lateral aspects of the brain. As noted in [Imaging parameters](#) section 2.3, our acquisition parameters ensured cerebellar coverage but did not always permit whole-brain coverage, and excluded these lateral regions in participants with wider brains.

### Functionally defined regions of interest

The localizer run was used to identify functional ROIs. Fig. 2A shows a group-level t-statistic map of the contrast Move vs. Rest during the localizer run, warped to MNI space and thresholded at  $p < .001$ , uncorrected (Table 1). Movements during this localizer activated a network of regions including the left pre- and post-central gyri, regions within the cingulate, bilateral insular, and right lateral frontal cortices, and the cerebellum bilaterally (Fig. 2A).

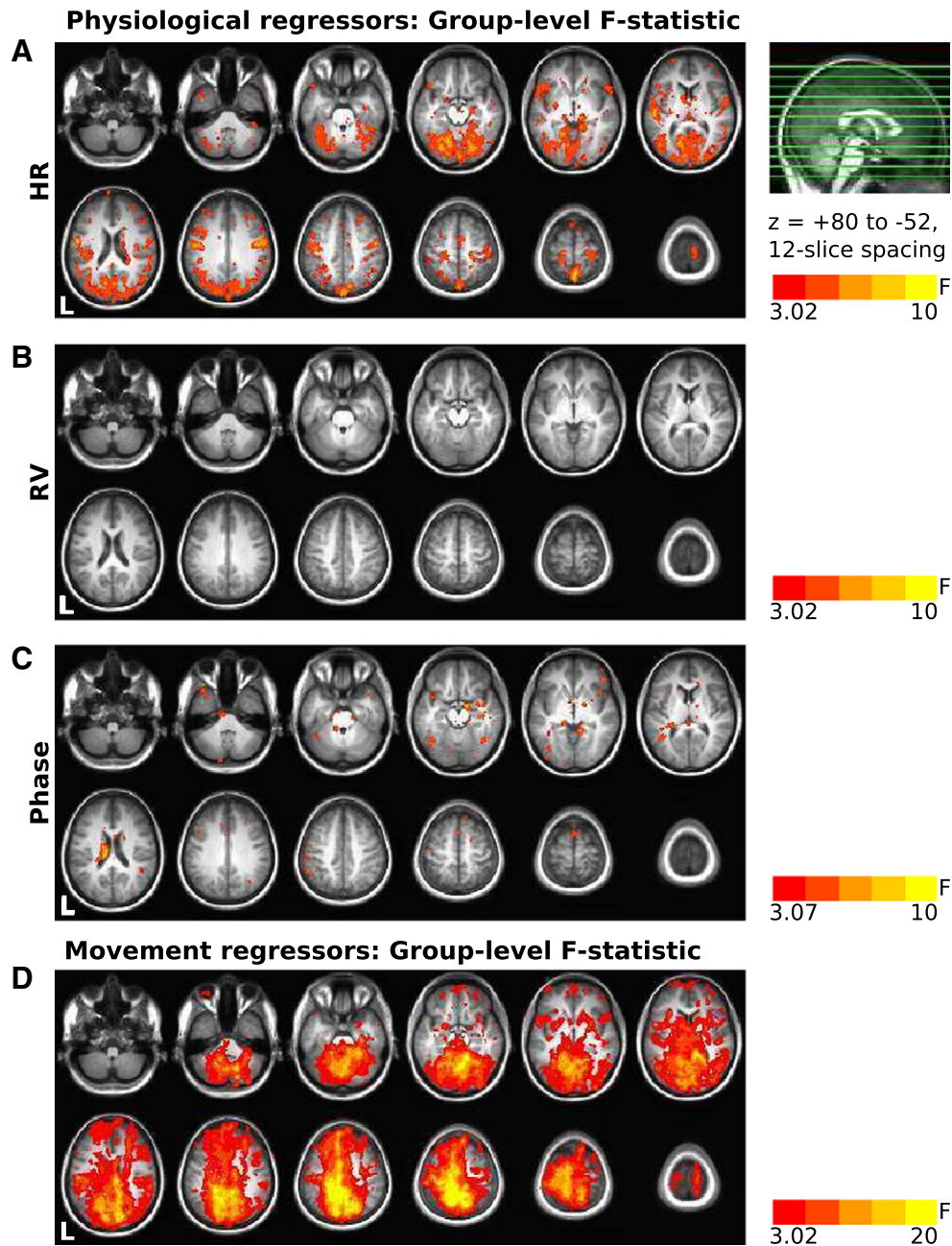
Table 1 lists the coordinates of the centers of the ROIs, along with the t-statistic thresholds used to restrict the ROIs to only active voxels. Fig. 2 show the locations and overlap of ROIs for M1 and cerebellum, respectively. Although these ROIs were defined using t-statistic maps for each individual with the contrast Move vs. Rest, the overlap map is in good alignment with the activation patterns seen at the group level.

### HRF estimation

Prior to estimating the shape of the movement-related hemodynamic response in our two ROIs, we normalized the BOLD time series by dividing the raw BOLD signal at each time point and at each voxel by the mean raw BOLD signal for that voxel in that run. We plotted the movement-evoked percent signal change, both the overall group average (thick black line, Fig. 2D&E) and average across two task runs for each individual (thin lines, one color per individual, Fig. 2D&E). The largest movement-evoked increase in the BOLD signal was 0.55% in M1, and 0.39% in the cerebellum. These magnitudes are well within the expected range of evoked BOLD responses at 3 T. There was variation across participants in both the amplitude and latency of the peak. This variation across participants contributed to (but did not wholly account for) the emergence of a double peak in the cerebellum.

The HRF within each M1 and cerebellar ROI was estimated using eight separate GLM analyses: the uncorrected model and the seven corrected models (GLMs that included HR, RV, and phase alone, all pairwise combinations, and the Fully Corrected model that includes all three variables). The mean parameter estimates at each lag in the





**Fig. 4.** Effect of physiological and movement regressors across the whole brain. Group-level F-statistic maps indicate where HR (A), RV (B), Phase (C), and movement (D) regressors had a positive or negative effect at any lag (the effect of physiological regressors represents the additional variance beyond movement regressors). These maps were created using a group-level F-contrast over the parameter estimates for all regressors within each set of regressors (12 lags of HR, RV, or movement onsets or 8 phase regressors). Color bar at right indicates group-level F-statistic. Midsagittal view of slices for all panels is shown on the right. All maps are thresholded at  $p < .001$ , uncorrected. Color bars indicate threshold for each variable based on associated degrees of freedom.

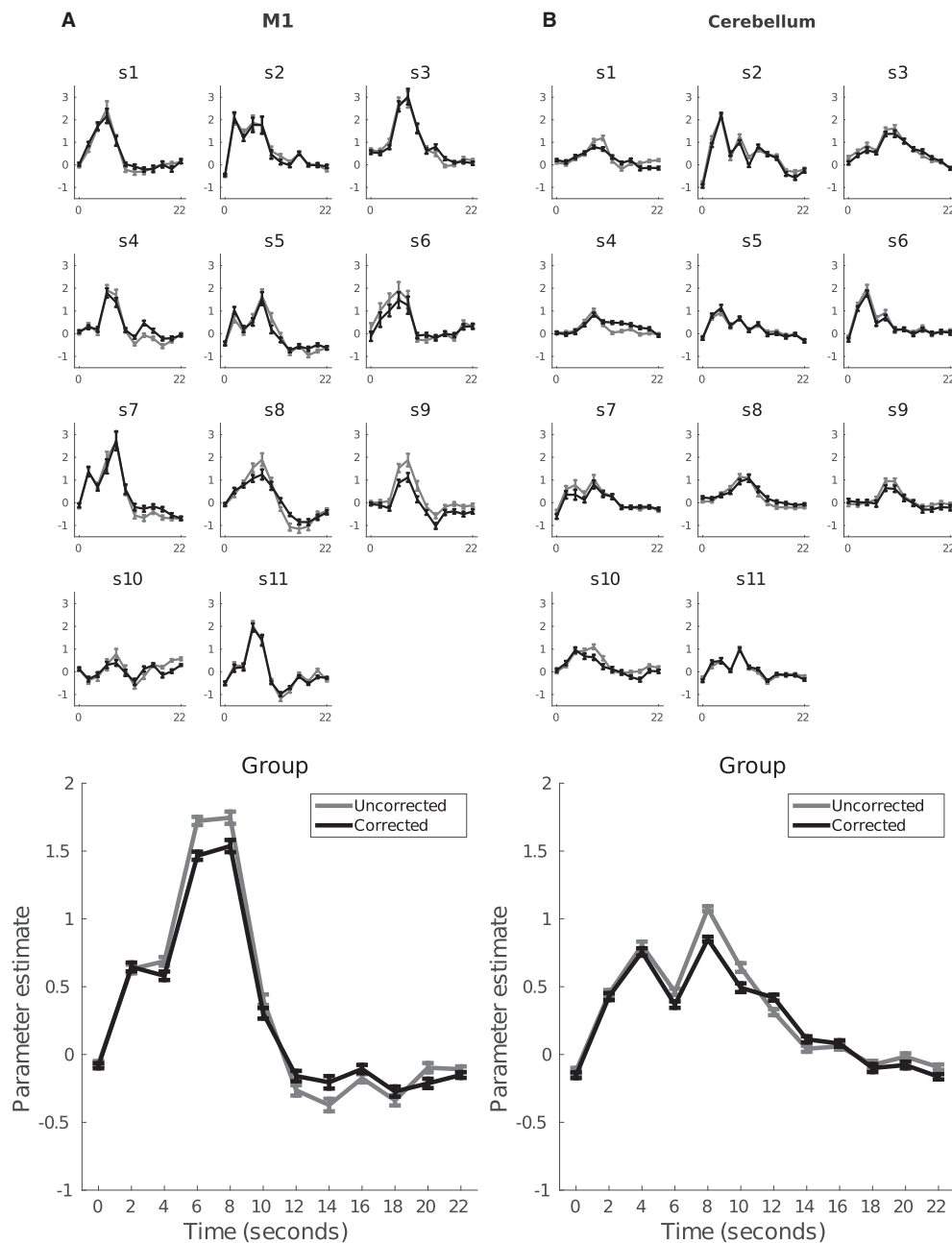
Move vs. Rest contrast are plotted in Fig. 5. These plots represent the estimated hemodynamic response to reaching movements in each ROI. Because differences among the models were minimal, we only show the results from the Uncorrected and Fully Corrected models, although all models were included in the statistical tests.

The HRFs for each of the Corrected models were compared to the Uncorrected model using seven 2-way repeated-measures ANOVAs, with factors Time (12 TRs, beginning with movement cue onset and extending for 22 s) and Model (Corrected vs. Uncorrected model), for a total of seven ANOVAs.

All ANOVAs revealed a significant main effect of time, as is expected when plotting a robust HRF. While the effect of Model was not

significant for any of the ANOVAs, all comparisons in which the Corrected model included HR as a regressor yielded a significant interaction effect for both the M1 and cerebellar ROIs ( $p < .0001$  in all cases where HR was included, 4 of 7 models; see Table 2).

Including HR resulted in a slightly lower peak in the cerebellar HRF (Fig. 5B). A similar pattern was observed for M1, with the addition of a more pronounced reduction in the post-stimulus undershoot in comparison with the cerebellar HRF. We compared individual time points between the corrected and uncorrected models using *post hoc* paired *t*-tests, using Bonferroni adjusted alpha levels of  $p < .004$  per test (.05/12). No time point reached significance at this threshold. As Bonferroni correction is conservative, however, we note that only a single time



**Fig. 5.** Deconvolved hemodynamic response functions (HRFs). HRFs deconvolved from each participant's ROI for the Uncorrected (gray) and Fully Corrected (black) models in M1 (A) and Cerebellum (B). Bottom row of each panel shows the mean HRF across all participants. Axes and legends for individual plots are as labeled in group plots. In individual plots, error bars represent the standard error of the mean across voxels; in group plot, error bars are within-participant (Loftus) error bars, given that our statistical test of interest (i.e., uncorrected vs. corrected) is conducted within-participant.

point contained a significant difference prior to correction, the fifth time point in the cerebellum (HR model:  $t(20) = 2.1731$ ,  $p < .05$ ; HR + phase:  $t(20) = 2.1214$ ,  $p < .05$ ; HR + RV + phase:  $t(20) = 2.4560$ ,  $p < .05$ ).

#### Percent variance explained and model comparisons

We next asked whether the inclusion of task-dependent effects on the physiological variables improved the fit of our eight GLMs. Table 3 reports the percent variance explained by each of the models, listing the mean, standard error of the mean, minimum, and maximum across participants. For all models, the percent variance explained was greater in M1 compared to the cerebellum. This result is not surprising given that the responses in M1 tended to have higher peaks, were more

consistent across participants, and bear a stronger resemblance to the canonical HRF. Inspection of whole-brain maps of temporal signal-to-noise ratio (tSNR) revealed slightly lower signal quality in the cerebellum, which may also have influenced our estimates. It is also possible that our ROI selection procedure resulted in greater heterogeneity of responses across voxels in the cerebellar ROI since this ROI could extend beyond the “classic” cerebellar motor region of the anterior lobe.

For each region, the inclusion of a new set of regressors (individually or in combination) yielded an increase of variance explained between 6.9 and 12.3%. The addition of HR and RV regressors tended to improve model fits more than phase regressors. Fig. 6 illustrates, in the percent-ages adjacent to each arrow, the mean additional percent variance explained by the addition of each set of regressors. Overall, for M1, the increase for the Fully Corrected model compared to the Uncorrected

**Table 3**

Percent variance explained.

Mean percent variance explained, standard error of the mean, and minima and maxima values for percent variance explained. Top is for M1 ROI and bottom is for cerebellar ROI.

| Model           | M1 mean | M1 sem | M1 min | M1 max |
|-----------------|---------|--------|--------|--------|
| Uncorrected     | 24.6    | 1.8    | 12.6   | 34.7   |
| RV              | 36.9    | 1.4    | 31.5   | 47.1   |
| HR              | 36.1    | 1.8    | 27.6   | 45.0   |
| Phase           | 32.5    | 1.9    | 19.4   | 41.4   |
| RV + HR         | 46.6    | 1.4    | 40.2   | 54.8   |
| RV + phase      | 43.7    | 1.3    | 36.1   | 51.1   |
| HR + phase      | 43.3    | 1.7    | 34.2   | 50.9   |
| RV + HR + phase | 52.6    | 1.2    | 45.7   | 58.6   |
| Model           | Ce mean | Ce sem | Ce min | Ce max |
| Uncorrected     | 15.8    | 1.3    | 10.2   | 23.4   |
| RV              | 26.3    | 1.8    | 20.3   | 38.3   |
| HR              | 26.2    | 1.8    | 19.9   | 36.0   |
| Phase           | 23.1    | 1.3    | 16.3   | 31.3   |
| RV + HR         | 26.2    | 1.9    | 29.2   | 45.9   |
| RV + phase      | 33.3    | 1.7    | 26.7   | 43.3   |
| HR + phase      | 33.2    | 1.8    | 25.8   | 42.5   |
| RV + HR + phase | 41.7    | 1.9    | 35.2   | 51.3   |

model ranged from 23.9% to 37.5% across participants. The comparable range for the cerebellum was 22.8% to 31.1%.

A nested series of pairwise F-tests were used to test, in the individual data sets, the impact of adding each set of physiological regressors. For each F-test, a given model was compared to a subordinate, reduced model from which one set of regressors had been removed (see Eq. (2)). This allowed quantification of the improvement in model fit due to the addition of a given variable as defined by a set of regressors. Fig. 6 presents, in the number beneath the additional percent variance explained by each correction, the number of participants for whom that correction resulted in a significant improvement in model fit. In M1, adding HR and RV regressors significantly improved the fit for a majority of participants, while adding phase did not. In the cerebellum, the number of participants with significantly improved fit was lower, and there were no substantial differences when adding HR, RV, or phase regressors.

Given the similarity of the mean additional percent variance explained in M1 and the cerebellum, the disparity in number of participants with significantly improved fit is somewhat surprising. Relaxing the criterion to  $p < .10$  adds 1–2 participants at most levels of correction (as diagrammed in Fig. 6); however, the disparity between M1 and the cerebellum is maintained. Paired *t*-tests on the additional percent variance explained for each correction were used to directly compare the

relative efficacy of these corrections in M1 and the cerebellum. No significant differences were found for any of the nine corrections ( $p > 0.15$  for all). Thus, while the effects tended to reach statistical significance for individual participants in M1 only, the overall magnitude and pattern of improvements was quite similar for both M1 and the cerebellum.

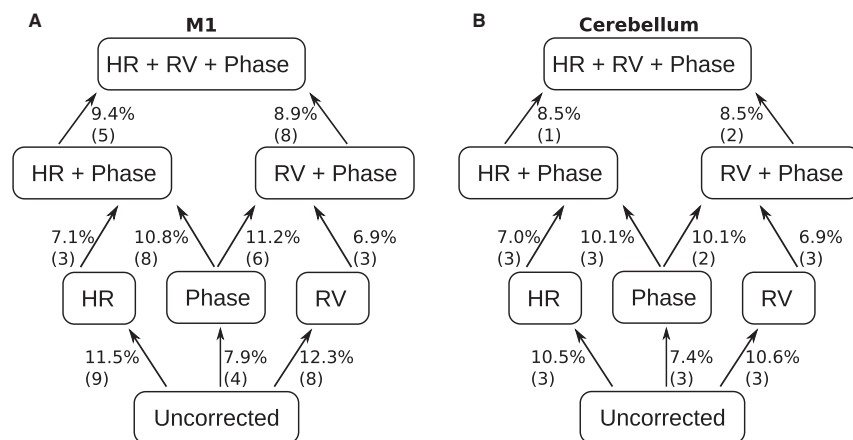
## Discussion

We examined how task-related changes in HR and RV influence the BOLD response and, as such, might impact the interpretation of neural activity. At the group level, reliable task-related fluctuations in HR resulted in subtle alterations in the shape of the HRF, while more variable, and nonsignificant, fluctuations in RV did not affect the shape of the HRF. The inclusion of these physiological regressors accounted for a significantly larger percent of the variance than analyses involving only the movement regressors, with the effect of HR and RV regressors producing roughly similar improvements in the individual data sets. This improvement in model fit was greater than that observed with regressors that allowed for phase-based corrections. The improvements in model fit were greater in M1 than in the cerebellum.

### Improvements in model fit obtained using nuisance variable regression

fMRI studies that seek to account for physiological fluctuations typically include HR and RV as nuisance regressors in the GLM. Indeed, in analyses of functional connectivity, these physiological corrections are increasingly commonplace (e.g. Birn et al., 2008, 2009; Chang et al., 2013; Khalili-Mahani et al., 2013). The benefits of nuisance variable regression for task-related fMRI, however, have not been widely considered (but see van 't Ent et al., 2014). Schlerf et al. (2012) provide an extreme example in which the failure to account for task-related changes in HR on fMRI responses led to radically different interpretation than that obtained when HR changes were included in a GLM.

The current study used a nested approach that looked at a set of physiological variables, either modeled individually or in combination within the GLMs. At the group level, improvements in model fit were assessed using the mean additional variance explained. With the fully corrected model, the mean percent variance explained increased from 24.6% to 52.6% in M1 and from 15.8% to 41.7% in the cerebellum (Table 3). In both regions, inclusion of RV, HR, and phase regressors in isolation resulted in mean improvements of around 10% (Fig. 6). These increases are considerably higher than in one previous resting state study in which the inclusion of cardiac rate regressors only explained an additional 1% of the variance (Shmueli et al., 2007). This difference



**Fig. 6.** Model comparisons: Increase in percent variance explained and nested F-tests. Percentage values represent the mean additional percent variance explained by each correction (percent variance explained in the full model minus that of the reduced model, as in Eq. (2)) for M1 (A) and Cerebellum (B) ROIs. Values in parentheses are the number of the 11 participants for whom each correction resulted in a significant improvement in model fit ( $p < 0.05$ ), accounting for the reduction in the degrees of freedom due to additional regressors.

is likely due to the restricted nature of our task- and ROI-based approach. Indeed, in our own previous study, the inclusion of HR regressors only accounted for 3% of BOLD variance (Schlerf et al., 2012), while the magnitude of the current correction effect is similar to that of Chang et al. (2009) who restricted their analysis to voxels where a significant proportion of the variance was explained by physiological regressors.

At the individual level, GLMs that included HR and RV as nuisance regressors resulted in improvements in a greater number of data sets than did those that included phase. Movement-evoked fluctuations in these variables did not predict their utility as nuisance regressors. Specifically, significant improvements in model fit occurred in the absence of significant task-related fluctuations for both variables and, in one instance (see s10, Fig. 3A), accounting for changes in HR did not predict an improvement in model fit despite a task-related fluctuation in HR.

Including cardiac and respiratory phases, as implemented using RETROICOR and other methods, has been shown to provide a significant improvement over uncorrected models (Hu et al., 1995; Glover et al., 2000). The rate of cardiac and respiratory events are independent regressors that may account for additional variance. By using a set of nested GLM analyses, we were able to determine whether additional sets of regressors accounted for redundant variance that could equally well be attributed to alternative regressors. Our analysis confirmed that these regressors linearly improve the model fit, suggesting that they do in fact manifest differently in the BOLD signal. For instance, tracking the influence across the levels of the M1 tree in Fig. 6, the inclusion of HR accounts about 10% additional variance when added to the uncorrected model, to a model that already includes phase, or to a model that already includes both RV and phase.

Prior studies have examined the redundancy of information in physiological regressors, and these have led to the recommendation that the regressors be consolidated, either by averaging correlated regressors or by replacing sets of shifted timeseries with single regressors convolved with an appropriate response function (respiratory, Birn et al., 2008, and cardiac, Chang et al., 2009). Our shape-agnostic FIR approach, with 11 additional regressors probing the lagging influence of these processes, reduced the potential impact of regional and temporal variation on analyses looking at the relationship between physiological variables and the BOLD signal. However, our approach does entail a greater reduction in the degrees of freedom, and this may have increased the prevalence of null results at the individual level. Evaluating the improvement of this approach over a canonical response function is left for future studies.

Overall, the inclusion of the physiological regressors at the individual level improved the fit more reliably in M1 than in the cerebellum. Adding HR to the GLM produced a significant improvement in fit in the cerebellar ROI for only three participants (compared to nine for M1), a result that was unexpected given the cerebellar findings of Schlerf et al. (2012). There are, however, a number of noteworthy differences between the studies. First, in the present study, the temporal signal-to-noise ratio (tSNR) was higher in M1 than in the cerebellum, and this may have limited the extent of improvement that was possible in the cerebellum. Second, our movement-related ROIs were comprised predominantly of voxels in lobules I–V, whereas the error-related activity in the previous study was predominantly located in lobule VI. There are likely notable differences in functionality between these regions, a point underscored by resting state connectivity data showing differentially connectivity patterns; for example, lobules I–V show greater connectivity with M1 whereas lobule VI shows greater connectivity with premotor and prefrontal cortex (Krienen and Buckner, 2009). Third, the movement task used in the current study did not employ visual targets or involve visual feedback, and as such had reduced cognitive demands. Finally, the movement vs. rest contrast in the current study typically produces more reliable activations in the cerebellum (and M1) than the errors vs. correct trials contrast examined by Schlerf et al. (2012). Accounting for a confounding factor such as HR may be

more important for statistical contrasts that produce less reliable results.

#### *Task-related changes in physiological variables and their impact on the hemodynamic response*

Using a simple movement task in a slow event-related paradigm, we observed a robust task-related increase in HR of about 1%. While the increase in HR (peaking around 4 s) occurred slightly faster than the canonical hemodynamic response (which peaks at around 6 s), the overlap in the time courses of these two responses is considerable and an issue of concern. Since each beat of the heart delivers oxygenated hemoglobin to the brain, and the BOLD signal is a measurement of the ratio of deoxygenated and oxygenated hemoglobin, changes in HR will modulate the BOLD signal. As such, HR-induced changes in the BOLD signal may be misattributed to neural activity. That is, task-related changes in HR could, at least in some cases, be so closely linked to the task parameters as to be indistinguishable from true neural changes. Given the relative paucity of studies examining cerebellar hemodynamic responses, combined with the previously observed susceptibility of the cerebellum to physiological changes (Schlerf et al., 2012), we sought to pinpoint sources of variability in the shape of the cerebellar HRF, as well as compare this to that observed in the cortex.

Many fMRI analysis packages, by default, convolve task regressors with a canonical HRF to model event-related responses. This may produce an inappropriate model of the data for some subjects, resulting in poor fit (for review, see Handwerker et al., 2012). Handwerker et al. (2004) provided an example where a 1 s error in an individual-specific HRF reduced the explained variance by 10%, and at 2 s error the explained variance was reduced by 38%. Given that the HRF varies greatly across individuals, it has been recommended that a separate task run be included to obtain individual-specific HRFs (Aguirre et al., 1998; D'Esposito et al., 1999; Handwerker et al., 2004). While this approach can account for relatively large inter-individual differences, smaller differences may remain in the shape of the HRF for different brain regions. Moreover, these regional differences may be differentially influenced by physiological factors depending on their proximity to vasculature and sulci (Birn et al., 2006, 2008).

Our analyses of the deconvolved hemodynamic response revealed subtle effects from the physiological corrections. The deconvolved HRFs in both the M1 and cerebellar ROIs were similar at the group level, although the inclusion of physiological regressors in the GLM yielded heterogeneous changes in the deconvolved HRFs of individual participants. Given the variation in the HRF across individuals, brain regions, and task parameters (Handwerker et al., 2004; Birn et al., 2008), as well as the spatial heterogeneity of susceptibility to physiological influences across the brain (Birn et al., 2008; Chang et al., 2009), differences caused by task-related physiological covariates are likely to be minor and irrelevant to the decision to use the canonical HRF to model task responses. However, the current results (Table 2) point to an important role for HR correction in accurately estimating the hemodynamic response, particularly when a task-related change in HR is present.

In the group data, the HR correction lowered the peak of the estimated HRF in both M1 and the cerebellum. The lower peaks indicate that some of the variance attributed to movement-related neural activity (via the Move vs. Rest contrast) was better accounted for by the HR correction. This observation is in accord with the known impact of HR changes on the BOLD signal (Shmueli et al., 2007). This relationship has been described by a cardiac response function (CRF), deconvolved from data recorded while subjects were lying at rest in the scanner. With a deconvolved CRF, HR is positively correlated with the BOLD signal at lags of around 4 s and negatively at lags around 12 s (Chang et al., 2009). The initial positive correlation, combined with the increase in HR induced by our movement task, offers an explanation for the reduced HRF peaks following addition of HR regressors (Fig. 5): Variance that had previously been misattributed to neural responses was re-



allocated to physiological variables, decreasing the parameter estimates for the movement regressors at these lags.

The generalizability of estimated HRFs can be improved through regularization (Casanova et al., 2008). Here, we chose to focus on the unconstrained HRFs, estimated from the FIR expansion model. If our estimated HRFs were to be used as basis functions with which to convolve task regressors in a separate GLM, regularization would be desirable. However, we were primarily interested in whether the inclusion of physiological variables in a GLM would result in a vastly diminished HRF peak due to the association of variance from these physiological variables to task variables. Given the small change in HRF shape produced by our corrections (Fig. 5), we simply reiterate the cautious recommendations of Aguirre et al. (1998) and Handwerker et al. (2004), and suggest that region- and subject-specific HRFs used as basis functions be regularized where doubt remains as to the accuracy of the deconvolved transfer function.

Because of the involuntary nature of the autonomic process, proper experimental control of HR is extremely difficult (for examples of experimental control of respiration, see Birn et al., 2006, 2008). However, the discrete nature of heartbeats has the advantage of reducing the number of possible sources of variance in the BOLD signal compared to changes in either the rate or depth of respiration. Importantly, a movement task such as that employed here can be a reliable modulator of HR. Indeed, a movement task might be used to drive a change in HR should an experimenter wish to explore individual physiological response functions. Convolving physiological regressors with this function avoids the loss of freedom imposed by an exhaustive lagged approach. Furthermore, modeling individual-specific changes in HR due to movements and other events could account for unwanted sources of variance across a range of tasks. Other cognitively or emotionally demanding tasks are also known to influence HR, and, as such, the BOLD response would likely be influenced by the inclusion of HR regressors in a GLM.

RV was not reliably modulated during our simple movement task. The lack of an effect of RV regressors on the shape of the hemodynamic response is consistent with our previous study involving an event-related analysis of movement errors (Schlerf et al., 2012). Naturally-occurring changes in respiration may result from some combination of the depth, duration, and frequency of breaths, necessitating tighter experimental control through cued breathing tasks in order to independently assess the sources of variance (Birn et al., 2006, 2008, 2009). However, it is unclear to what extent cued changes in respiration mimic naturally-occurring changes in respiration (Birn et al., 2008). Further research on the impact of RV on fMRI signals is needed, given the lack of effect of RV on the HRF but comparable efficacy of HR and RV in nuisance variable regression in the current study (Table 3, Fig. 6).

## Conclusions

Sources of variance in fMRI studies that are unaccounted for may influence the shape of the hemodynamic response function (HRF) and/or model fit. These unmodeled variables may be particularly problematic when they are correlated with task factors. We modeled a set of physiological variables as nuisance regressors to examine their impact on the deconvolved hemodynamic response and on model fit. Regressing out task-related changes in heart rate (HR) resulted in subtle changes in the HRF in both the cerebellum and M1, with or without the presence of respiratory variation (RV) and phase regressors. These results indicate that the BOLD response may be more accurately estimated by including HR in a GLM even when changes in HR are correlated with changes in the contrast of interest. As assessed by model fit at the group level, the inclusion of both RV and HR regressors resulted in significant improvements in model fit exceeding those of phase regressors, further underscoring the need to account for both task-related and task-independent changes in the rate of physiological processes when inferring patterns of neural activity from fMRI time series.

## Acknowledgments

The authors are grateful to Ben Inglis and Rick Redfern.

This work was supported by the National Science Foundation Graduate Research Fellowship Program and the National Institutes of Health (NS074917, HD060306). Data were collected at the Henry J. Wheeler, Jr. Brain Imaging Center at the University of California, Berkeley, which receives support from the National Science Foundation through their Major Research Instrumentation Program, award number BCS-0821855.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2016.02.068>.

## References

- Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998. The variability of human, BOLD hemodynamic responses. *NeuroImage* 8, 360–369. <http://dx.doi.org/10.1006/nimg.1998.0369>.
- Birn, R.M., 2012. The role of physiological noise in resting-state functional connectivity. *NeuroImage* 62, 864–870. <http://dx.doi.org/10.1016/j.neuroimage.2012.01.016>.
- Birn, R.M., Diamond, J.B., Smith, M.A., Bandettini, P.A., 2006. Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. *NeuroImage* 31, 1536–1548. <http://dx.doi.org/10.1016/j.neuroimage.2006.02.048>.
- 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. <http://dx.doi.org/10.1016/j.neuroimage.2007.11.059>.
- Birn, R.M., Murphy, K., Handwerker, D.A., Bandettini, P.A., 2009. fMRI in the presence of task-correlated breathing variations. *NeuroImage* 47, 1092–1104. <http://dx.doi.org/10.1016/j.neuroimage.2009.05.030>.
- 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.
- Buxton, R.B., Frank, L.R., 1997. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *J. Cereb. Blood Flow Metab.* 17, 64–72. <http://dx.doi.org/10.1097/00004647-199701000-00009>.
- Casanova, R., Ryali, S., Serences, J., Yang, L., Kraft, R., Laurienti, P.J., Maldjian, J.A., 2008. The impact of temporal regularization on estimates of the BOLD hemodynamic response function: a comparative analysis. *NeuroImage* 40, 1606–1618. <http://dx.doi.org/10.1016/j.neuroimage.2008.01.011>.
- Chang, C., Glover, G.H., 2009. Relationship between respiration, end-tidal CO<sub>2</sub>, and BOLD signals in resting-state fMRI. *NeuroImage* 47, 1381–1393. <http://dx.doi.org/10.1016/j.neuroimage.2009.04.048>.
- Chang, C., Cunningham, J.P., Glover, G.H., 2009. Influence of heart rate on the BOLD signal: the cardiac response function. *NeuroImage* 44, 857–869. <http://dx.doi.org/10.1016/j.neuroimage.2008.09.029>.
- Chang, C., Metzger, C.D., Glover, G.H., Duyn, J.H., Heinze, H.-J., Walter, M., 2013. Association between heart rate variability and fluctuations in resting-state functional connectivity. *NeuroImage* 68, 93–104. <http://dx.doi.org/10.1016/j.neuroimage.2012.11.038>.
- Critchley, H.D., 2009. Psychophysiology of neural, cognitive and affective integration: fMRI and autonomic indicators. *Int. J. Psychophysiol.* 73, 88–94. <http://dx.doi.org/10.1016/j.ijpsycho.2009.01.012>.
- Crone, E.A., van der Veen, F.M., van der Molen, M.W., Somsen, R.J.M., van Beek, B., Jennings, J.R., 2003. Cardiac concomitants of feedback processing. *Biol. Psychol.* 64, 143–156.
- Crone, E.A., Bunge, S.A., de Klerk, P., van der Molen, M.W., 2005. Cardiac concomitants of performance monitoring: context dependence and individual differences. *Brain Res. Cogn. Brain Res.* 23, 93–106. <http://dx.doi.org/10.1016/j.cogbrainres.2005.01.009>.
- Dagli, M.S., Ingelholm, J.E., Haxby, J.V., 1999. Localization of cardiac-induced signal change in fMRI. *NeuroImage* 9, 407–415. <http://dx.doi.org/10.1006/nimg.1998.0424>.
- Dale, A.M., 1999. Optimal experimental design for event-related fMRI. *Hum. Brain Mapp.* 8, 109–114.
- Dale, A.M., Buckner, R.L., 1997. Selective averaging of rapidly presented individual trials using fMRI. *Hum. Brain Mapp.* 5, 329–340. [http://dx.doi.org/10.1002/\(SICI\)1097-0193\(1997\)5:5<329::AID-HBM1>3.0.CO;2-5](http://dx.doi.org/10.1002/(SICI)1097-0193(1997)5:5<329::AID-HBM1>3.0.CO;2-5).
- Damen, E.J., Brunia, C.H., 1987. Changes in heart rate and slow brain potentials related to motor preparation and stimulus anticipation in a time estimation task. *Psychophysiology* 24, 700–713.
- D'Esposito, M., Zarahn, E., Aguirre, G.K., Rypma, B., 1999. The effect of normal aging on the coupling of neural activity to the bold hemodynamic response. *NeuroImage* 10, 6–14. <http://dx.doi.org/10.1006/nimg.1999.0444>.
- Diedrichsen, J., Balsters, J.H., Flavell, J., Cussans, E., Ramnani, N., 2009. A probabilistic MR atlas of the human cerebellum. *NeuroImage* 46, 39–46. <http://dx.doi.org/10.1016/j.neuroimage.2009.01.045>.
- Diedrichsen, J., Hashambhoy, Y., Rane, T., Shadmehr, R., 2005. Neural correlates of reach errors. *J. Neurosci.* 25, 9919–9931. <http://dx.doi.org/10.1523/JNEUROSCI.1874-05.2005>.
- Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S.J., 1994. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210. <http://dx.doi.org/10.1002/hbm.460020402>.

- Friston, K.J., Frith, C.D., Turner, R., Frackowiak, R.S., 1995. Characterizing evoked hemodynamics with fMRI. *NeuroImage* 2, 157–165.
- Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R., 1998. Event-related fMRI: characterizing differential responses. *NeuroImage* 7, 30–40. <http://dx.doi.org/10.1006/nimg.1997.0306>.
- Glover, G.H., Li, T.Q., Ress, D., 2000. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn. Reson. Med.* 44, 162–167.
- Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V., 2003. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 100, 253–258. <http://dx.doi.org/10.1073/pnas.0135058100>.
- 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. <http://dx.doi.org/10.1016/j.neuroimage.2003.11.029>.
- Handwerker, D.A., Gonzalez-Castillo, J., D'Esposito, M., Bandettini, P.A., 2012. The continuing challenge of understanding and modeling hemodynamic variation in fMRI. *NeuroImage* 62, 1017–1023. <http://dx.doi.org/10.1016/j.neuroimage.2012.02.015>.
- Hernandez, L., Badre, D., Noll, D., Jonides, J., 2002. Temporal sensitivity of event-related fMRI. *NeuroImage* 17, 1018–1026.
- Hu, X., Le, T.H., Parrish, T., Erhard, P., 1995. Retrospective estimation and correction of physiological fluctuation in functional MRI. *Magn. Reson. Med.* 34, 201–212.
- Iacovella, V., Hasson, U., 2011. The relationship between BOLD signal and autonomic nervous system functions: implications for processing of “physiological noise”. *Magn. Reson. Imaging* 29, 1338–1345. <http://dx.doi.org/10.1016/j.mri.2011.03.006>.
- Jennings, J.R., van der Molen, M.W., 2002. Cardiac timing and the central regulation of action. *Psychol. Res.* 66, 337–349. <http://dx.doi.org/10.1007/s00426-002-0106-5>.
- Jennings, J.R., van der Molen, M.W., Brock, K., Somsen, R.J., 1991. Response inhibition initiates cardiac deceleration: evidence from a sensory-motor compatibility paradigm. *Psychophysiology* 28, 72–85.
- Jennings, J.R., van der Molen, M.W., Brock, K., Somsen, R.J., 1992. On the synchrony of stopping motor responses and delaying heartbeats. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 422–436.
- Khalili-Mahani, N., Chang, C., van Osch, M.J., Veer, I.M., van Buchem, M.A., Dahan, A., Beckmann, C.F., van Gerven, J.M.A., Rombouts, S.A.R.B., 2013. The impact of “physiological correction” on functional connectivity analysis of pharmacological resting state fMRI. *NeuroImage* 65, 499–510. <http://dx.doi.org/10.1016/j.neuroimage.2012.09.044>.
- Kobayashi, N., Yoshino, A., Takahashi, Y., Nomura, S., 2007. Autonomic arousal in cognitive conflict resolution. *Auton. Neurosci.* 132, 70–75. <http://dx.doi.org/10.1016/j.autneu.2006.09.004>.
- Krienen, F.M., Buckner, R.L., 2009. Segregated fronto-cerebellar circuits revealed by intrinsic functional connectivity. *Cereb. Cortex* 19, 2485–2497. <http://dx.doi.org/10.1093/cercor/bhp135>.
- Logothetis, N.K., 2003. MR imaging in the non-human primate: studies of function and of dynamic connectivity. *Curr. Opin. Neurobiol.* 13, 630–642.
- Logothetis, N.K., Wandell, B.A., 2004. Interpreting the BOLD signal. *Annu. Rev. Physiol.* 66, 735–769.
- Ogawa, S., Lee, T.M., 1990. Magnetic resonance imaging of blood vessels at high fields: in vivo and in vitro measurements and image simulation. *Magn. Reson. Med.* 16, 9–18.
- Poline, J.-B., Brett, M., 2012. The general linear model and fMRI: does love last forever? *NeuroImage* 62, 871–880. <http://dx.doi.org/10.1016/j.neuroimage.2012.01.133>.
- Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. *Proc. Natl. Acad. Sci. U. S. A.* 98, 676–682. <http://dx.doi.org/10.1073/pnas.98.2.676>.
- Raj, D., Anderson, A.W., Gore, J.C., 2001. Respiratory effects in human functional magnetic resonance imaging due to bulk susceptibility changes. *Phys. Med. Biol.* 46, 3331–3340.
- Schlerf, J., Ivry, R.B., Diedrichsen, J., 2012. Encoding of sensory prediction errors in the human cerebellum. *J. Neurosci.* 32, 4913–4922. <http://dx.doi.org/10.1523/JNEUROSCI.4504-11.2012>.
- Shmueli, K., van Gelderen, P., de Zwart, J.A., Horowitz, S.G., Fukunaga, M., Jansma, J.M., Duyn, J.H., 2007. Low-frequency fluctuations in the cardiac rate as a source of variance in the resting-state fMRI BOLD signal. *NeuroImage* 38, 306–320. <http://dx.doi.org/10.1016/j.neuroimage.2007.07.037>.
- Tursky, B., Shapiro, D., Crider, A., Kahneman, D., 1969. Pupillary, heart rate, and skin resistance changes during a mental task. *J. Exp. Psychol.* 79, 164–167.
- van 't Ent, D., den Braber, A., Rotgans, E., de Geus, E.J.C., de Munck, J.C., 2014. The use of fMRI to detect neural responses to cognitive interference and planning: evidence for a contribution of task related changes in heart rate? *J. Neurosci. Methods* 229, 97–107. <http://dx.doi.org/10.1016/j.jneumeth.2014.04.013>.
- van Buuren, M., Gladwin, T.E., Zandbelt, B.B., van den Heuvel, M., Ramsey, N.F., Kahn, R.S., Vink, M., 2009. Cardiorespiratory effects on default-mode network activity as measured with fMRI. *Hum. Brain Mapp.* 30, 3031–3042. <http://dx.doi.org/10.1002/hbm.20729>.
- Verstynen, T.D., Deshpande, V., 2011. Using pulse oximetry to account for high and low frequency physiological artifacts in the BOLD signal. *NeuroImage* 55, 1633–1644. <http://dx.doi.org/10.1016/j.neuroimage.2010.11.090>.