

RESEARCH ARTICLE

Temporal stability of the hemodynamic response function across the majority of human cerebral cortex

Amanda J. Taylor | Jung Hwan Kim  | David Ress 

Department of Neuroscience, Baylor College of Medicine, Houston, Texas, USA

Correspondence

David Ress, Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA.

Email: ress@bcm.edu**Funding information**

National Heart, Lung, and Blood Institute, Grant/Award Number: K25 HL131997; National Institute of Neurological Disorders and Stroke, Grant/Award Number: R01 NS095933

Abstract

The hemodynamic response function (HRF) measured with functional magnetic resonance imaging is generated by vascular and metabolic responses evoked by brief (<4 s) stimuli. It is known that the human HRF varies across cortex, between subjects, with stimulus paradigms, and even between different measurements in the same cortical location. However, our results demonstrate that strong HRFs are remarkably repeatable across sessions separated by time intervals up to 3 months. In this study, a multisensory stimulus was used to activate and measure the HRF across the majority of cortex (>70%, with lesser reliability observed in some areas of prefrontal cortex). HRFs were measured with high spatial resolution (2-mm voxels) in central gray matter to minimize variations caused by partial-volume effects. HRF amplitudes and temporal dynamics were highly repeatable across four sessions in 20 subjects. Positive and negative HRFs were consistently observed across sessions and subjects. Negative HRFs were generally weaker and, thus, more variable than positive HRFs. Statistical measurements showed that across-session variability is highly correlated to the variability across events within a session; these measurements also indicated a normal distribution of variability across cortex. The overall repeatability of the HRFs over long time scales generally supports the long-term use of event-related functional magnetic resonance imaging protocols.

KEYWORDS

BOLD fMRI, cerebral hemodynamics, longitudinal, multisensory stimulation, temporal dynamics

1 | INTRODUCTION

In functional magnetic resonance imaging (fMRI) experiments using brief (<4-s duration) stimuli, the blood-oxygen-level-dependent (BOLD) signal largely depends on the coupling between oxygen metabolism and blood flow after neural activation (Boynton et al., 1996; Kim & Ress, 2016; Logothetis et al., 2001; Ogawa et al., 1990; Ogawa et al., 1992). The hemodynamic response function (HRF) is the BOLD dynamics generated by a brief stimulus (Boynton et al., 1996). For excitatory stimuli, experiments show that

stereotypical HRFs can be characterized with three temporal phases (Aguirre et al., 1998; Friston et al., 2003; Woolrich et al., 2001) that are broadly evident across human cerebral cortex (Taylor et al., 2018). Upon stimulus onset, a transient (<2 s) initial dip or delay is followed by a hyperoxic peak (3–10 s after stimulus onset), and then an undershoot that lasts ~15 s until it returns to the signal baseline (Boynton et al., 1996; Buxton et al., 1998; Kim et al., 2013; Menon et al., 1995; Thompson et al., 2003; Yacoub & Hu, 1999).

Many BOLD studies assume a heuristic, time-invariant form for the HRF (typically a difference of gamma-variate functions), which is

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

often convolved with a stimulus function to fit the BOLD signal (Boynton et al., 2012; Friston, Fletcher, et al., 1998; Glover, 1999; Handwerker et al., 2004; Li et al., 2019; Rosa et al., 2015). However, several other studies have raised questions about the validity of this assumption due to variability of the HRF. For example, a human study shows that the variability of the HRFs may arise from neural and non-neural mechanisms (Fox & Raichle, 2007). Within a subject, the HRFs may vary in magnitude and timing across trials, sessions, seasons, brain regions, and tasks (Aguirre et al., 1998; Cohen et al., 2002; Gonzalez-Castillo et al., 2015; Handwerker et al., 2004; Meyer et al., 2016; Neumann et al., 2003; Puckett et al., 2014; Taylor et al., 2018; Truccolo et al., 2002). Comparison between HRFs across subjects and brain conditions can show even larger variability (Bonakdarpour et al., 2007; D'Esposito et al., 2003; Elbau et al., 2018; Handwerker et al., 2004). Some of the variability in many early experiments could be attributed to the use of coarse spatial resolution (3–5 mm anisotropic voxels), which creates partial volume effects that mix gray matter, white matter, and superficial vascular signals from draining veins (Kim & Ress, 2017; Turner, 2002). Nevertheless, our recent work using 2-mm voxels also showed significant and substantial spatial variability of HRFs across cortex (Taylor et al., 2018).

Moreover, negative HRFs (negative BOLD responses evoked by a brief stimulus) often shown in fMRI experiments limit the utility of a heuristic form. Negative HRFs can resemble inverted HRFs; however, their dynamics suggest suppression, deactivation, or neurovascular uncoupling (Harel et al., 2002; Pasley et al., 2007; Puckett et al., 2014; Shmuel et al., 2002). Measurements of the negative HRFs show that they have even greater variability and lower signal-to-noise ratio than positive HRFs (Olman et al., 2007). It has also been shown that the mixed HRF signals near the borders of regions of predominantly positive and predominantly negative HRF are highly unstable (Klingner et al., 2011). Moreover, HRFs associated with negative BOLD can take on more complex forms suggesting disparate physiology between excitatory and suppressive cortical activity (DeLaRosa et al., 2021). Altogether, there is much evidence that the HRF exhibits substantial variability across the cortical surface, tasks, and subjects.

Animal experiments also report HRF variability across trials, days, and subjects (Belloy et al., 2020; Hillman et al., 2007) using optical and fMRI methods. Rodents have frequently been used as models to study the HRF, producing stable temporal dynamics in somatosensory cortex with forepaw stimulation (Bailey et al., 2013; de Zwart et al., 2005; Silva et al., 2007), although these HRF dynamics are significantly faster than in humans. While rodents have smaller vascular structures and greater blood flow speed than humans, the BOLD signal in rodents is regionally specific with substantial variability across cortex (Hirano et al., 2011; Keilholz et al., 2004). Furthermore, most of these studies used anesthetization, which strongly affects the hemodynamic response and therefore confounds study of HRF variability (Hillman et al., 2007; Magnuson et al., 2014).

Although HRFs can vary in shape, amplitude, and dynamics within a human subject for a given stimulus paradigm, the variation within a session (across trials) has been shown to be smaller than that between sessions and much smaller than variation across subjects or across

cortical regions, (Aguirre et al., 1998; Dale & Buckner, 1997; Kim et al., 1997; Miezin et al., 2000). It was also found that HRFs were reliable across session events with the time-to-peak (TTP) being the most stable HRF parameter (Neumann et al., 2003). Two parameters of the HRF, onset time (defined as 1 SD above baseline) and poststimulus undershoot amplitude were the most variable among HRF parameters (Miezin et al., 2000). When the HRF was evoked by a brief Stroop task over long periods of functional experiments (120 min), there was little significant temporal variability over these time scales (Menz et al., 2006). However, these experiments examined only small regions of cortex, and generally were performed with coarse spatial resolution that did not resolve the gray matter.

Previously, we characterized the HRF time series with 2-mm-cubic voxels across the majority of human cortex (Taylor et al., 2018). We analyzed several HRF parameters in 20 healthy subjects—peak amplitude, TTP, peak full-width at half-maximum (FWHM), undershoot amplitude, and time-to-undershoot (TTU). Sensory, associative, and executive regions, driven by our multimodal stimulus and task, showed strong activation amplitudes and stable temporal dynamics. Despite the variability of HRF amplitudes across these activated regions, spatial variations of those parameters had consistent patterns across subjects. Long-term stability of the local HRF would simplify linear analysis of fMRI experiments. Moreover, if these HRF parameters are consistent over long time scales, they could serve as biomarkers to enable comparison of neurovascular and neurometabolic function between healthy and pathological populations. In this study, we examine the long-term reliability of these HRFs.

Many other experiments have evaluated the temporal reliability of the BOLD response (Grady & Garrett, 2014; Leontiev & Buxton, 2007; Miller et al., 2002; Tjandra et al., 2005). Most of these utilized sustained stimuli that evoke strong BOLD responses over portions of cortex (Gonzalez-Castillo et al., 2015) but may also generate remote vascular responses that can unpredictably modulate local cerebral metabolism and blood flow (Drew et al., 2011; Hirano et al., 2011; Iadecola et al., 1997). These studies then assume a linear model to estimate the neural response and its reliability. Experiments making use of very long scanning durations requiring multiple sessions have indeed confirmed the variability and complexity of BOLD responses to sustained stimuli (Bandettini & Cox, 2000). This linear deconvolution approach has two main drawbacks.

First, the assumption of linearity is questionable; many nonlinearities have been characterized for the BOLD response (Boynton et al., 1996; Vazquez & Noll, 1998). The clearest nonlinearity in BOLD response corresponds to the stimulus duration, which saturates for stimuli greater than 2 s (Friston, Josephs, et al., 1998; Huettel et al., 2004; Miller, Luh, et al., 2001). This motivated our use of a 2-s-duration stimulus to evoke a localized neural response and HRF. Such a brief stimulus is beneficial for several reasons. We previously showed that a brief stimulus uniformly excites the majority of cortex with spatially variable but temporally stable dynamics (Taylor et al., 2018). Also, because neuronal activity closely follows the temporal window of the stimulus, the use of a brief stimulus separates the fast timescales of the neural activity (~milliseconds) from the much

slower evoked vascular and metabolic responses (~seconds). Furthermore, the underlying physiology corresponding to the BOLD dynamics evoked by a short stimulus is simpler than for longer stimuli. In general, BOLD dynamics are believed to involve a combination of cerebral blood flow, oxygen metabolism, and blood volume (Buxton et al., 1998; Buxton et al., 2004; Davis et al., 1998; Kim & Ress, 2016). However, recent experiments have not observed significant venous blood volume effects (Drew et al., 2011; Gagnon et al., 2015; Hillman et al., 2007; Lindvere et al., 2013), particularly for the short stimuli used in event-related fMRI experiments, including those that evoke the HRF. Thus, HRF dynamics associated with a brief neural activity are dominated by neurovascular and neurometabolic coupling without venous volume effects.

Second, the linear deconvolution approach finds a linear filter that optimizes some simple reliability metric. This has the drawback that the fits vary across voxels and subjects, with noisier or weaker responses generally corresponding to a more strongly low-pass filter, so that temporal resolution becomes variable and confounds characterization of spatial and subject-by-subject variability. To avoid these issues, we used a simple time-locked averaging approach to characterize the HRF without the use of deconvolution. Also, we use a model-free parameterization scheme. Our fMRI measurements use a sampling interval (TR) of 1.25 s (Nyquist frequency of 0.4 Hz) that is sufficient to resolve BOLD signal variations driven by the slow hemodynamic and metabolic responses. Although time-locked approaches have been noted to have low estimation efficiency when noise is spectrally white (Liu, 2004), the strongest source of nuisance in BOLD measurements are pulse and respiration (Glover et al., 2000; Krüger & Glover, 2001; Triantafyllou et al., 2005; Wald &

Polimeni, 2017). Because these nuisance components are fast relative to our sampling frequency, they are effectively reduced by averaging. In fact, we have obtained excellent results in terms of contrast-to-noise ratio (CNR) using a time-locked averaging and model-free design in previous work (Taylor et al., 2018), with similar data quality reported here. Our results are generally relevant to all short-stimulus experiment designs, for example, event-related fMRI experiments.

Using these approaches, our present goal is to evaluate the long-term temporal reliability of the HRF across the majority of cerebral cortex. Specifically, the analysis of the HRF data was motivated by several questions (Figure 1). First, how stable are the HRF time series? Second, how stable are the HRF parameters? Third, what are the statistical distributions of parameter variability? These questions were addressed separately for positive and negative HRFs using several noise and variation metrics that are summarized in Figure 1. We examined BOLD HRF stability across four sessions obtained at intervals of 3 h, 3 days, and 3 months. High-resolution HRF data from each session were compared directly on a voxel-wise basis within the subject to quantify variability across sessions. Nonparametric statistical methods were then used to estimate and compare the distributions of variability both within and across sessions. Over long time scales within cortical gray matter, we observed remarkable temporal stability of HRF amplitudes and their spatial patterns, as well as HRF dynamics.

2 | METHODS

Twenty healthy subjects from two age groups (20–35 years ($N = 10$), and 50–65 years ($N = 10$)) participated in the study; half of each

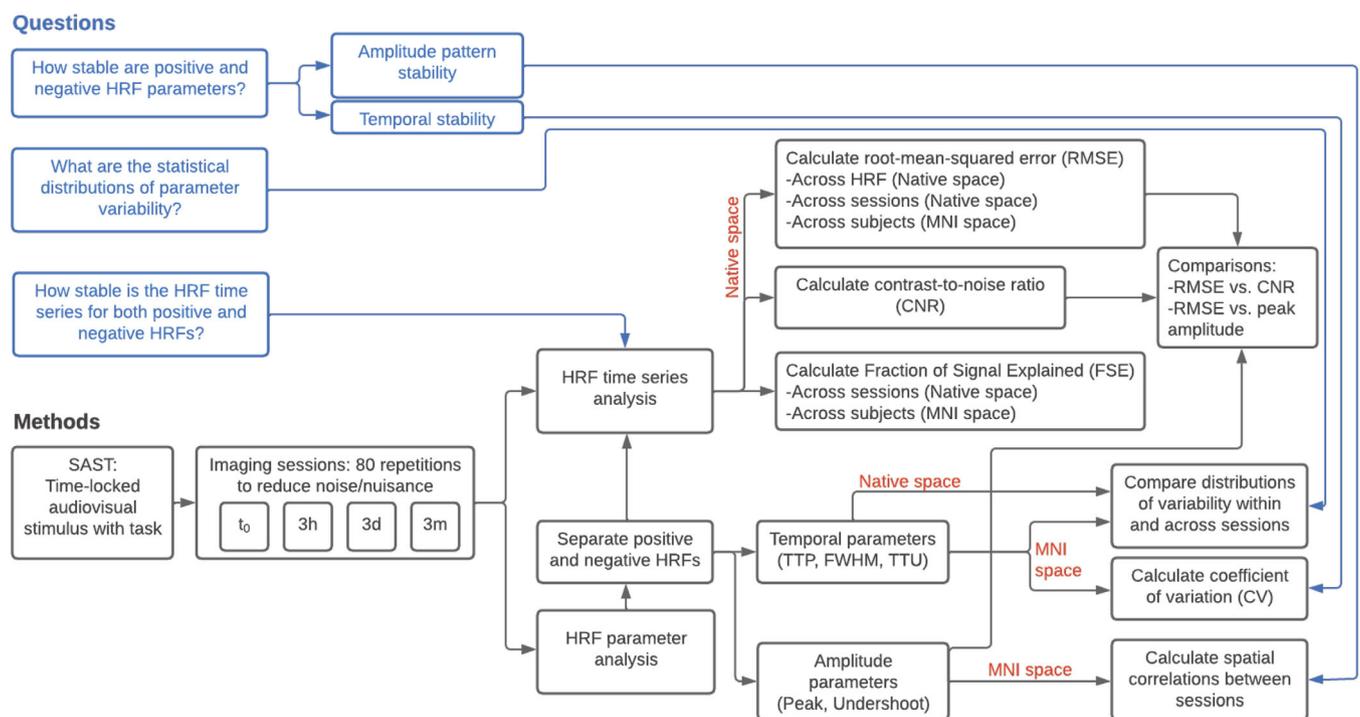


FIGURE 1 Concept map summarizing analysis methods and the questions motivating their application

group were female. Subject ages were balanced into two groups to enable later studies of age-related HRF differences. Subjects were recruited from the local community after providing informed consent according to a protocol approved by the local ethics committee for human study, Baylor College of Medicine Institutional Review Board. Our human-subjects experimentation protocol conforms to Baylor College of Medicine's "Ethical and Regulatory Mandate for Protecting Human Subjects," which places emphasis on the principles in the Belmont Report. On their first visit, each subject was trained on the task until their performance stabilized, then performed their first fMRI session; this became their $t_0 - 3$ days session (3d). They returned 3 days later for two sessions separated by three hours (3h). Although the 3d time-point was acquired first, the second scanning session is chosen as the reference session, t_0 . A fourth scanning session was obtained 3 months (3m) after the reference session (Figure 2(a)). The four session time-points are hereafter referred to as t_0 (reference session), 3d, 3h, and 3m.

Imaging was performed at 3 Tesla on a MAGNETOM Trio (Siemens Healthcare, Erlangen, Germany) scanner using the product 32-channel head coil. fMRI data were collected using an SMS-accelerated echo-planar imaging sequence (Breuer et al., 2005; Setsompop et al., 2012), with acquisition parameters: TR = 1.25 s, TE = 30 ms, GRAPPA factor = 2, SMS factor = 3, 2-mm pixels with 2-mm slice thickness, and 57 slices. During functional runs, stimulus timing was programmed to align with the TR of the scanner (24 TR periods per trial).

During each fMRI scan, HRFs were evoked by a 2-s speeded audiovisual sequence-following task (SAST) followed by a 28-s nondemanding fixation task. Stimulus onset was cued by a change of fixation dot color for 0.5 s before the 2-s stimulation period (Figure 2(b)). The stimulus had three components: visual, audio, and task. Visual stimulation consisted of three consecutive presentations of flickering (6 Hz) colored dots, half brightly colored and half darkly colored to enhance contrast. The dots were presented in one of three circular (5° radius) regions for 667 ms. The regions were uniformly distributed horizontally across the width of the display with each position having a specific color: yellow on the left, green in the center, and red on the right. Vertical positions were randomly varied over a $\pm 3^\circ$ range. The spatial order of presentation was random without sequential repetition. Each dot-region display was accompanied by audio stimulus of filtered white noise: medium pitch during yellow dots; low pitch for red; and high pitch for green. Subjects were instructed to follow the colored dot regions with eye movements and quickly (within the 667-ms period) press a button that matched the position, color, and sound presented. Reaction time and response accuracy were tracked during each run. Between these strong audiovisual stimuli, subjects performed a nondemanding fixation task to maintain attention and fixation. During the fixation period, subjects attended a central colored dot (0.15° diameter) that changed color every 0.6 s. Subjects pressed a button at the appearance of a single target color, which appeared on average every 6 s throughout the scan (exponential distribution, truncated at a minimum of 1 s) (Taylor et al., 2018). A stimulus and a fixation period constituted a 30-s HRF measurement. Then, 16 HRFs were collected in each of 5 runs to yield 80 HRF

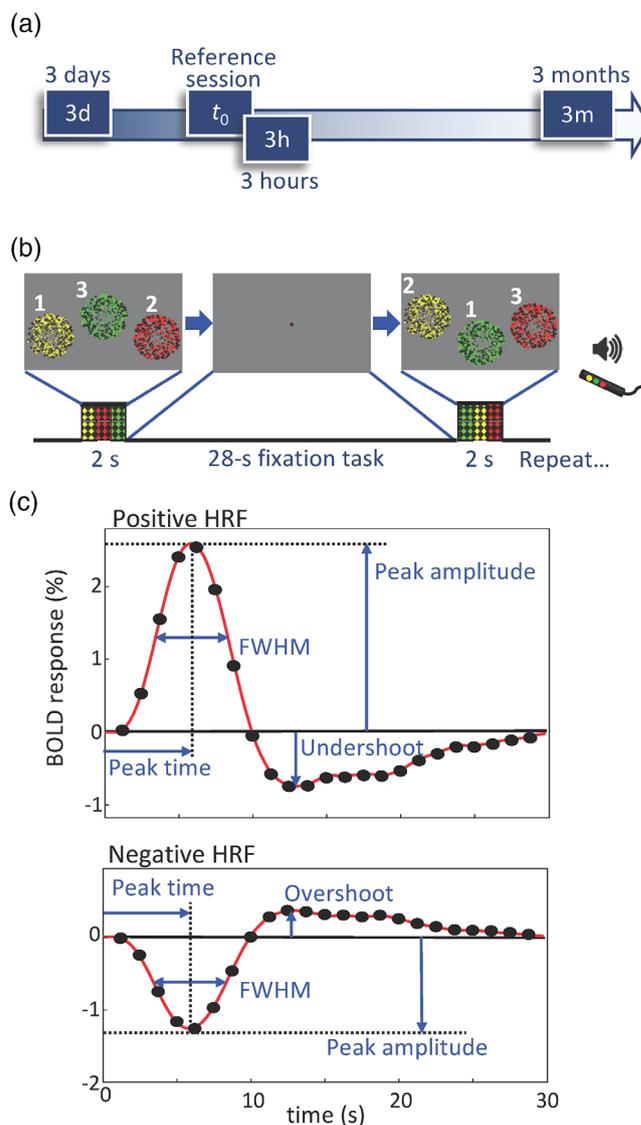


FIGURE 2 (a) Four sessions were obtained for each subject: The first was the 3-day session, then reference session 3 days later (t_0), then a third session at $t_0 + 3$ hour, finally a fourth session at $t_0 + 3$ months. (b) Schematic shows the speeded audiovisual sequence-following task (SAST) paradigm. Three regions consisting of flickering colored dots appear consecutively for 667 ms each to create a 2-s visual stimulus accompanied by auditory stimuli with frequencies corresponding to the position of the dot regions, while subjects performed a sequence-following task. The sequence order is randomized. A single hemodynamic response function (HRF) event spans the 2-s stimulus task and 28-s resting period fixation task. (c) HRF parameters are extracted from positive and negative blood-oxygen-level-dependent (BOLD) responses including the HRF peak amplitude, peak time, full-width at half-maximum (FWHM), and under(over)-shoot amplitude

measurements for each session. At the end of each session, a T1-weighted volume (3D FLASH sequence, minimum TE and TR, 15° flip angle) was collected on the functional slice prescription to facilitate alignment to a high-resolution reference anatomy (described below).

For each subject, a MP-RAGE T1-weighted structural anatomy was also collected (TR = 2300 ms, TI = 900 ms, flip angle 9°, 1-mm cubic voxels) and segmented using FreeSurfer (Dale et al., 1999) to distinguish gray and white matter. These segmented volumes were subsequently aligned to the functional data using the T1-weighted volume obtained in each session to enable further analysis.

Functional images produced time-series data with a total of 1920 temporal samples per session (80 HRFs \times 30-s/HRF \times 1TR/1.25-s). After motion correction and registration to high-resolution anatomy, the time series were depth-averaged across the central layers of gray matter (0.2–0.8 normalized depths) to reduce partial-volume effects (Salminen et al., 2016), and mapped onto the gray-white surface vertices (Khan et al., 2011; Kim & Ress, 2017). Note that this approach maximizes use of the native MRI spatial resolution to select gray matter voxels, avoiding lower reliability samples from adjacent white matter and superficial vascular regions superficial to the gray matter (Kim & Ress, 2017). Thus, for each vertex on this surface, the 80 HRFs were averaged to create a single HRF time series (black dots in Figure 2(c)). The surface data were spatially smoothed along the surface using manifold-distance coordinates (8-mm-FWHM Gaussian kernel). Each vertex's time series was then upsampled to 0.1-s sampling with Hermite-spline interpolation and characterized by four HRF parameters (Figure 2(c)): peak amplitude, TTP, FWHM, and undershoot (in positive HRFs) or overshoot (in negative HRFs). Peak amplitude was required to be in the range of 2–14 s to minimize the effects of occasional outliers from noisy events. The HRFs were separated into positive and negative HRFs based on the sign of the peak amplitude parameter. Positive and negative HRFs were analyzed separately.

For across-subject comparisons, the smoothed time series of all the subjects were resampled and registered to the MNI-152 brain surface using FreeSurfer surface-to-surface transformations (Collins et al., 1994; Fischl et al., 1999; Mazziotta et al., 2001). To properly reflect the smoothing in statistical comparisons, each surface was correspondingly downsampled using MATLAB's *reducepatch* function so that 1/64th (corresponding to the 8-mm-FWHM smoothing) of the vertex data was sampled. The HRF parameters for these downsampled data were used to measure the correlations, coefficients of variation, and variability across session time-points for each subject. This yielded 1989 ± 232 vertex correlations per subject that were spatially averaged separately for positive and negative HRFs.

First, we examined the stability of the HRF time series (Figure 1). To do this, we calculated the root-mean-squared error (RMSE) values across cortical surface vertices. Specifically, we took the RMSE of the mean HRF time series in each vertex between the reference session t_0 and the other session time-point (3h, 3d, 3m) measurements, then we took the average across the three sessions. To quantify the temporal variations in variability during the HRF, RMSE between sessions was also calculated at each timeframe of the HRF and then averaged across vertices. Distributions of RMSE values were calculated by bootstrapping (discussed below) the mean RMSE values across subjects. This was performed for positive and negative HRFs separately. RMSE calculations provided a raw estimate of the HRF variation

between sessions, as well as a metric for how variability temporally evolved over the course of each HRF. However, this metric is sensitive to absolute amplitude shifts between sessions. Because of this amplitude dependence, we compared RMSE values to a related metric, CNR, of the time series. We define CNR as the ratio of the absolute value of greatest amplitude to its standard-error-of-the-mean (SEM) across the 80 events. This metric also varies with signal amplitude, so we expected similar results between RMSE and CNR. With this metric, $CNR > 3$ represents data with significant activation at $p < 0.002$. Finally, to provide a more normalized metric of variability, fraction-of-signal-explained (FSE) was used to compare HRF time series at each vertex between t_0 and the other time-point intervals. FSE was calculated by taking the differences between each pair of time series and then dividing by the reference session (t_0) time series values (or the square root of one minus the fraction of unexplained variance). The FSE metric characterizes linear variability that ignores scaling and offset differences, while RMS is a more stringent metric that includes all forms of variability.

Next, we tested the longitudinal stability of the HRF parameters. In our previous work, temporal HRF parameters did not display apparent patterns across cortex but were relatively homogeneous compared to amplitude patterns (Taylor et al., 2018). Because of this, we chose to measure the absolute variability of these parameters using coefficient of variation (CV). CV was used to compare HRF temporal parameters at each vertex on the surface across the four time-points. CVs, also known as relative standard deviations (Bland & Altman, 1996; Snedecor & Cochran, 1972), were calculated by measuring the absolute value of the percent standard deviation divided by the mean. CVs were calculated separately for positive and negative HRFs from the downsampled set of surface vertices. SEM values of the CVs across cortex were also obtained to demonstrate the distribution of CVs. To obtain across-subject CVs, temporal parameters from all sessions were resampled onto the MNI surface. For each session time-point, the bootstrapped means (discussed below) of temporal data were averaged across subjects. Then, CV was calculated across session time-points for each parameter.

Variability was estimated using bootstrapping (resampling with replacement) to determine the mean distributions of each HRF parameter for each vertex (Efron, 1981). Within-subject variability for each session was estimated by bootstrapping across the 80 HRF events ($N_{boot} = 500$) to obtain means and 68% confidence intervals. Variability was defined as half the difference between the confidence intervals, which is equal to the SEM in normally distributed data. To obtain time-series variability, this bootstrapping was performed for each time sample in the HRF. Variability for each HRF parameter was similarly estimated by bootstrapping across the 80 HRF events.

Across-subject variability was estimated by resampling and registering each subject's HRF parameters onto the MNI surface. Parameters from each of the 20 subjects were resampled 500 times; then the bootstrapped mean values of each parameter were used to create surface overlays on the MNI-152 surface. Variability across subjects was again calculated as half the difference between the 68% confidence intervals obtained from the bootstrapped data.

To further examine the character of the temporal variability, we compared variability between sessions to the variability across events within the t_0 session. Comparisons were calculated across the HRF time series and HRF temporal parameters (TTP and FWHM) for vertices of a downsampled cortical mesh in each subject. Within-session variability was calculated at each vertex by bootstrapping the parameter values measured from all events and then subtracting the mean parameter value, providing a distribution of δ values. We applied Gaussian fits to measure the variability, σ , in each distribution. Similarly, we fit Gaussians to the distributions of differences of bootstrapped parameters between the largest interval (3-month) time-point and the initial session, t_0 , to estimate across-session variability.

3 | RESULTS

3.1 | Task performance

The subjects performed well in the main stimulus task with $92 \pm 8\%$ valid responses with average reaction times of 0.39 ± 0.10 s. Subjects performed the fixation task with $46 \pm 11\%$ valid responses with average reaction times of 0.41 ± 0.05 s. As a respite period between main stimulus periods, subjects were generally less attentive during baseline fixation. As an expansion of our previous work from one to four sessions per subject (Taylor et al., 2018), we found that the task generated strong activations ($\text{CNR} > 3$) in $70 \pm 9.1\%$ of cortex across all subjects and sessions.

3.2 | Temporal stability of the HRF

After parameterization of each averaged HRF from cortical surface, we used the peak amplitude to separate positive and negative HRFs for each session; positive HRFs covered $74 \pm 11\%$ of strongly responsive cortex ($\text{CNR} > 3$) across all sessions; the remainder ($26 \pm 11\%$) responded with negative HRFs. Reliability (p value) of the peak amplitude classification of positive and negative HRFs is depicted in the Supporting Information (Figure S1). Both positive and negative HRF amplitudes were very reliable except in areas of low CNR and at the boundaries of positive and negative HRF regions. Then, we compared the HRF time series across session time-points. Mean HRF examples from each session time-point are shown (Figure 3) for 6-mm-diameter regions of positive and negative peak amplitudes. Peak amplitude parameter maps and corresponding time series across session time-points are shown for two representative subjects (Subjects 9 and 17) from younger and older age groups, respectively. The HRF time series for both positive and negative HRFs are quite similar across all session time-points.

RMSE calculations of the HRFs between session time-points quantify the absolute HRF variation for positive and negative HRFs (Figure 4). First, RMSE values were calculated across the whole HRF for each vertex on the cortical surface across session time-points.

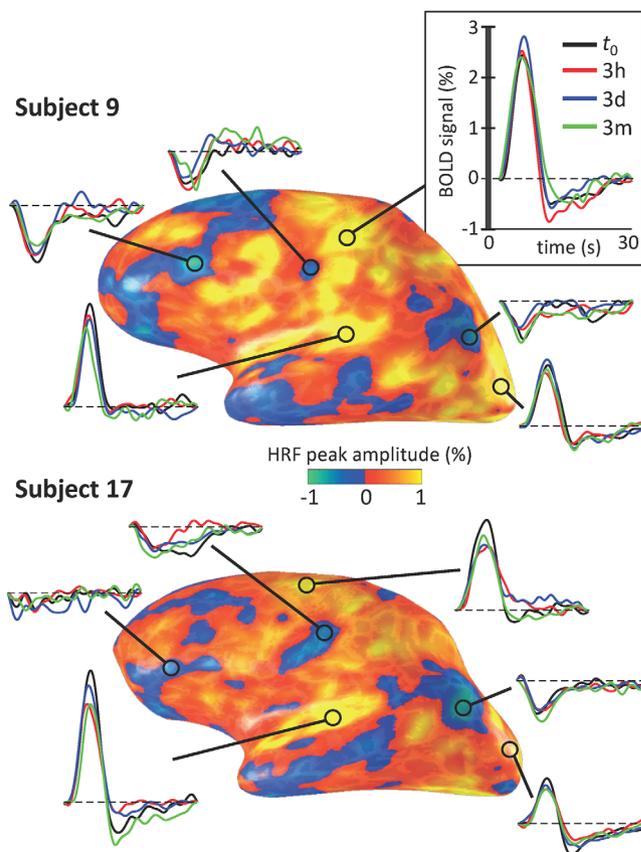


FIGURE 3 Overlays of hemodynamic response function (HRF) amplitudes, red-yellow for positive HRFs and blue-green for negative HRFs, after 8-mm-full-width at half-maximum (FWHM) smoothing, on cortical surfaces of two subjects. Circles mark sample time series from 6-mm-diameter gray-matter disks showing variations across session time-points (t_0 , 3 h, 3 days, and 3 months)

These showed patterns of higher variation in a few cortical regions, such as superior temporal lobe and frontal lobes (Figure 4(a)). Average RMSE values and standard deviations for each subject are compared in bar plots (Figure 4(b)). Subjects are arranged in the bar plots by ascending age. Bootstrapped confidence intervals showed no significant differences ($p < .05$) between the RMSE values of positive and negative HRFs across subjects. No significant differences were found between the two age-group averages (shown as Young and Older in the Native space means in the bar plots). Distributions of the RMSE values are shown for positive and negative HRFs across subjects (Figure 4(c)). However, when evaluated across cortex (MNI space), the RMSE of positive and negative HRFs were significantly different ($p < .01$). RMSE was 0.201 ± 0.166 for positive HRFs and 0.214 ± 0.174 for negative HRFs. RMSE was also measured during the HRF to evaluate variations in temporal dynamics (Figure 3(d)). Averaged across subjects, both positive and negative HRFs had a baseline RMSE around 0.15 throughout the HRF. Moreover, during the time corresponding to the HRF peak (3–10 s), the RMSE increased, suggesting a stimulus-driven effect. The variation of the RMSE over vertices (shown as SEM in the line plot) suggests this is true across cortex. Similar behavior was observed in all subjects (Figure S2(a)). To further

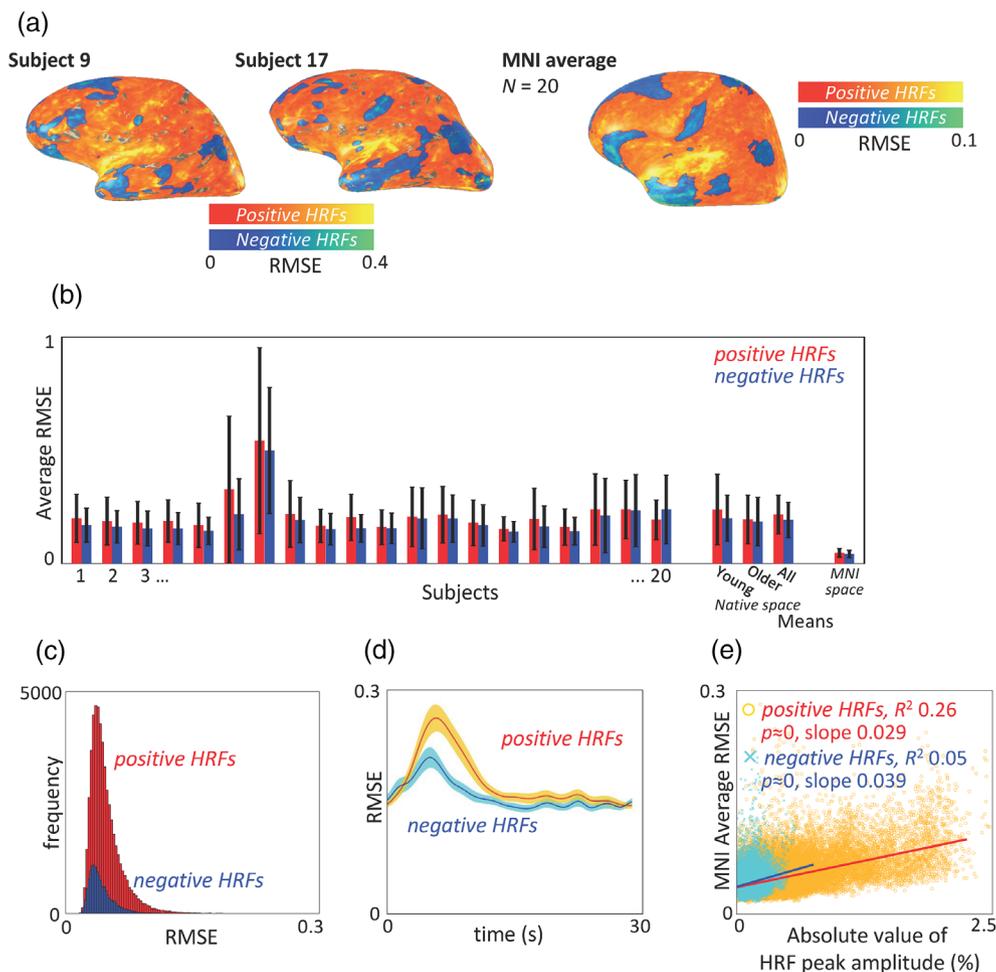


FIGURE 4 (a) Root-mean-squared error (RMSE) averaged across the whole hemodynamic response function (HRF) and the four session time-points with respect to t_0 for each vertex on the cortical surfaces of Subjects 9 and 17, and the average across 20 subjects in the MNI-152 standard space. (b) Average (\pm standard deviation) RMSE across cortex of 20 subjects for positive (red) and negative (blue) HRFs. The average (\pm standard deviation) across subject values is shown as the *Native space* mean. Subjects are ordered by age from youngest to oldest and then averaged by groups (Young, Older, and All subjects). There is no significant difference in spatial mean RMSE between the two age groups. The right-most bar shows the four-session RMSE average after resampling all 20 subjects to the MNI-152 standard space. (c) Distributions of RMSE values in MNI-152 standard space separated for positive and negative HRFs. (d) RMSE was averaged across vertices for each time frame of the HRF time series for all 20 subjects. The average RMSE (\pm SEM across subjects) is shown for positive HRFs (red/yellow) and negative HRFs (blue/cyan). (e) Average RMSE across subjects in MNI-152 standard space measured at peak time compared to the absolute value of the peak amplitude for each vertex along the cortical surface shows positive trendlines for both positive HRFs (yellow markers/red line) and negative HRFs (cyan markers, blue line). Correlations and p -values are shown for each comparison

explore the stimulus-driven effect, we measured the correlation between RMSE and absolute value of the peak amplitude on the MNI-152 average surface (Figure 4(e)). RMSE values of both positive and negative HRFs show slight positive trends with amplitude. These data indicate that there are two sources of session-to-session variability: a baseline, time-independent component, and an early, stimulus-evoked component. We also measured the correlation between RMSE and absolute value of all signal amplitudes during the HRF and found a similar trend (Figure S2(b)).

The CNR similarly varies with peak amplitude. CNR shows patterns of strongly activated regions of cortex (Figure 5(a)) in the reference session. Bar plots show the average (\pm standard deviation) CNR for positive (red) and negative (blue) HRFs across cortex of 20 subjects, *Native space* mean, and the MNI-152 standard space mean

(Figure 5(b)). CNR varies over the course of the HRF which has maximum value and variation around the peak, shown for the average across subjects (Figure 5(c)). Similar results were observed in all subjects (Figure S3). A comparison of RMSE and CNR shows two regimes of correlation in both positive and negative HRFs (Figure 5(d)). To capture this behavior, we regressed MNI average data with piecewise linear models and adjusted the CNR breakpoint between them to maximize variance explained by all the HRFs. Choosing $\text{CNR} = 5.0$, we then observe $R^2 = .265$ for positive HRFs, and $R^2 = .201$ for negative HRFs. Thus, for low CNR, <5.0 , RMSE decreases with CNR for both positive HRFs (slope = -0.016) and negative HRFs (slope = -0.013). For high CNR, >5.0 , RMSE increases with CNR for positive HRFs (slope = 0.004) and negative HRFs (slope = 0.012). All four correlations have negligible p value.

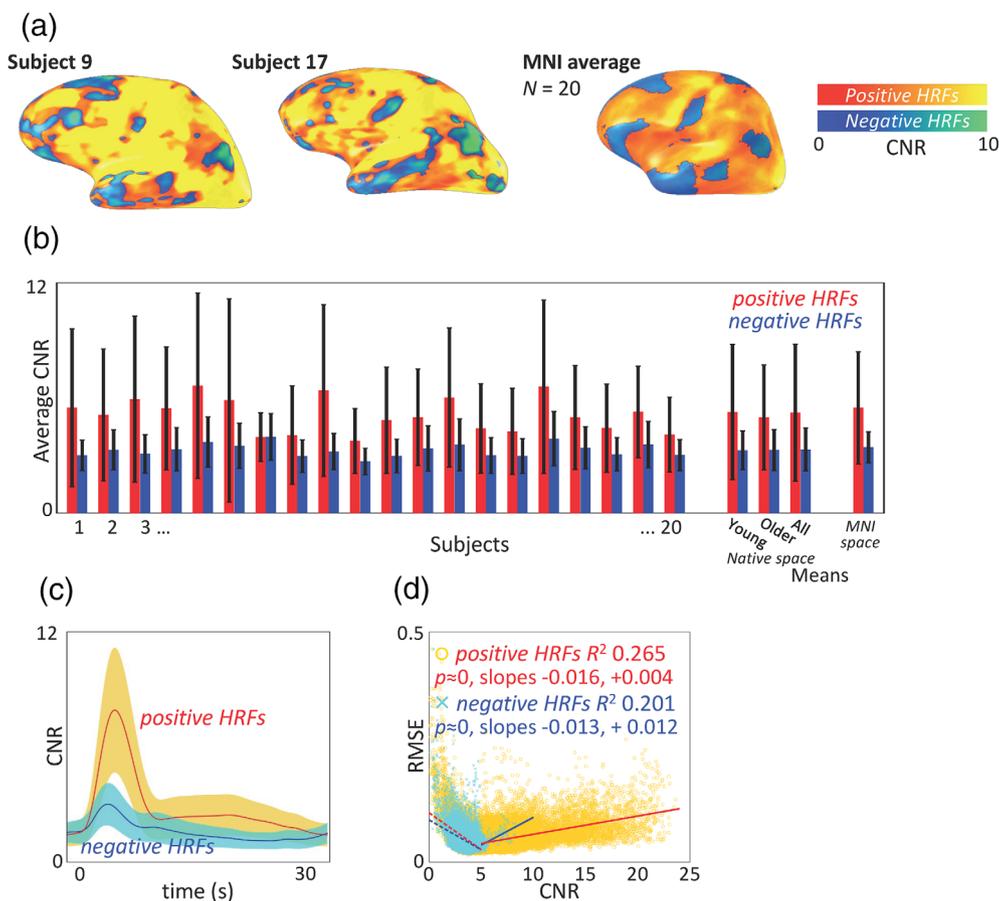


FIGURE 5 (a) Contrast-to-noise ratio (CNR) at t_0 for each vertex on the cortical surfaces of Subjects 9 and 17, and the average across 20 subjects in the MNI-152 standard space. (b) Average (\pm standard deviation) CNR across cortex of 20 subjects for positive (red) and negative (blue) hemodynamic response functions (HRFs). The averages (\pm standard deviation) across subject values and age groups are shown as the *Native space* means. The right-most bar shows the CNR average after resampling all 20 subjects to the MNI-152 standard space. There is no significant difference in spatial mean CNR between the two age groups. (c) The CNR was averaged across vertices for each time frame of the HRF time series for all 20 subjects. The average CNR (\pm standard deviation) across subjects is shown for positive HRFs (red/yellow) and negative HRFs (blue/cyan). (d) Correlations between average RMSE and CNR across vertices in MNI space are bilinear for both positive (red/yellow) and negative (blue/cyan) HRFs. Bilinear fits are modeled for CNR < 5.0 (dashed lines) and CNR > 5.0 (solid lines) for each population

As another metric for the similarity of the HRFs between session time-points, we calculated the FSE relative to the HRFs of the reference time-point, t_0 . Most of the signal was explained by the HRF obtained at the reference time-point for each vertex on the cortical surface for the 3-h, 3-day, and 3-month time-points (Figure 6). The general patterns of variability are qualitatively similar for the FSE and RMSE metrics. In the surface maps, we observe that FSE tends to be lower near the transitions between positive and negative HRFs. FSE values, spatially averaged across the downsampled mesh in individual subjects, are compared in bar plots. Mean FSE values were calculated across subject means (denoted *Native space* in Figure 6) and again after registering onto the MNI-152 surface (denoted *MNI space*). A large fraction the signal in each session time-point was explained by the initial session for both positive and negative HRFs. In the *MNI* average the signal explained at 3h, 3d, and 3m time-points by the t_0 HRFs were $64 \pm 15\%$, $65 \pm 15\%$, and $63 \pm 15\%$, respectively, for positive HRFs and $49 \pm 15\%$, $50 \pm 14\%$, and $48 \pm 15\%$ for negative HRFs. Negative HRFs showed significantly less explained signal than positive

HRFs across sessions, $p < .01$. Mean FSE values were slightly higher in the participants' native spaces than after transformation to MNI space.

3.3 | Spatial HRF amplitude patterns

Overlays of HRF peak and undershoot amplitudes show similar cortical patterns across sessions for the two representative subjects (Figure 7(a)). While both positive and negative HRFs are strongly evoked by the multisensory stimulus paradigm, negative HRFs are generally weaker and more variable across session time-points. Similar results were observed in all individual subjects, and in the average across subjects. Consistent with our previous work (Taylor et al., 2018), undershoot amplitude patterns (or overshoot amplitudes for negative HRFs) were less stable over time than peak amplitude patterns. Furthermore, under/overshoots were only consistently observed in a subset (78.2%) of vertices. For each time-point (3h, 3d,

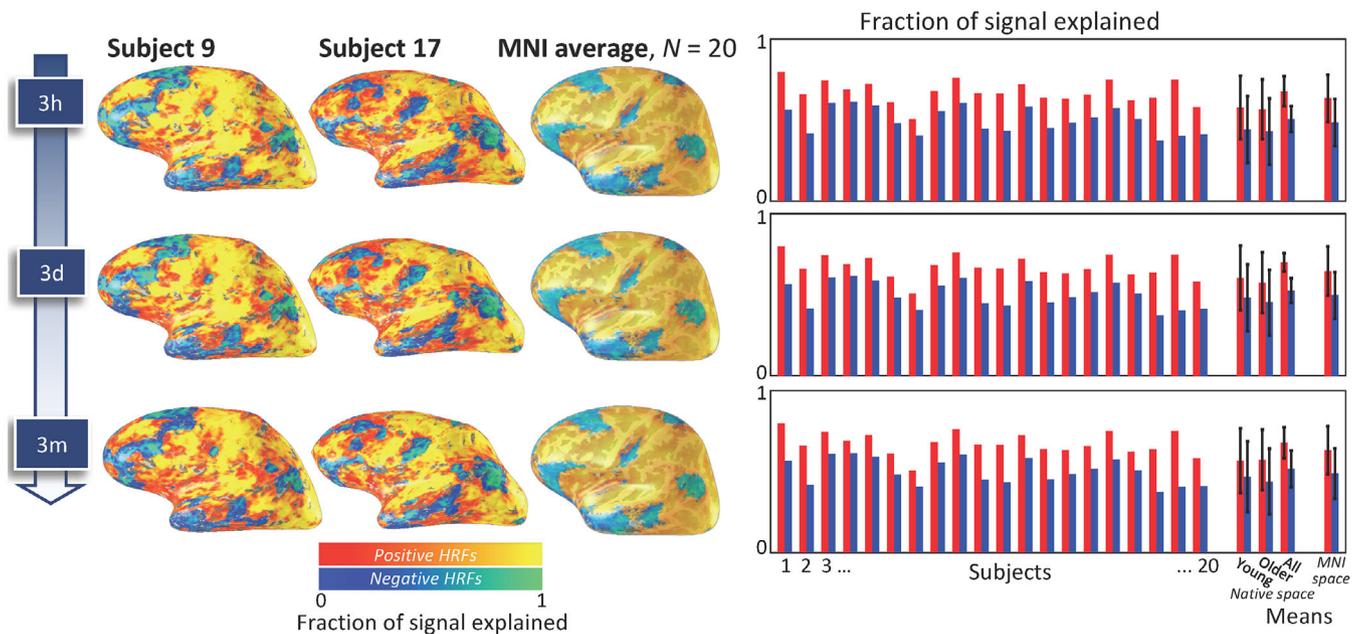


FIGURE 6 For each session time-point (3 h, 3 days, and 3 months), the hemodynamic response function (HRF) of the reference session time (t_0) explains a large fraction of the signal. Fraction-of-signal-explained (FSE) is depicted as overlays on inflated cortical surfaces for two example subjects and averaged across subjects ($N = 20$) in the MNI-152 brain. Color maps are used to distinguish correlations for positive (red-yellow scale) and negative (blue-green scale) HRFs. Spatial mean FSE across the decimated cortex (\pm standard deviation) is shown in bar plots for each subject, separated into positive (yellow) and negative (blue) HRFs. Across-subject means are also shown for both native and MNI spaces. There is no significant difference in spatial mean FSE between the two age groups

3m), we calculated the spatial correlation (Pearson's R) between parameters to t_0 on the subjects' downsampled surfaces. Mean correlations (\pm standard deviations) across time-points were calculated for each subject (Figure 7(b)). The mean peak amplitude correlations across subjects on their native surfaces were 0.86 ± 0.07 for positive HRFs and 0.45 ± 0.16 for negative HRFs. Undershoot amplitude correlations (for those vertices that showed clear under/overshoots) were slightly weaker than peak amplitudes, averaging 0.74 ± 0.07 for positive HRFs and 0.35 ± 0.09 for the overshoots associated with negative HRFs. When resampled onto the MNI surface and then correlated, the correlation values increased for both peak amplitude and undershoot/overshoot. Thus, the patterns of activation were remarkably consistent across the session time-points, particularly for the positive HRFs.

3.4 | Stability of HRF dynamics

Consistent with our previous work, the temporal HRF parameters did not display apparent patterns across cortex but were relatively homogeneous compared to amplitude patterns. Parameters were obtained separately for positive and negative HRFs, with mean values obtained by averaging across both native and MNI surfaces. TTP was 6.2 ± 0.8 s for positive HRFs, 6.6 ± 0.8 s for negative HRFs. FWHM was 3.9 ± 0.7 s for positive, 3.7 ± 0.8 s for negative HRFs. TTU was 13.6 ± 1.8 s for positive, 14.2 ± 2.0 s for negative HRFs. Because temporal parameters are quite homogeneous across cortex, we quantify their

variability using CV. It is worth noting that this metric is opposite to the correlational metric that we used for amplitude. We calculated CV across the four session time-points for each vertex on the surface. This process was repeated for all subjects in Figure 8. Figure 8(a) shows the variability of TTP in three forms: surface overlays, whisker plots, and histograms.

First, the surface overlays of CV for TTP are shown for two subjects, S1 and S2, and the average across subjects in MNI space (Figure 8(a)). The CVs were generally small, $<30\%$ across cortex. CVs were also compared across subjects in MNI space. Because smoothed data were averaged across subjects, variability was much smaller, $CV < 10\%$ across the MNI surface, which was also true for the other parameters.

Next, whisker plots, separated for positive (red) and negative (blue) HRFs, show the median CV and distributions across cortex for each subject. CVs were smaller in positive HRFs than in negative HRFs. CV distributions from the combined subject data are shown in the plots labeled *Native space*. For TTP, the mean CV values were $14.1 \pm 2.7\%$ for positive HRFs and $18.8 \pm 3.6\%$ for negative HRFs. Finally, the CVs from the MNI surface average across subjects are shown in the plots labeled *MNI space*.

Histograms show the distributions of CVs for positive (red) and negative (blue) HRFs compiled from all subjects across cortex. The 68th percentile interval of the CV distributions ranged 6.4–21.2% for positive HRFs, which corresponds to <1.3 -s dispersion of TTP between session time-points, and 9.0–27.0% for negative HRFs, corresponding to <1.8 -s dispersion.

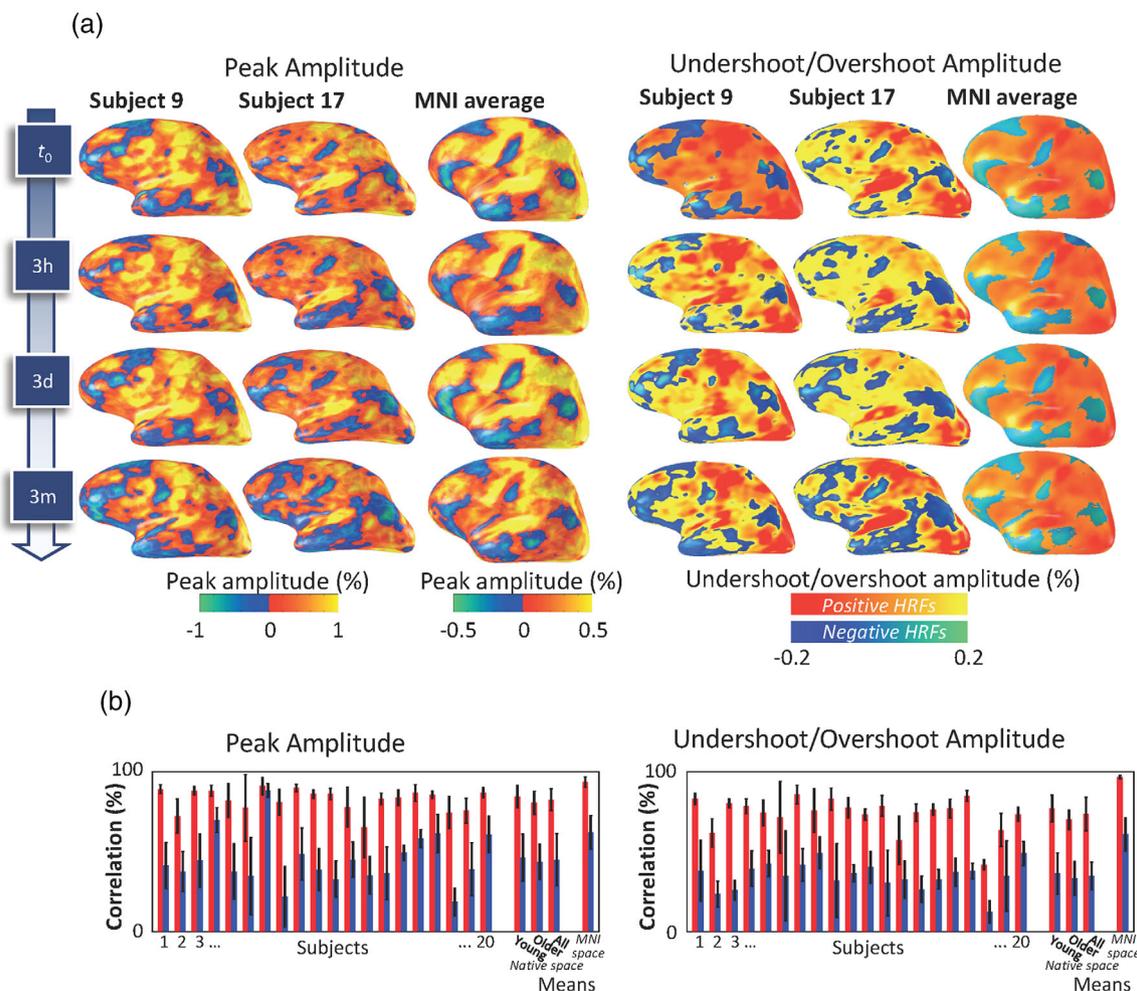


FIGURE 7 (a) Spatial patterns of hemodynamic response function (HRF) amplitudes are overlaid for two example subjects on their native surfaces and the mean across subjects on the MNI-152 surface for each longitudinal session time-point. Left overlays show peak amplitude; right overlays show undershoot/overshoot amplitudes. The spatial patterns are very similar in both native and MNI spaces. (b) Correlations for peak amplitude (left) and undershoot/overshoot amplitude (right) between the spatial patterns at t_0 and 3 h, 3 days, and 3 months were averaged and are shown in bar plots separated by positive HRF (red) and negative (blue) bars (\pm standard deviation across session time-points). Cross-subject means are also shown in native and MNI spaces

Similarly, CVs were measured for FWHM and TTU. FWHM showed the greatest fractional variation among HRF dynamical parameters (Figure 8(b)). The *Native space* mean CV for FWHM was $18.0 \pm 4.6\%$ across subjects for positive HRFs and $23.8 \pm 4.7\%$ for negative HRFs. Histograms of the distribution of FWHM CV values ranged 8.1–27.3% at the 68th percentile, corresponding to <1.1 s of dispersion in positive HRFs. Negative HRFs varied more as CVs ranged 10.9–35.6% (1.3-s dispersion).

Of the temporal parameters, TTU (Figure 8(c)) showed the lowest CV on average: $11.4 \pm 2.6\%$ for positive HRFs and $14.2 \pm 3.0\%$ for negative. However, when returned to units of seconds, TTU dispersion surpasses the other parameters with across-subject averages of 1.6 s (positive) and 2.0 s (negative). Positive TTU CV distributions ranged from 5.5 to 16.4% (<2.2 -s dispersion) in the 68th percentile, while negative TTU CV distributions ranged from 6.5 to 19.2% (<2.7 -s dispersion).

3.5 | Comparison of within-session and across-session variability

Within-session variability distributions, *delta* values, across vertices were modeled well by a single Gaussian, $R > 0.94$ for all subjects (Figure 9(a)). Distributions were less normal in character for the temporal parameters, $R > 0.69$ for TTP and $R > 0.88$ for FWHM, but the Gaussian fits still characterized the distributions reasonably well. Deviations from normality were largely associated with greater kurtosis, without substantial skew (Figure S4). Fits of the across-session distributions to a single Gaussian had R values greater than 0.92 for the time series, 0.73 for TTP, and 0.91 for FWHM across subjects (Figure 9(b)).

For single Gaussian fits to the time series noise distributions, within-session variability was very slightly more normally distributed, $R = .98 \pm .01$, than the across-session variability across subjects, $R = .97 \pm .01$, (Figure 9(c)). The across-session variability in dynamical

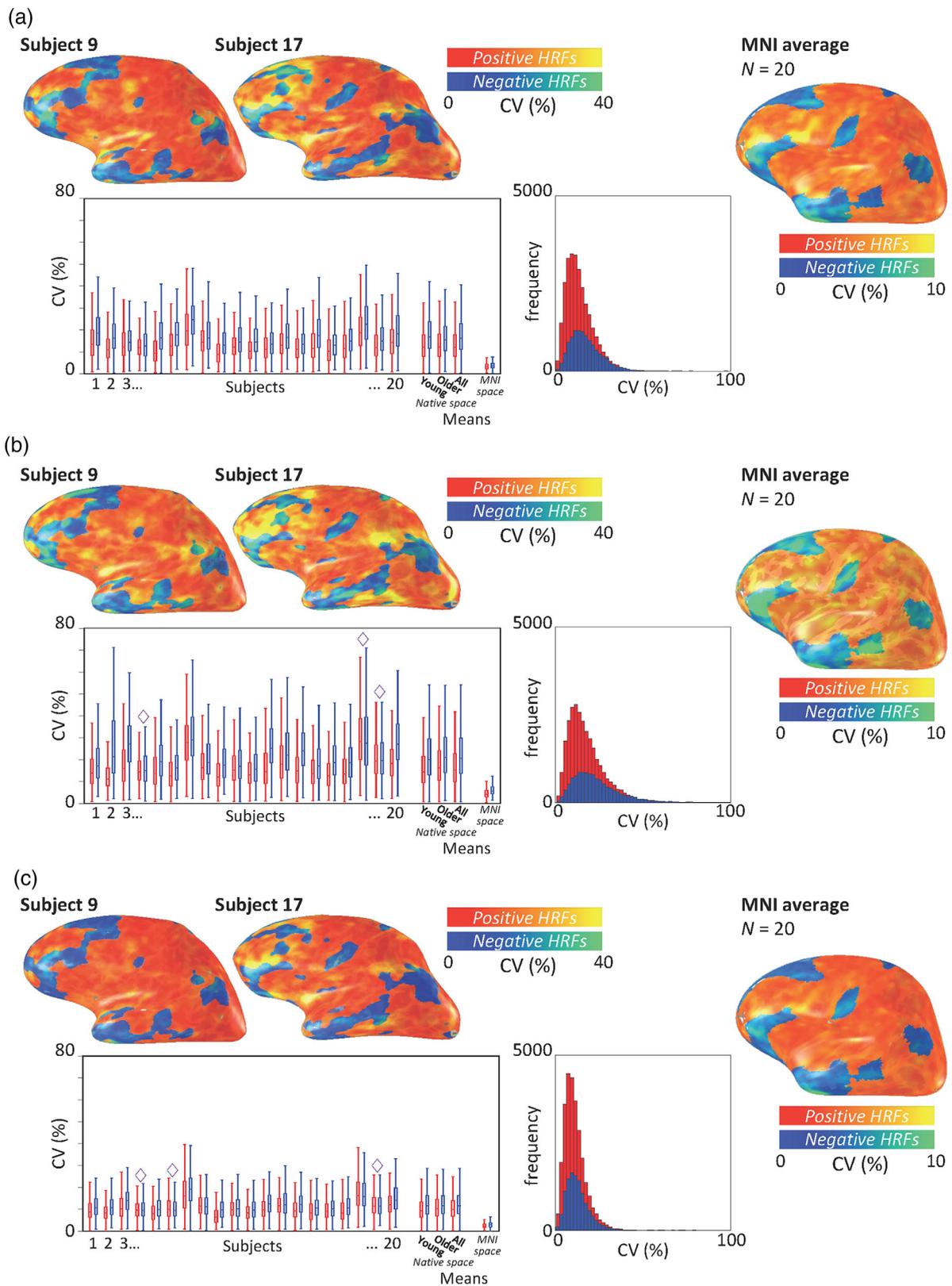
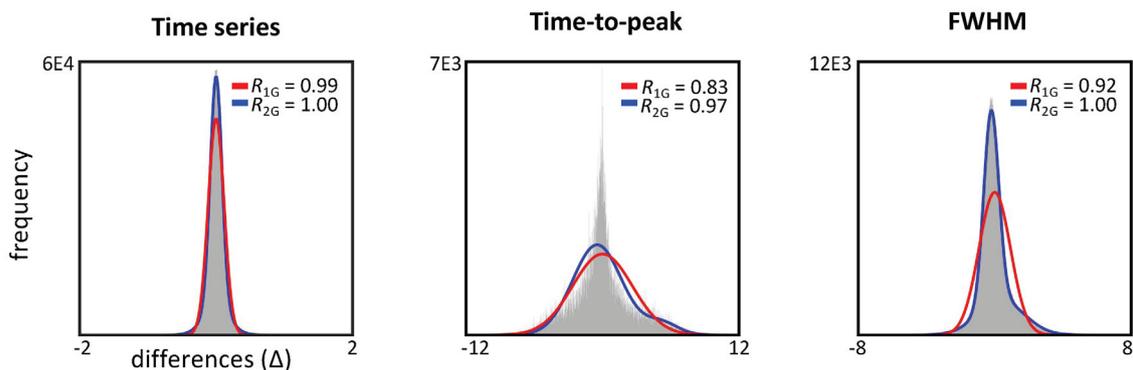
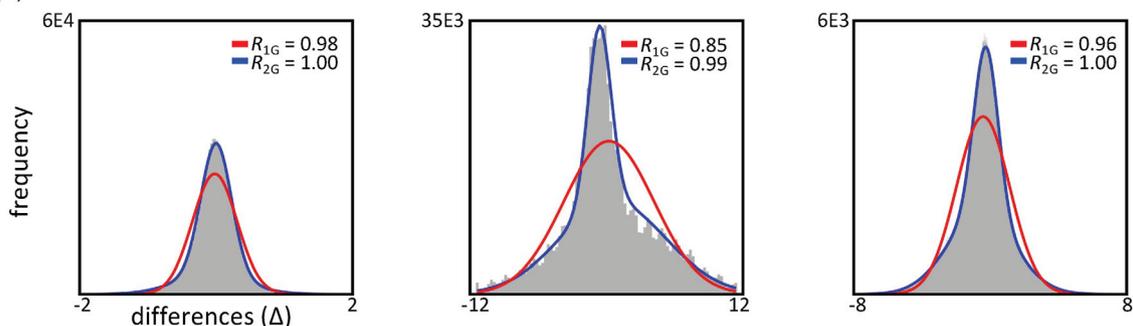


FIGURE 8 Coefficient of variation (CV) for hemodynamic response function (HRF) temporal parameters: (a) peak time, (b) full-width at half-maximum (FWHM), and (c) undershoot time. Overlays show CV with separate color scales for positive and negative HRFs in two example subjects and for the across-subject average on the MNI surface. Whisker plots show the median CV and distributions across cortex for each subject (positive HRFs in red-yellow; negative HRFs in blue-green). CV for positive and negative HRFs is significantly different for all subjects, $p < .05$, except in those subjects marked with purple diamonds. Histograms (positive, red; negative, blue) show the distributions of CV across all subjects. There is no significant difference in spatial mean CV between the two age groups

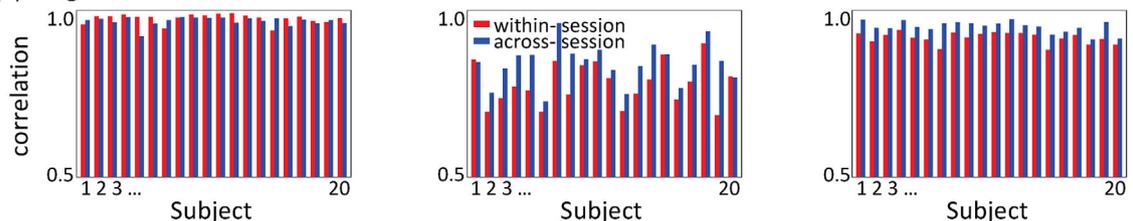
(a) Within-session differences



(b) Across-session differences



(c) Single Gaussian fit correlations



(d) Across- vs. within-session variability

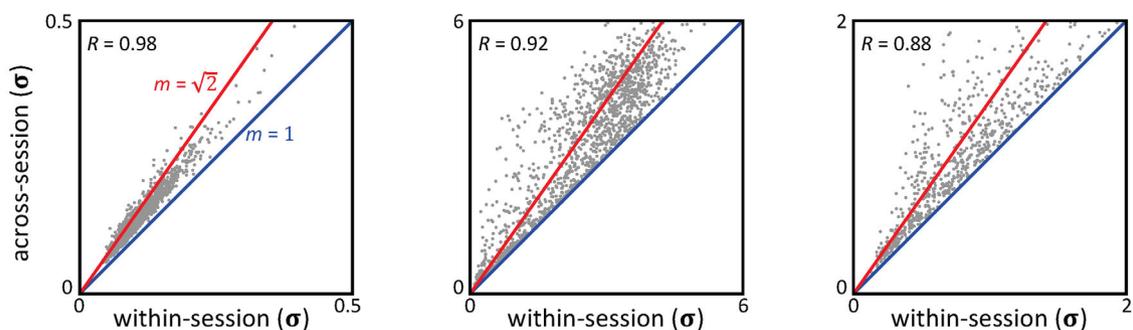


FIGURE 9 (a) Histograms and scatterplots show distributions of variability in the hemodynamic response function (HRF) time series, time-to-peak (TTP), and full-width at half-maximum (FWHM) for Subject 9. The first-row histograms show variability distributions for the time series, TTP, and FWHM calculated across events *within* the 3-month session. A single Gaussian fit (red curve) and a sum-of-two-Gaussians fit (blue curve) is shown for each distribution with Pearson correlation for each fit. (b) The second-row histograms show the distribution of variability calculated *across* time-points t_0 and 3m; Gaussian fits are shown for each distribution. (c) Correlations to the single-Gaussian fits of within-session variability (red bars) and across-session variability (blue bars) for all subjects. (d) Scatterplots compare the variability *across* the 3m and t_0 time-points to variability *within* the t_0 time-point for Subject 9; correlations between sigma values are also shown. Line of identity (blue) and the square root of two (red) line are plotted for reference

parameters (TTP and FWHM) was more normally distributed with slightly better fits, $R_{TTP} = .84 \pm .06$ and $R_{FWHM} = .95 \pm .02$ than within-session noise, $R_{TTP} = .78 \pm .06$ and $R_{FWHM} = .92 \pm .02$.

We then compared the sigma values from single Gaussian models at each vertex between the across-session and within-session variability, (Figure 9(d)). Results are shown for Subject 9, but all subjects were

similar. Sigma values representing across-session variability are highly correlated with the sigma values for within-session variability for most vertices. The time series correlation was strongest, $R = .93 \pm .07$ averaged across subjects. Correlation was somewhat lower for the dynamical parameters; for TTP, $R = .88 \pm .04$, and for FWHM, $R = .83 \pm .08$. Most of the across-to-within-session variability points in Figure 9(d) fall between the line of identity and a line with a square root of two slopes. Vertices in this range suggest that the noise processes that create the observed variability are not fully independent but somewhat correlated between sessions.

Because the Gaussian fits did not fully explain the measurements, distributions were also modeled by a sum-of-two Gaussians to test the hypothesis that two independent sources of variability were present. For both within- and across-session variability, this somewhat improved the fit quality (signal explained increased by $\leq 8\%$ for time series, $\leq 43\%$ for TTP, and $\leq 13\%$ for FWHM). However, when correlations between within- and across-session variability were examined for the dual-Gaussian fits, they proved to be much weaker than the single Gaussian fits (Figure S5), suggesting that the hypothesis of dual independent sources of variability was not descriptive.

3.6 | Comparison of positive and negative HRF parameter distributions

Finally, we performed a simple comparison of the positive and negative HRF parameters: peak amplitude, TTP, and FWHM (Figure 10). The absolute values of the peak amplitudes are presented across all subjects in MNI space for all session time-points combined. The distributions of negative HRF parameters are clearly distinct from those of positive HRF parameters representing significantly different populations ($p \approx 0$ for all parameters).

4 | DISCUSSION

4.1 | Major discussion topics

This study compared longitudinal fMRI data measured across four sessions ranging up to a 3-month interval. We used high-resolution fMRI

(2-mm cubic voxels) and a stimulus paradigm that consistently activates $>70\%$ of the cortical surface including auditory, motor, visual, default-mode network, and executive cortices. We used depth mapping to select only the central gray matter to minimize partial-volume effects. Our approach utilizes the full 2-mm resolution of the fMRI acquisition to map responses onto the gray matter, with subsequent smoothing performed only upon the gray-matter manifold. This is essentially a form of anisotropic spatial resolution that greatly avoids contamination of gray-matter functional signals with weaker white-matter signals and less-reliable signals from superficial vascular structures (Kim & Ress, 2017).

We found that the HRF time series across the cortical surface were quite consistent across the sessions. The high repeatability indicated that the HRF evoked by this paradigm has excellent long-term reliability when measured primarily in the central gray matter. In general, all characteristics of the positive HRFs were more stable than in negative HRFs. FSE measurements showed that HRFs of the reference session explained $>60\%$ of the positive HRFs and $>45\%$ of negative HRFs in the fourth session, 3 months later. Session-to-session variability that was observed in the HRF time series was primarily evident in the late-time behavior of the BOLD signal as it returns to baseline. This is consistent with our previous findings in which late temporal dynamics are more variable across events than early BOLD dynamics (Aguirre et al., 1998; Handwerker et al., 2004; Kim et al., 2019). Both sustained oxygen metabolism and oscillatory blood flow mechanisms may contribute to this variability. However, we also discovered that variations between sessions were not unique to the latter portion of the HRF. RMSE and CNR calculations revealed a baseline variation for both positive and negative HRFs throughout the HRF and an increase near the HRF TTP. These results suggest that variation across sessions has two sources: a baseline (time-independent) source observed in RMSE correlations with low CNR values, and a stimulus-driven source observed in RMSE correlations to high CNR. These two sources are consistent with previous studies that evaluated the sources of BOLD noise (Krüger & Glover, 2001; Wald & Polimeni, 2017). Specifically, the stimulus-evoked component could correspond to physiological noise that scales with changes in the measured signal that produce BOLD contrast. Results show that the stimulus-driven effects are noisier in negative HRFs with larger CNR dependence of the RMSE. Meanwhile the baseline component could

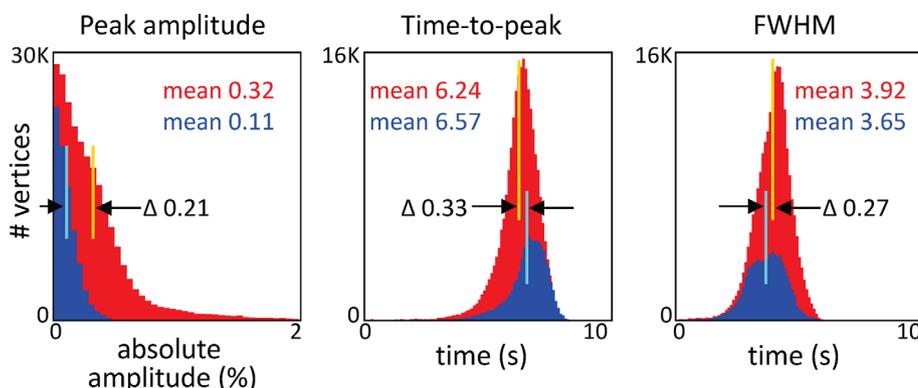


FIGURE 10 Parameters from all session time-points for all the MNI average surface vertices were separated into positive (red) and negative (blue) hemodynamic response function (HRF) peak amplitudes. Absolute value of the peak amplitudes, time-to-peak, and full-width at half-maximum (FWHM) distributions are shown for each set of HRFs. The means (positive—yellow, negative—cyan) and difference values (Δ) are shown for each distribution. All differences are very significant ($p < 10^{-6}$)

reflect other physiological noise sources, such as cardiac pulsatility and respiration, which are proportional to the overall signal amplitude but not signal fluctuations. Considering this, we conclude that the apparent late-time variability is not a stimulus-evoked effect, but rather a consequence of its weaker signal amplitude competing with time-independent, baseline noise sources.

Previous work indicates that variations between sessions in a subject can be attributed to differences in local activation pattern, with cognitive state and local vascularization playing major roles (Lee et al., 1995; Neumann et al., 2003; Rangaprakash et al., 2018). In our stimulus paradigm, differences in baseline cognitive state were minimized by using a nondemanding fixation task contrasted with a brief but demanding sequence-following task. This generated consistent activation patterns; more than 70% of cortex was activated in each session with highly correlated HRF peak amplitudes. Under(over-)shoot amplitudes were only slightly less correlated than their peak counterparts, suggesting that both blood flow and relative oxygen metabolism were consistent over these long timescales. The strong consistency of the HRF responses, is at least partly the consequence of our use of a short stimulus protocol that avoids slow changes neural responses such as adaptation or variable attention. The weak variability across sessions suggests that infrequent presentations of a speeded task evoke a remarkably stereotypical pattern of response. It may be that the brain has a relatively nonspecific vascular response to a speeded task, delivering metabolic substrates to multiple intrinsic connectivity networks in preparation for possible cognitive demands.

Within-subject amplitude variations may stem from different cognitive states (arousal and attention) and learning effects between and across measurements (Miller et al., 2002; Miller, Handy, et al., 2001; Munneke et al., 2008; Petersson et al., 1999). To measure the consistency of fMRI data, many studies have examined test–retest reliability across sessions (Bennett & Miller, 2010; Noble et al., 2019; Vul et al., 2009). However, as noted previously, most of these experiments used sustained stimuli that are likely subject to additional neuronal and cognitive variability that are partially avoided by our short-stimulus approach. Quantitative comparison with these studies is additionally confounded by the many methods used to quantify reliability such as CV, Dice coefficient of similarity, and Jaccard overlap (Bennett & Miller, 2013). Nevertheless, most of these measurements found moderate to fair within-session reliability (Bennett & Miller, 2010), which is consistent with our results here and previously (Taylor et al., 2018). However, other investigators found that between-session variability was generally worse, except for motor and sensory regions. Our results are partially consistent with these observations, in that regions of strong activation were generally more reliable across sessions than in regions of weaker activation (Bennett & Miller, 2010; Holiga et al., 2018). Our results differ in that between-session variability was generally quite low, probably because of the globally high activation amplitudes evoked across the majority of cortex.

Our subjects were evenly divided two age ranges: young (20–35) and late middle age (50–64). For all spatially averaged metrics of variability, we observed no significant differences between these two age

groups. However, detailed analysis reveals spatially variant regions of significant differences, but analysis of this topic will be reserved for later work.

Besides CNR, one of the strongest predictors of variability was the polarity of the HRF. Almost 75% of activated cortex generated positive HRFs—found in sensory, motor, association, and executive areas. These areas were specifically targeted by the stimulus and were highly consistent across sessions. However, the amplitudes and dynamics of negative HRFs displayed higher variability across session time-points than the positive HRFs. Negative HRFs were nearly half as correlated for amplitude parameters, and they were about 30% more variable in dynamical parameters than positive HRFs. Positive and negative HRFs were classified based on the polarity of the largest signal magnitude of the bootstrapped-mean HRF across events. Reliability of the negative–positive discrimination across cortex (Supporting Information, Figure S1), shows that a minority of HRFs are quite ambiguous of sign—these fall into two classes: weak HRFs, or HRFs in boundary regions between strongly positive and negative areas. Metrics that used a CNR threshold usually cull these HRFs. More broadly, our simple bipartite HRF classification was likely enabled by our use of a short stimulus that separates the timescales of neuronal response from the evoked neurovascular and neurometabolic responses.

Negative BOLD response has been observed in many studies as a task-evoked response, often surrounding strongly activated regions of positive BOLD (DeLaRosa et al., 2021; Shmuel et al., 2002; Shmuel et al., 2006; Shulman et al., 1997; Wade & Rowland, 2010). It has been suggested that negative HRFs are not simple inversions of positive HRFs; instead they have other neurophysiological origins, such as poststimulus neural responses (Mullinger et al., 2014), or strong ringing of the flow response (DeLaRosa et al., 2021). Negative BOLD is likely an effect of neural suppression (Devor et al., 2007; Maggioni et al., 2016; Pasley et al., 2007; Stefanovic et al., 2004) but may also be linked to decreases in local blood supply (Puckett et al., 2014; Smith et al., 2004) or a combination of these effects via neurovascular coupling (Shmuel et al., 2006; Stefanovic et al., 2004). Negative HRFs observed by Puckett et al. (2014), in the visual cortex had notably higher variability than positive counterparts, consistent with our own results. However, our data indicate that negative HRFs generally appear to show more variability than positive HRFs because of their weaker peak amplitude.

In this study, negative signals were not specifically observed adjacent to strong positive activation, but rather in large clusters. It is likely that this is partially the consequence of the broad activation evoked by our stimulus paradigm. Most negative HRFs were observed in areas often attributed to the default mode network (DMN), a task-negative functional network found across human subjects (Buckner et al., 2008; Greicius et al., 2003; Raichle et al., 2001). Negative BOLD signals in the DMN consistently suspend activity upon cognitive activation independent of the task, although the extent of deactivation depends on the cognitive load (Esposito et al., 2006). The higher variability measured in these regions is consistent with previous studies, which showed that mechanisms of downregulation in the DMN have

confounding transient behavior modulated by attention (McCormick & Telzer, 2018; Ossandón et al., 2011). However, part of the increased variability may simply be the consequence of the relatively weaker amplitudes of the negative HRFs, supported by our RMSE calculations.

We found that all dynamical parameters (TTU, FWHM, and TTU) were quite stable across sessions. In absolute temporal units, the least variation was observed in TTP, and largest variation in TTU. However, lowest CV was observed for TTU, which is consistent with the lower RMSE observed during the later period of the HRF. Our measurement of mean across-cortex variation in TTP was only 0.8 s, while it was found to be 0.5 s in visual cortex-only HRFs in a previous study (de Zwart et al., 2005). FWHM varied up to 0.8 s in our study, whereas in the previous study FWHM varied up to 0.4 s. The larger variability observed in our study may correspond to the much larger region of cortex evaluated. Notably, the fractional variation of TTU was the most stable of dynamical variables. This further supports a linkage between the dynamics that create the peak and undershoot, such as underdamped flow oscillations that we proposed previously (Kim & Ress, 2016), and supported by direct flow measurements (Kim et al., 2019). Altogether, temporal dynamics were repeatable with low relative variability across session time-points.

We suspected that within-session variability was a major factor in the observed longitudinal variability of HRF dynamics, that is, voxels having less trial-by-trial variability typically show more consistent results across the four sessions. We found strong correlation between within-sessions variability and across-session variability with most data lying between lines of identity and $\sqrt{2}$. This suggests that most variability can be explained by local fluctuations of the hemodynamic response, which are observed across HRF measurements within a session. Furthermore, the metric (difference calculation) used in the correlations suggests that much of the observed variability should not be attributed to white noise. Rather, spontaneous fluctuations, as reported in resting-state functional experiments, may generate notable differences across trials (Petridou et al., 2013; Wu & Marinazzo, 2016). This hypothesis is also supported by the observed temporal structure of the RMSE (Figure 4(d)), indicating baseline and stimulus-evoked components; the baseline component could partially arise from stimulus-independent spontaneous fluctuations.

While fitting the within- and across-session variability, we noted that a single Gaussian function fit the data relatively well, and a sum-of-two-Gaussians function slightly improved the fit quality. However, the dual-Gaussian model parameters showed weak correlation between the variability metrics, unlike the well-correlated single Gaussian fits. This suggests that a single, strong source of variability is found both within and across sessions. Despite our efforts to engage and control subject attention and behavior during the fMRI measurements, low-frequency fluctuations, the focus of resting-state fMRI studies, likely still contribute to the observed time series (Tong et al., 2019). There is evidence that these fluctuations are repeatable over time (Biswal et al., 1995), which is consistent with our observation of strong correlation between the across-session and within-session variabilities observed at many vertices in each brain.

The observed Gaussian distributions of variability were somewhat surprising because typical fMRI noise is non-Gaussian. Aside from thermal noise, significant amounts of noise arise in high magnetic field (≥ 3 T) measurements from cardiac and respiratory artifacts (about 10% of total noise) as well as fluctuations in blood flow, blood volume, and oxygen metabolism (about 25% of total noise) (Krüger & Glover, 2001). The distribution of noise is mostly uniform across cortex with variations in certain gray-matter regions due to local mechanisms (Krüger & Glover, 2001; Tong & Frederick, 2014), such as cardiac pulsatility (Glover et al., 2000). Variability was minimized by our use of full native spatial resolution to focus on the central gray-matter parenchyma, avoiding lower-reliability signals in superficial vasculature and tissues adjacent to the gray matter (Kim & Ress, 2016). Additionally, our use of large surface smoothing whitens the noise, increasing its normality. This observed normal character of noise makes the statistical model associated with fMRI experiments more tractable. Thus, our approach of using full MRI resolution to resolve the gray matter, followed by smoothing along the gray-matter manifold appears to offer a CNR advantage as compared to uniform spatial smoothing across the volume.

4.2 | Limitations

Our stimulus paradigm with a brief, demanding stimulus and task followed by a nondemanding fixation task, has shown remarkable reliability across subjects and sessions. The contrast between tasks activates the majority of cortex allowing the study of global activity. However, the fixation task evokes a less demanding but nonzero baseline signal that may introduce confounds to the HRF. The baseline task is not designed to be difficult, but it does require modest vigilance. Poor baseline behavior suggests that it may be of interest to use eye tracking in future studies. This rigorous stimulus and task paradigm is suited to a general cohort of subjects; however, it may not serve certain clinical settings or subjects.

Even with our high spatial sampling, 2-mm isotropic voxels cannot resolve the thinnest regions of gray matter, which range 1–4.5 mm (Fischl & Dale, 2000). However, the thinnest regions are usually found in the sulcal fundi and make up less than 10% of cortex (Fischl & Dale, 2000; Markowitsch & Tulving, 1994). In our measurements $30.0 \pm 5.4\%$ of cortex is thinner than 2 mm and may contribute partial-volume effects. A thickness-dependent study of HRF variability within and across subjects is reserved for future work. Moreover, we relied on structural volumes with 1-mm sampling, and segmentation and thickness calculations at higher resolution should yield more precise results (van der Kouwe et al., 2008).

Temporal resolution was limited by the 1.25-s TR of the fMRI measurements, which we interpolated to 0.1-s sampling prior to averaging across events. This temporal sampling should be sufficient to capture the sluggish blood flow and metabolic responses that give rise to the BOLD response (Buxton et al., 1998; Kim & Ress, 2016). Moreover, the intensive averaging over many events filters out frequencies above 0.4 Hz, including cardiac and respiratory artifacts (Lewis

et al., 2016). It has been noted that a time-locked approach, such as ours, provides poor estimation efficiency (Burock & Dale, 2000; Liu, 2004). Nevertheless, our results show excellent CNR, suggesting that our approach effectively removes high-frequency nuisance such as pulse and respiration.

During HRF parameterization, TTP was chosen as the time-point with largest magnitude between 2 and 14 s of the HRF time series. This range should be large enough to cover TTP measurements from previous fMRI studies and should not affect most of the data (Hirano et al., 2011; Martindale et al., 2003; Uludağ et al., 2004). However, the artificial boundaries combined with discretization of the temporal measurements generated discontinuous distributions in the bootstrapped averaged data and may have been the source of some of the observed kurtosis in the distributions.

Seasonal and circadian variations in the HRF were not fully balanced in our study. While measurements were obtained in all months of the year, some months were more strongly represented than others; January had the fewest, while June the largest number of measurements. Moreover, measurements were not explicitly balanced across the day, but were, in fact, generally obtained during normal business hours. These issues may have slightly confounded our results.

The analysis that we present here was somewhat limited in scope because of space constraints, and much further analysis is possible. In particular, our rich dataset will enable a far more complete analysis of across-subject characteristics that will enable rankings of spatial, temporal, and across-subject noise for the majority of cortex. In an effort to balance variation associated with age, we scanned adults from young and late-middle-age populations. There were interesting, spatially localized differences between the two age groups, but this topic exceeds the scope of this article. Moreover, we plan to evaluate the data for differences associated with sex. Finally, it should be possible to examine variability as a function of cortical geometry, specifically gray-matter thickness and curvature. These topics will all be evaluated in future work.

5 | CONCLUSIONS

Our results demonstrate that our stimulus and task protocol evoked strong HRFs across the majority of cortex that are remarkably repeatable across sessions separated by long time intervals, up to 3 months. HRF time series and parameters (amplitudes and dynamics) are highly repeatable within a subject. Amplitude patterns were not only consistent across sessions in a subject but also highly correlated across subjects, consistent with our previous findings. Regions of positive and negative HRFs are consistently observed across sessions and subjects. Both amplitude and dynamical parameters differ significantly between positive and negative HRFs. The HRF time series and temporal parameters were also repeatable across session time-points. Early temporal dynamics are most stable across sessions. Positive HRFs are generally more stable than negative HRFs. Across-session variability is highly correlated with within-session variability. In general, the observed long-term reliability across subjects of the HRFs evoked by our

stimulus and task suggests its utility as a diagnostic tool for activation of the majority of cortex to study pathology or healthy brain function. More generally, the results help to quantify the reliability of event-related fMRI experimental measurement across sessions.

ACKNOWLEDGMENT

The authors thank Qureshi Asma, Rahul Goel, Xiaowei Zou, and Lacey Berry for providing thoughtful comments and assistance with data acquisition and analysis.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All software used in our analysis is part of the mrHirbil package, which is available for download upon request. Data from these experiments can also be made available for download in the form of the original DICOM image sets from the scanner. Such data have already been deidentified as required by our IRB protocol. Raw data were generated at Baylor College of Medicine (Houston, TX, USA). Derived data supporting the findings of this study are available from the corresponding author, David Ress on request.

ORCID

Jung Hwan Kim  <https://orcid.org/0000-0003-0053-8321>

David Ress  <https://orcid.org/0000-0002-7326-4889>

REFERENCES

- Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *NeuroImage*, 8, 360–369.
- Bailey, C. J., Sanganahalli, B. G., Herman, P., Blumenfeld, H., Gjedde, A., & Hyder, F. (2013). Analysis of time and space invariance of BOLD responses in the rat visual system. *Cerebral Cortex*, 23, 210–222.
- Bandettini, P. A., & Cox, R. W. (2000). Event-related fMRI contrast when using constant interstimulus interval: Theory and experiment. *Magnetic Resonance in Medicine*, 43, 540–548.
- Belloy, M. E., Billings, J., Abbas, A., Kashyap, A., Pan, W.-J., Hinz, R., Vanreusel, V., Van Audekerke, J., Van der Linden, A., Keilholz, S. D., Verhoye, M., & Keliris, G. A. (2020). Resting brain fluctuations are intrinsically coupled to visual response dynamics. *Cerebral Cortex*, 31, 1511–1522.
- Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences*, 1191, 133–155.
- Bennett, C. M., & Miller, M. B. (2013). fMRI reliability: Influences of task and experimental design. *Cognitive, Affective, & Behavioral Neuroscience*, 13, 690–702.
- Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, 34, 537–541.
- Bland, J. M., & Altman, D. G. (1996). Measurement error proportional to the mean. *BMJ*, 313, 106.
- Bonakdarpour, B., Parrish, T., & Thompson, C. (2007). Hemodynamic response function in patients with stroke-induced aphasia: Implications for fMRI data analysis. *NeuroImage*, 36, 322–331.
- Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience*, 16, 4207–4221.
- Boynton, G. M., Engel, S. A., & Heeger, D. J. (2012). Linear systems analysis of the fMRI signal. *NeuroImage*, 62, 975–984.

- Breuer, F. A., Blaimer, M., Heidemann, R. M., Mueller, M. F., Griswold, M. A., & Jakob, P. M. (2005). Controlled aliasing in parallel imaging results in higher acceleration (CAIPIRINHA) for multi-slice imaging. *Magnetic Resonance in Medicine*, *53*, 684–691.
- Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, *1124*, 1–38.
- Burock, M. A., & Dale, A. M. (2000). Estimation and detection of event-related fMRI signals with temporally correlated noise: A statistically efficient and unbiased approach. *Human Brain Mapping*, *11*, 249–260.
- Buxton, R. B., Uludag, K., Dubowitz, D. J., & Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *NeuroImage*, *23*(Suppl 1), S220–S233.
- Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magnetic Resonance in Medicine*, *39*, 855–864.
- Cohen, E. R., Ugurbil, K., & Kim, S. G. (2002). Effect of basal conditions on the magnitude and dynamics of the blood oxygenation level-dependent fMRI response. *Journal of Cerebral Blood Flow and Metabolism*, *22*, 1042–1053.
- Collins, D. L., Neelin, P., Peters, T. M., & Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography*, *18*, 192–205.
- Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, *5*, 329–340.
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, *9*, 179–194.
- Davis, T. L., Kwong, K. K., Weisskoff, R. M., & Rosen, B. R. (1998). Calibrated functional MRI: Mapping the dynamics of oxidative metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 1834–1839.
- de Zwart, J. A., Silva, A. C., van Gelderen, P., Kellman, P., Fukunaga, M., Chu, R., Koretsky, A. P., Frank, J. A., & Duyn, J. H. (2005). Temporal dynamics of the BOLD fMRI impulse response. *NeuroImage*, *24*, 667–677.
- DeLaRosa, N., Ress, D., Taylor, A. J., & Kim, J. H. (2021). Retinotopic variations of the negative blood-oxygen-level dependent hemodynamic response function in human primary visual cortex. *Journal of Neurophysiology*, *125*, 1045–1057.
- D'Esposito, M., Deouell, L. Y., & Gazzaley, A. (2003). Alterations in the BOLD fMRI signal with ageing and disease: A challenge for neuroimaging. *Nature Reviews Neuroscience*, *4*, 863–872.
- Devor, A., Tian, P., Nishimura, N., Teng, I. C., Hillman, E. M. C., Narayanan, S. N., Ulbert, I., Boas, D. A., Kleinfeld, D., & Dale, A. M. (2007). Suppressed neuronal activity and concurrent arteriolar vasoconstriction may explain negative blood oxygenation level-dependent signal. *The Journal of Neuroscience*, *27*, 4452–4459.
- Drew, P. J., Shih, A. Y., & Kleinfeld, D. (2011). Fluctuating and sensory-induced vasodynamics in rodent cortex extend arteriole capacity. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 8473–8478.
- Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, *68*, 589–599.
- Elbau, I. G., Bruckmeier, B., Uhr, M., Arloth, J., Czamara, D., Spormaker, V. I., Czisch, M., Stephan, K. E., Binder, E. B., & Samann, P. G. (2018). The brain's hemodynamic response function rapidly changes under acute psychosocial stress in association with genetic and endocrine stress response markers. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E10206–E10215.
- Esposito, F., Bertolino, A., Scarabino, T., Latorre, V., Blasi, G., Popolizio, T., Tedeschi, G., Cirillo, S., Goebel, R., & Di Salle, F. (2006). Independent component model of the default-mode brain function: Assessing the impact of active thinking. *Brain Research Bulletin*, *70*, 263–269.
- Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 11050–11055.
- Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, *9*, 195–207.
- Friman, O., Borge, M., Lundberg, P., & Knutsson, H. (2003). Adaptive analysis of fMRI data. *NeuroImage*, *19*, 837–845.
- 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.
- Friston, K. J., Josephs, O., Rees, G., & Turner, R. (1998). Nonlinear event-related responses in fMRI. *Magnetic Resonance in Medicine*, *39*, 41–52.
- Fox, M., & Raichle, M. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci*, *8*, 700–711. <https://doi.org/10.1038/nrn2201>
- Gagnon, L., Sakadzic, S., Lesage, F., Musacchia, J. J., Lefebvre, J., Fang, Q., Yucel, M. A., Evans, K. C., Mandeville, E. T., Cohen-Adad, J., Polimeni, J. R., Yaseen, M. A., Lo, E. H., Greve, D. N., Buxton, R. B., Dale, A. M., Devor, A., & Boas, D. A. (2015). Quantifying the microvascular origin of BOLD-fMRI from first principles with two-photon microscopy and an oxygen-sensitive nanoprobe. *The Journal of Neuroscience*, *35*, 3663–3675.
- Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, *9*, 416–429.
- Glover, G. H., Li, T. Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic Resonance in Medicine*, *44*, 162–167.
- Gonzalez-Castillo, J., Hoy, C. W., Handwerker, D. A., Roopchansingh, V., Inati, S. J., Saad, Z. S., Cox, R. W., & Bandettini, P. A. (2015). Task dependence, tissue specificity, and spatial distribution of widespread activations in large single-subject functional MRI datasets at 7T. *Cerebral Cortex*, *25*, 4667–4677.
- Grady, C. L., & Garrett, D. D. (2014). Understanding variability in the BOLD signal and why it matters for aging. *Brain Imaging and Behavior*, *8*, 274–283.
- 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. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 253–258.
- 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.
- Harel, N., Lee, S.-P., Nagaoka, T., Kim, D.-S., & Kim, S.-G. (2002). Origin of negative blood oxygenation level-dependent fMRI signals. *Journal of Cerebral Blood Flow & Metabolism*, *22*, 908–917.
- Hillman, E. M., Devor, A., Bouchard, M. B., Dunn, A. K., Krauss, G. W., Skoch, J., Bacskai, B. J., Dale, A. M., & Boas, D. A. (2007). Depth-resolved optical imaging and microscopy of vascular compartment dynamics during somatosensory stimulation. *NeuroImage*, *35*, 89–104.
- Hirano, Y., Stefanovic, B., & Silva, A. C. (2011). Spatiotemporal evolution of the functional magnetic resonance imaging response to ultrashort stimuli. *The Journal of Neuroscience*, *31*, 1440–1447.
- Holiga, Š., Sambataro, F., Luzy, C., Greig, G., Sarkar, N., Renken, R. J., Marsman, J.-B. C., Schobel, S. A., Bertolino, A., & Dukart, J. (2018). Test-retest reliability of task-based and resting-state blood oxygen level dependence and cerebral blood flow measures. *PLoS One*, *13*, e0206583.
- Huettel, S. A., McKeown, M. J., Song, A. W., Hart, S., Spencer, D. D., Allison, T., & McCarthy, G. (2004). Linking hemodynamic and electrophysiological measures of brain activity: Evidence from functional MRI and intracranial field potentials. *Cerebral Cortex*, *14*, 165–173.
- Iadecola, C., Yang, G., Ebner, T. J., & Chen, G. (1997). Local and propagated vascular responses evoked by focal synaptic activity in cerebellar cortex. *Journal of Neurophysiology*, *78*, 651–659.

- Keilholz, S. D., Silva, A. C., Raman, M., Merkle, H., & Koretsky, A. P. (2004). Functional MRI of the rodent somatosensory pathway using multislice echo planar imaging. *Magnetic Resonance in Medicine*, *52*, 89–99.
- Khan, R., Zhang, Q., Darayan, S., Dhandapani, S., Katyal, S., Greene, C., Bajaj, C., & Ress, D. (2011). Surface-based analysis methods for high-resolution functional magnetic resonance imaging. *Graphical Models*, *73*, 313–322.
- Kim, J. H., Khan, R., Thompson, J. K., & Ress, D. (2013). Model of the transient neurovascular response based on prompt arterial dilation. *Journal of Cerebral Blood Flow and Metabolism*, *33*, 1429–1439.
- Kim, J. H., & Ress, D. (2016). Arterial impulse model for the BOLD response to brief neural activation. *NeuroImage*, *124*, 394–408.
- Kim, J. H., & Ress, D. (2017). Reliability of the depth-dependent high-resolution BOLD hemodynamic response in human visual cortex and vicinity. *Magnetic Resonance Imaging*, *39*, 53–63.
- Kim, J. H., Taylor, A. J., Wang, D. J. J., Zou, X., & Ress, D. (2019). Dynamics of the cerebral blood flow response to brief neural activity in human visual cortex. *Journal of Cerebral Blood Flow & Metabolism*, *40*, 1823–1837.
- Kim, S. G., Richter, W., & Ugurbil, K. (1997). Limitations of temporal resolution in functional MRI. *Magnetic Resonance in Medicine*, *37*, 631–636.
- Klingner, C. M., Ebenau, K., Hasler, C., Brodoehl, S., Gorlich, Y., & Witte, O. W. (2011). Influences of negative BOLD responses on positive BOLD responses. *NeuroImage*, *55*, 1709–1715.
- Krüger, G., & Glover, G. H. (2001). Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magnetic Resonance in Medicine*, *46*, 631–637.
- Lee, A. T., Glover, G. H., & Meyer, C. H. (1995). Discrimination of large venous vessels in time-course spiral blood-oxygen-level-dependent magnetic-resonance functional neuroimaging. *Magnetic Resonance in Medicine*, *33*, 745–754.
- Leontiev, O., & Buxton, R. B. (2007). Reproducibility of BOLD, perfusion, and CMRO₂ measurements with calibrated-BOLD fMRI. *NeuroImage*, *35*, 175–184.
- Lewis, L. D., Setsompop, K., Rosen, B. R., & Polimeni, J. R. (2016). Fast fMRI can detect oscillatory neural activity in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, E6679–E6685.
- Li, M., Newton, A. T., Anderson, A. W., Ding, Z., & Gore, J. C. (2019). Characterization of the hemodynamic response function in white matter tracts for event-related fMRI. *Nature Communications*, *10*, 1140.
- Lindvere, L., Janik, R., Dorr, A., Chartash, D., Sahota, B., Sled, J. G., & Stefanovic, B. (2013). Cerebral microvascular network geometry changes in response to functional stimulation. *NeuroImage*, *71*, 248–259.
- Liu, T. T. (2004). Efficiency, power, and entropy in event-related fMRI with multiple trial types. Part II: Design of experiments. *NeuroImage*, *21*, 401–413.
- 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.
- Maggioni, E., Zucca, C., Reni, G., Cerutti, S., Triulzi, F. M., Bianchi, A. M., & Arrigoni, F. (2016). Investigation of the electrophysiological correlates of negative BOLD response during intermittent photic stimulation: An EEG-fMRI study. *Human Brain Mapping*, *37*, 2247–2262.
- Magnuson, M., Thompson, G., Pan, W., & Keilholz, S. (2014). Time-dependent effects of isoflurane and dexmedetomidine on functional connectivity, spectral characteristics, and spatial distribution of spontaneous BOLD fluctuations. *NMR in Biomedicine*, *27*, 291–303.
- Markowitsch, H. J., & Tulving, E. (1994). Cognitive processes and cerebral cortical fundi: Findings from positron-emission tomography studies. *Proceedings of the National Academy of Sciences of the United States of America*, *91*, 10507–10511.
- Martindale, J., Mayhew, J., Berwick, J., Jones, M., Martin, C., Johnston, D., Redgrave, P., & Zheng, Y. (2003). The hemodynamic impulse response to a single neural event. *Journal of Cerebral Blood Flow & Metabolism*, *23*, 546–555.
- Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., ... Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (ICBM). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *356*, 1293–1322.
- McCormick, E. M., & Telzer, E. H. (2018). Contributions of default mode network stability and deactivation to adolescent task engagement. *Scientific Reports*, *8*, 18049.
- Menon, R. S., Ogawa, S., Hu, X., Strupp, J. P., Anderson, P., & Ugurbil, K. (1995). BOLD based functional MRI at 4 tesla includes a capillary bed contribution: Echo-planar imaging correlates with previous optical imaging using intrinsic signals. *Magnetic Resonance in Medicine*, *33*, 453–459.
- Menz, M. M., Neumann, J., Muller, K., & Zysset, S. (2006). Variability of the BOLD response over time: An examination of within-session differences. *NeuroImage*, *32*, 1185–1194.
- Meyer, C., Muto, V., Jaspar, M., Kussé, C., Lambot, E., Chellappa, S. L., Degueldre, C., Balteau, E., Luxen, A., Middleton, B., Archer, S. N., Collette, F., Dijk, D.-J., Phillips, C., Maquet, P., & Vandewalle, G. (2016). Seasonality in human cognitive brain responses. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 3066–3071.
- Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., & Buckner, R. L. (2000). Characterizing the hemodynamic response: Effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *NeuroImage*, *11*, 735–759.
- Miller, K. L., Luh, W. M., Liu, T. T., Martinez, A., Obata, T., Wong, E. C., Frank, L. R., & Buxton, R. B. (2001). Nonlinear temporal dynamics of the cerebral blood flow response. *Human Brain Mapping*, *13*, 1–12.
- Miller, M. B., Handy, T. C., Cutler, J., Inati, S., & Wolford, G. L. (2001). Brain activations associated with shifts in response criterion on a recognition test. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *55*, 162–173.
- Miller, M. B., Van Horn, J. D., Wolford, G. L., Handy, T. C., Valsangkar-Smyth, M., Inati, S., Grafton, S., & Gazzaniga, M. S. (2002). Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. *Journal of Cognitive Neuroscience*, *14*, 1200–1214.
- Mullinger, K. J., Mayhew, S. D., Bagshaw, A. P., Bowtell, R., & Francis, S. T. (2014). Evidence that the negative BOLD response is neuronal in origin: A simultaneous EEG-BOLD-CBF study in humans. *NeuroImage*, *94*, 263–274.
- Munneke, J., Heslenfeld, D. J., & Theeuwes, J. (2008). Directing attention to a location in space results in retinotopic activation in primary visual cortex. *Brain Research*, *1222*, 184–191.
- Neumann, J., Lohmann, G., Zysset, S., & von Cramon, D. Y. (2003). Within-subject variability of BOLD response dynamics. *NeuroImage*, *19*, 784–796.
- Noble, S., Scheinost, D., & Constable, R. T. (2019). A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *NeuroImage*, *203*, 116157.
- Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, *87*, 9868–9872.
- Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, *89*, 5951–5955.
- Olman, C. A., Inati, S., & Heeger, D. J. (2007). The effect of large veins on spatial localization with GE BOLD at 3 T: Displacement, not blurring. *NeuroImage*, *34*, 1126–1135.

- Ossandón, T., Jerbi, K., Vidal, J. R., Bayle, D. J., Henaff, M.-A., Jung, J., Minotti, L., Bertrand, O., Kahane, P., & Lachaux, J.-P. (2011). Transient suppression of broadband gamma power in the default-mode network is correlated with task complexity and subject performance. *The Journal of Neuroscience*, 31, 14521–14530.
- Pasley, B. N., Inglis, B. A., & Freeman, R. D. (2007). Analysis of oxygen metabolism implies a neural origin for the negative BOLD response in human visual cortex. *NeuroImage*, 36, 269–276.
- Petersson, K. M., Elfgrén, C., & Ingvar, M. (1999). Learning-related effects and functional neuroimaging. *Human Brain Mapping*, 7, 234–243.
- Petridou, N., Gaudes, C. C., Dryden, I. L., Francis, S. T., & Gowland, P. A. (2013). Periods of rest in fMRI contain individual spontaneous events which are related to slowly fluctuating spontaneous activity. *Human Brain Mapping*, 34, 1319–1329.
- Puckett, A. M., Mathis, J. R., & DeYoe, E. A. (2014). An investigation of positive and inverted hemodynamic response functions across multiple visual areas. *Human Brain Mapping*, 35, 5550–5564.
- 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. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 676–682.
- Rangaprakash, D., Wu, G. R., Marinazzo, D., Hu, X., & Deshpande, G. (2018). Hemodynamic response function (HRF) variability confounds resting-state fMRI functional connectivity. *Magnetic Resonance in Medicine*, 80, 1697–1713.
- Rosa, P. N., Figueiredo, P., & Silvestre, C. J. (2015). On the distinguishability of HRF models in fMRI. *Frontiers in Computational Neuroscience*, 9, 54.
- Salminen, L. E., Conturo, T. E., Bolzenius, J. D., Cabeen, R. P., Akbudak, E., & Paul, R. H. (2016). Reducing CSF partial volume effects to enhance diffusion tensor imaging metrics of brain microstructure. *Technology and Innovation*, 18, 5–20.
- Setsompop, K., Gagoski, B. A., Polimeni, J. R., Witzel, T., Wedeen, V. J., & Wald, L. L. (2012). Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic Resonance in Medicine*, 67, 1210–1224.
- Shmuel, A., Augath, M., Oeltermann, A., & Logothetis, N. K. (2006). Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nature Neuroscience*, 9, 569–577.
- Shmuel, A., Yacoub, E., Pfeuffer, J., Van de Moortele, P. F., Adriany, G., Hu, X., & Ugurbil, K. (2002). Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. *Neuron*, 36, 1195–1210.
- Shulman, G., Fiez, J., Corbetta, M., Buckner, R., Miezin, F., Raichle, M., & Petersen, S. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *Journal of Cognitive Neuroscience*, 9, 648–663.
- Silva, A. C., Koretsky, A. P., & Duyn, J. H. (2007). Functional MRI impulse response for BOLD and CBV contrast in rat somatosensory cortex. *Magnetic Resonance in Medicine*, 57, 1110–1118.
- Smith, A. T., Williams, A. L., & Singh, K. D. (2004). Negative BOLD in the visual cortex: Evidence against blood stealing. *Human Brain Mapping*, 21, 213–220.
- Snedecor, G. W., & Cochran, W. G. (1972). *Statistical Methods* (6th ed.). Iowa State University Press.
- Stefanovic, B., Warnking, J. M., & Pike, G. B. (2004). Hemodynamic and metabolic responses to neuronal inhibition. *NeuroImage*, 22, 771–778.
- Taylor, A. J., Kim, J. H., & Ress, D. (2018). Characterization of the hemodynamic response function across the majority of human cerebral cortex. *NeuroImage*, 173, 322–331.
- Thompson, J. K., Peterson, M. R., & Freeman, R. D. (2003). Single-neuron activity and tissue oxygenation in the cerebral cortex. *Science*, 299, 1070–1072.
- 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.
- Tong, Y., & Frederick, B. (2014). Studying the spatial distribution of physiological effects on BOLD signals using ultrafast fMRI. *Frontiers in Human Neuroscience*, 8, 196.
- Tong, Y., Hocke, L. M., & Frederick, B. B. (2019). Low frequency systemic hemodynamic “noise” in resting state BOLD fMRI: Characteristics, causes, implications, mitigation strategies, and applications. *Frontiers in Neuroscience*, 13, 787.
- Triantafyllou, C., Hoge, R. D., Krueger, G., Wiggins, C. J., Potthast, A., Wiggins, G. C., & Wald, L. L. (2005). Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. *NeuroImage*, 26, 243–250.
- Truccolo, W. A., Ding, M., Knuth, K. H., Nakamura, R., & Bressler, S. L. (2002). Trial-to-trial variability of cortical evoked responses: Implications for the analysis of functional connectivity. *Clinical Neurophysiology*, 113, 206–226.
- Turner, R. (2002). How much cortex can a vein drain? Downstream dilution of activation-related cerebral blood oxygenation changes. *NeuroImage*, 16, 1062–1067.
- Uludağ, K., Dubowitz, D. J., Yoder, E. J., Restom, K., Liu, T. T., & Buxton, R. B. (2004). Coupling of cerebral blood flow and oxygen consumption during physiological activation and deactivation measured with fMRI. *NeuroImage*, 23, 148–155.
- van der Kouwe, A. J. W., Benner, T., Salat, D. H., & Fischl, B. (2008). Brain morphometry with multiecho MPRAGE. *NeuroImage*, 40, 559–569.
- Vazquez, A. L., & Noll, D. C. (1998). Nonlinear aspects of the BOLD response in functional MRI. *NeuroImage*, 7, 108–118.
- Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, 4, 274–290.
- Wade, A. R., & Rowland, J. (2010). Early suppressive mechanisms and the negative blood oxygenation level-dependent response in human visual cortex. *The Journal of Neuroscience*, 30, 5008–5019.
- Wald, L. L., & Polimeni, J. R. (2017). Impacting the effect of fMRI noise through hardware and acquisition choices — Implications for controlling false positive rates. *NeuroImage*, 154, 15–22.
- Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage*, 14, 1370–1386.
- Wu, G. R., & Marinazzo, D. (2016). Sensitivity of the resting-state hemodynamic response function estimation to autonomic nervous system fluctuations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374, 20150190.
- Yacoub, E., & Hu, X. (1999). Detection of the early negative response in fMRI at 1.5 tesla. *Magnetic Resonance in Medicine*, 41, 1088–1092.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Taylor, A. J., Kim, J. H., & Ress, D. (2022). Temporal stability of the hemodynamic response function across the majority of human cerebral cortex. *Human Brain Mapping*, 43(16), 4924–4942. <https://doi.org/10.1002/hbm.26047>