



# Early Stopping in Experimentation With Real-Time Functional Magnetic Resonance Imaging Using a Modified Sequential Probability Ratio Test

Sarah J. A. Carr<sup>1,2</sup>, Weicong Chen<sup>3</sup>, Jeremy Fondran<sup>4</sup>, Harry Friel<sup>5</sup>, Javier Sanchez-Gonzalez<sup>6</sup>, Jing Zhang<sup>4</sup> and Curtis Tatsuoka<sup>2,4\*</sup>

<sup>1</sup> Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom, <sup>2</sup> Department of Neurology, Case Western Reserve University, Cleveland, OH, United States, <sup>3</sup> Department of Computer and Data Sciences, Case Western Reserve University, Cleveland, OH, United States, <sup>4</sup> Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, United States, <sup>5</sup> Philips Healthcare, Highland Heights, OH, United States, <sup>6</sup> Philips Healthcare, Madrid, Spain

## OPEN ACCESS

### Edited by:

Nikolaus Weiskopf,  
Max Planck Institute for Human  
Cognitive and Brain Sciences,  
Germany

### Reviewed by:

Michael Lührs,  
Maastricht University, Netherlands  
Amelie Haugg,  
Psychiatric University Hospital Zurich,  
Switzerland  
Joram Soch,  
Bernstein Center for Computational  
Neuroscience, Germany

### \*Correspondence:

Curtis Tatsuoka  
cmt66@case.edu

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 18 December 2020

**Accepted:** 13 October 2021

**Published:** 04 November 2021

### Citation:

Carr SJA, Chen W, Fondran J,  
Friel H, Sanchez-Gonzalez J, Zhang J  
and Tatsuoka C (2021) Early Stopping  
in Experimentation With Real-Time  
Functional Magnetic Resonance  
Imaging Using a Modified Sequential  
Probability Ratio Test.  
*Front. Neurosci.* 15:643740.  
doi: 10.3389/fnins.2021.643740

**Introduction:** Functional magnetic resonance imaging (fMRI) often involves long scanning durations to ensure the associated brain activity can be detected. However, excessive experimentation can lead to many undesirable effects, such as from learning and/or fatigue effects, discomfort for the subject, excessive motion artifacts and loss of sustained attention on task. Overly long experimentation can thus have a detrimental effect on signal quality and accurate voxel activation detection. Here, we propose dynamic experimentation with real-time fMRI using a novel statistically driven approach that invokes early stopping when sufficient statistical evidence for assessing the task-related activation is observed.

**Methods:** Voxel-level sequential probability ratio test (SPRT) statistics based on general linear models (GLMs) were implemented on fMRI scans of a mathematical 1-back task from 12 healthy teenage subjects and 11 teenage subjects born extremely preterm (EPT). This approach is based on likelihood ratios and allows for systematic early stopping based on target statistical error thresholds. We adopt a two-stage estimation approach that allows for accurate estimates of GLM parameters before stopping is considered. Early stopping performance is reported for different first stage lengths, and activation results are compared with full durations. Finally, group comparisons are conducted with both early stopped and full duration scan data. Numerical parallelization was employed to facilitate completion of computations involving a new scan within every repetition time (TR).

**Results:** Use of SPRT demonstrates the feasibility and efficiency gains of automated early stopping, with comparable activation detection as with full protocols. Dynamic stopping of stimulus administration was achieved in around half of subjects, with typical time savings of up to 33% (4 min on a 12 min scan). A group analysis produced similar

patterns of activity for control subjects between early stopping and full duration scans. The EPT group, individually, demonstrated more variability in location and extent of the activations compared to the normal term control group. This was apparent in the EPT group results, reflected by fewer and smaller clusters.

**Conclusion:** A systematic statistical approach for early stopping with real-time fMRI experimentation has been implemented. This dynamic approach has promise for reducing subject burden and fatigue effects.

**Keywords:** real-time fMRI, adaptive fMRI, dynamic experimentation, SPRT (sequential probability ratio test), early stopping fMRI

## INTRODUCTION

Analysis of task-based fMRI scans is typically performed with fixed, predetermined experimental designs. As a result, subjects must often endure stimulus protocols that are overly long in order to ensure the neural activity can be statistically discerned in the noisy data. However, this can lead to fatigue, learning effects and excessive motion, such as from agitation, as well as being costlier to administer due to longer scan times and potentially less reliable measurement. Also, the experimenter does not know if the neural activity is detectable until long after the scanning session is over. Real-time functional MRI (RT-fMRI) provides an opportunity to ameliorate these issues. RT-fMRI has been successfully applied in the field of neurofeedback and biofeedback from neural responses, where subjects may be trained to alter their brain activity based on real-time information provided from the fMRI scans. This has been reported in ADHD (Alegria et al., 2017), healthy subjects with no psychiatric or neurological disorders (Lawrence et al., 2014; Sherwood et al., 2016), Alzheimer's disease (Hohenfeld et al., 2017) and Parkinson's disease (Subramanian et al., 2011, 2016). Its uses have also been described in psychoradiology to aid diagnosis and treatment planning in psychiatric disorders (Lui et al., 2016). Real-time resting state fMRI has for instance been studied and implemented as well using TurboFIRE (Vakamudi et al., 2020). A largely unexplored application of RT-fMRI is to dynamically and statistically determine when a stimulus has been sufficiently presented in terms of replication of blocks to terminate early. The magnitude of effort and variability in neural activity while completing a task will vary from person to person. Trial administration within a block design can be stopped early if sequentially updated statistical inference on activation can be determined with sufficient accuracy based on the observed BOLD (blood oxygen level dependent) signal response up to that point. This application will be explored in detail.

The benefits of adaptive RT-fMRI include: (1) **Shorter scan times for fMRI testing:** Shorter scan times cannot only save in technology and personnel costs, but fatigue and learning effects can be avoided, improving signal quality. Scanning becomes less burdensome on the subject as well, which is an especially important consideration for children or elderly subjects. (2) **Real-time quality control:** greater consistency in activation classification error can be obtained, through statistical error-based benchmarks for stopping rules and real-time

feedback on classification performance and adjustment of stimulus durations. (3) **Richer information:** Paradigms can become more complex and sophisticated. With greater time efficiency and flexibility, more variations of a stimulus, such as reflected by a broader range of difficulty levels, can be administered in the same amount of time. (4) **Wide applicability:** Dynamic adjustment of stimuli based on BOLD response in real-time can be generally applied across a range of focus areas that investigate localization of brain activity, including cognition and motor functioning.

Since the advent of RT-fMRI in the mid 1990s (Cox et al., 1995), a handful of mainstream software packages have been developed for use by the fMRI community. These include Turbo BrainVoyager (Goebel, 2012), AFNI's real-time plugin (Cox et al., 1995), FSL-based FRIEND (Sato et al., 2013) and Python-based OpenNFT (Koush et al., 2017). There have been a few previous studies that have employed adaptive task-based RT-fMRI. It has been used to determine 'good' and 'bad' brain states to optimize learning (Yoo et al., 2012); to determine a person's brain state to judge their attention to a task (Debettencourt et al., 2015); to elicit activity in particular brain regions by presenting stimuli chosen based on the response to the previous stimulus (Lorenz et al., 2016); and to estimate when brain activity was mapped to a particular network (Lorenz et al., 2018). Each study used different methods to perform the real-time analysis. These include real-time general linear model (GLM) methods (Yoo et al., 2012; Lorenz et al., 2016), multivariate pattern analysis (Debettencourt et al., 2015) and a Bayesian optimization algorithm (Lorenz et al., 2018).

Here, we extend the use of a statistically based dynamic approach to RT-fMRI experimentation described in Feng et al. (2015), addressing issues related to practical implementation. This approach involves the sequential updating of voxel-level likelihood ratio tests, known as sequential probability ratio tests (SPRTs) and assessing after each scan whether there is sufficient statistical evidence to determine whether or not an associated parameter value indicates task activation. Such results, considered in aggregate across a collection of voxels, can be used as a basis for early stopping of experimentation. Most off-line, *post-hoc* analyses of fMRI data use the general linear model to test statistical associations of voxel activation magnitude to task administration (Cox, 1996; Lazar, 2008; Jenkinson et al., 2012). This approach involves the voxel-level estimation of task-related regression parameters that indicate magnitude of association

between an expected hemodynamic response signal from a task and the observed BOLD signal. We have adapted this general method for real-time fMRI by incrementally updating GLM regression parameter estimates as soon as the brain volumes are collected. At the individual voxel level, we can then assess hypothesis tests related to activation that are based on these estimates. In aggregate, the voxel level analyses inform decisions on early stopping and the tailoring of fMRI experimentation (Feng et al., 2015).

In comparison to Feng et al. (2015), we adopt a two-stage estimation approach that allows for parameter values that represent activation thresholds to be formulated in terms of  $z$ -score scale at the voxel level. This adaptive specification avoids the intractable problem of pre-specifying magnitudes of GLM parameter values that would be considered as “active.” Such magnitudes need to be scaled relative to estimation variance, which should be stably estimated after a first stage. We determine an appropriate duration of the first stage by monitoring estimation convergence of key GLM parameters. Also, while in Feng et al. (2015) serial independence was assumed, here we use the “sandwich” estimator to recognize potential serial covariance in inference (Carroll et al., 1998; Kauermann and Carroll, 2001). The impact of early stopping on group analysis is considered here as well. Importantly, we now present a novel workflow to apply and implement these methods on a Philips scanner, with a dynamic feedback system that allows for real-time dynamic adjustment of the experimentation with subjects. This was facilitated with adoption of numerical parallelization techniques. This work supports the premise that adaptive, individualized experimentation is feasible and can lead to practical and useful savings in scan times by reducing experimental redundancy.

Another novel aspect of this work is the application of adaptive RT-fMRI in a sample group of 12 healthy adolescent subjects and 11 adolescents born extremely preterm (EPT). The fMRI stimulus was a mathematical version of the well-known 1-back task. Early stopping was implemented using sequential probability ratio test (SPRT) statistics and our server was a Linux workstation located in a nearby building. Processing of RT-fMRI was completed within 3 s, always before the next scan arrived. We observed time savings of up to 33% based on early stopping when 80% of voxels were classified, which equals up to 4-min savings with a 12-min scan. The impact on activation analysis from the selection of early stopping criteria is assessed, as described in detail below. Finally, as an illustration, we conduct a comparison of group analyses between difficulty levels for EPT and healthy control subjects, to assess the effects of early stopping in this context.

## BACKGROUND INFORMATION

### General Linear Model

Briefly, the general linear model involves convolving a double gamma hemodynamic response function (HRF) with task indicator variables that denote timing of administration to reflect

expected task-related BOLD responses. Voxel-level task-related regression parameters are estimated and represent the association of the observed response to expected task-activated BOLD signal. Thus, activation is assessed through statistical inference on regression parameters. For a given voxel up to time  $t$  (i.e., for scans 1 through  $t$ ), the GLM takes the form:

$$Y_t = X_t\beta + E_t \quad (1.1)$$

where  $Y_t$  is a  $t \times 1$  vector of observed BOLD signal intensities for the voxel up to time  $t$ ,  $E_t$  is a  $t \times 1$  vector that represents the error terms,  $X_t$  is a  $t \times p$  design matrix and includes the expected BOLD signal values per condition and  $\beta = [b_1 \dots b_j \dots b_p]'$  is a  $p \times 1$  regression coefficients vector. In this formulation, a regression parameter  $b_j$  can represent magnitude of association with condition  $j$ .  $E_t$  is assumed to be distributed as multivariate normal with mean zero and covariance  $W_t$ , where  $W_t$  is a  $t \times t$  covariance matrix. Error variances comprise the diagonal elements. For spatial correlation, we conduct spatial smoothing, so do not explicitly model the spatial correlation structure.  $Y_t$  is assumed to have a multivariate normal probability distribution as follows:

$$f(Y_t, \beta, W_t) = \frac{1}{(2\pi)^{t/2} |W_t|} \exp\left(-\frac{1}{2}(Y_t - X_t\beta)' W_t^{-1} (Y_t - X_t\beta)\right) \quad (1.2)$$

where  $|W_t|$  is the determinant of  $W_t$ . At each voxel, the focus is on estimation of  $c\beta$ , where  $c$  is some  $1 \times p$  linear contrast. Hence, we consider the more general scenario when inferential interest can also be a linear combination of the task parameters, which includes the single task parameter case. The standard linear estimate for  $c\beta$  at time  $t$  is:

$$c\hat{\beta}_t = c(X_t'X_t)^{-1} X_t'Y_t \quad (1.3)$$

This estimator is normally distributed and unbiased under (1.2). We fit regression models in parallel for all voxels under consideration in a target region of interest (ROI), which could be a particular structure or include the whole brain. Real-time analysis requires signal and image processing steps, as well as the continual updating of statistical estimates as new scan data are received from the scanner. Hence, given the large number of voxels to be analyzed, real-time fMRI presents “big data” computational challenges.

### Sandwich Estimator

In our previous work (Feng et al., 2015), we assumed serial independence for computational simplicity. Here we recognize potential for more general correlation structure using the nonparametric sandwich estimator  $\widehat{\text{var}}[c\hat{\beta}_t]$  (Carroll et al., 1998; Kauermann and Carroll, 2001). The sandwich estimator is a robust, model-free variance estimator for  $c\hat{\beta}_t$  that does not require assumptions such as homoscedastic error variance or serial independence. Importantly, it still provides asymptotically consistent variance estimates, although convergence rates can

be slow (Carroll et al., 1998; Kauermann and Carroll, 2001). The approach is computationally feasible for real-time analysis. This is in contrast to our initial autoregressive modeling approaches for the covariance matrix  $W$ , which were too slow to estimate for real-time implementations given our available computational resources.

## Sequential Probability Ratio Test

At the voxel level, we can use the sequential analytic framework of Wald (1947); Wald and Wolfowitz (1948); Cox (1963); Tantaratana and Thomas (1977); Li (2010); Feng et al. (2015), to adaptively assess activation status using real-time fMRI. As we will demonstrate, Wald's SPRT test statistic can serve as the basis of an efficient, incremental GLM testing approach that can greatly reduce the need for experimental block administrations compared with fixed designs while attaining similar classification performance in simulation, and activation patterns with subject data. This approach relies on a SPRT statistic to conduct hypothesis testing, with the null hypothesis representing no activation with respect to a task, and the alternative hypothesis representing some threshold of activation, as represented by a GLM parameter value (Feng et al., 2015). This statistic is updated with each new observation, and its value is compared with thresholds for stopping.

The general procedure of Wald's SPRT is as follows. Consider a one-sided hypothesis of the form  $H_0 : c\beta = c\beta_0$  versus  $H_a : c\beta \geq c\beta_1$ , where  $c(\beta_1 - \beta_0) > 0$ . Two-sided formulations are described in Wald (1947) and Feng et al. (2015). Implementation of Wald's SPRT involves updating Wald's likelihood ratio statistic as new data are observed (Wald, 1947):

$$\Lambda_t = \log \left( \frac{f(\widehat{c\beta}_t | c\beta_1, \widehat{\text{var}}[c\widehat{\beta}_t])}{f(\widehat{c\beta}_t | c\beta_0, \widehat{\text{var}}[c\widehat{\beta}_t])} \right) \quad (1.4)$$

where  $f(\widehat{c\beta}_t | c\beta_0, \widehat{\text{var}}[c\widehat{\beta}_t])$  and  $f(\widehat{c\beta}_t | c\beta_1, \widehat{\text{var}}[c\widehat{\beta}_t])$  are the respective normal probability density functions of  $c\widehat{\beta}_t$  given  $c\beta_0$  or  $c\beta_1$  is the true value of the parameter of interest and conditioning on the estimated variance. After  $Y_t$  is observed up to time point,  $t$ , one of three possible decisions is made according to the following rules:

1. Continue sampling, if  $B < \Lambda_t < A$
2. Stop sampling and accept  $H_0$ , if  $\Lambda_t < B$
3. Stop sampling and accept  $H_a$ , if  $A < \Lambda_t$

where stopping boundaries are:

$$(A, B) = \left( \log \left( \frac{1 - \beta_E}{\alpha_E} \right), \log \left( \frac{\beta_E}{1 - \alpha_E} \right) \right) \quad (1.5)$$

and the target Type I and Type II error levels are, respectively, denoted as  $\alpha_E$  and  $\beta_E$ . Note that both the Type I and Type II error levels are controlled for with SPRT, as opposed to standard hypothesis test formulations that only control for Type I error level. Multiple SPRTs are conducted concurrently across voxels. This can in principle be adjusted for by Bonferroni correction to account for this simultaneous testing. If the specified Type I error levels are too small, however, this can affect the feasibility of early

stopping. Although not done here, it is possible to also consider *post-hoc* multiple comparisons methods after experimentation is completed (Lindquist and Mejia, 2015).

A practical modification of the original SPRT formulation for stopping is to consider the truncated SPRT (Sawasd Tantaratana, 1977), which will additionally call for stopping if an upper bound for the number of observations is reached. In our case, this is reached when the upper limit of blocks have been administered. Additional modifications include conducting two-stage estimation to allow sufficient observation for preliminary estimates of the voxel-level error variance from a first stage where stopping is not yet considered (Hall, 1962). With these estimates, we can derive an alternative hypothesis value for a linear contrast of task parameters  $c\beta$  that will correspond to a desired  $z$ -statistic value, denoted as  $z_t$ . As an illustration, suppose a  $z$ -statistic threshold value of 3.10 is selected, as will be done below in our studies. (Note  $z_t = 3.10$  is associated with the one-sided  $p$ -value = 0.001). Given an estimated value  $\widehat{\text{var}}[c\widehat{\beta}_t]$  from a first stage of length  $t$  scans, we solve for the value of  $\theta_1 = c\beta_1$  that satisfies:

$$\frac{\theta_1}{\sqrt{\widehat{\text{var}}[c\widehat{\beta}_t]}} = z_t \quad (1.6)$$

This value becomes the alternative hypothesis, and it represents the voxel-level targeted activation magnitude threshold. For subsequent scans we fix the value of  $\theta_1$  as the alternative hypothesis value for  $c\beta$ , so that the activation magnitude threshold in the alternative hypothesis is held constant. Note that  $z_t$  will thus tend to increase across scans, since  $\widehat{\text{var}}[c\widehat{\beta}_t]$  decreases with more data. This approach allows for more comparable activation patterns across different scan lengths, since respective activations reflect similar magnitudes.

Ultimately, we aggregate the findings of the voxel-level SPRTs to determine whether or not experimentation within a block design should be terminated early. A "global" stopping rule that considers all voxels in a region of interest (can be whole brain or smaller ROIs) that we have adopted is to terminate task administration when a predetermined percentage of voxels have been classified by their respective SPRTs. For instance, we have used 80% as a global stopping criterion. Note that 80% classified means either as active or non-active. We choose this cut-off as it is fairly strict, and yet approximately one half of the participants still stop early. As we will see, it also facilitates correspondence with full scan data results, particularly if the activation threshold is adjusted to recognize longer scan durations. We also consider other global stopping criterion here, 70 and 90%, and assess impact on stopping times and resultant images arising from early stopping. We also choose the SPRT target Type I and Type II error levels that are relatively more stringent for Type I error. Note that for  $c\beta$  parameter values that are "in-between" the null and alternative hypothesis values, the SPRT is indifferent to preferring one hypothesis over the other. This leads to larger numbers of scans needed before a stopping boundary is crossed. So, we have to accept a lack of decisive stopping decisions for these cases in order for overall experimentation to stop early. This can be an acceptable trade-off for shorter experimental scan times and the ability to tailor experimentation.

## METHODOLOGY

### Participants

Twelve healthy subjects were recruited, 7 males. They were aged 15–16 years old and 11 were right-handed. They had no known neurological conditions and a normal developmental history. A group of 11 adolescents born EPT were also recruited, 1 male. EPT was defined as being born at <26-week gestation and weighing < 1000 g. All were aged 15–17 years old and 8 were right-handed, 2 left-handed and 1 ambidextrous. All subjects were recruited as part of a larger study to evaluate functional and structural differences associated with mathematical abilities and working memory between those born EPT and those born at normal term. The aim of the larger study is to improve our understanding of mathematics disabilities and potentially lead to improvements in pedagogical practices for young people experiencing problems acquiring mathematics skills. Adolescents were recruited as they can handle the stress of fMRI experimentation, are mathematically advanced enough and have had time to master the subject area compared to younger children. This age range is also an advantageous time to implement interventions to improve mathematical abilities before leaving school, hence adults were not studied. EPT subjects were included to show that differences with patient populations are detectable with our methods. A subsection of the full study is reported here to demonstrate the real-time analysis.

The subjects made one two-h visit to the MRI department at University Hospitals Cleveland Medical Center (UHCMC). Ethics approval was obtained from the UHCMC Institutional Review Board office prior to the study and complied with the Declaration of Helsinki for human subject research. Subjects and their parents gave informed consent prior to taking part.

As part of our wider study, subjects also made another, separate 3-h visit to the study offices to undergo neuropsychological testing and a refresher of fraction calculations. In the interests of brevity, the full neuropsychological testing results are not reported here. One finding that is particularly relevant to the fMRI task considered here is that nearly two thirds (63.6%) of the EPT cohort have lower working memory function, compared to just over one third (35.7%) of controls subjects. Working memory level was determined by sample quartiles of standardized z-scores based on the Wide Range Assessment of Memory and Learning test (WRMAL<sup>1</sup>, Sheslow and Adams, 2003). We denote “lower” working memory function as having a standardized score that is less than the sample median.

### MRI Protocols

The subjects were positioned head-first supine on the scanner bed with their head fixed in position using inflatable pads. An 8-channel head coil was used for data acquisition. Echo planar imaging scans were acquired on a Philips Ingenuity 3T PET/MR imager at UHCMC. The following fMRI scan parameters were used: TR = 3.0 s, TE = 35 ms, in-plane resolution was 1.797 mm<sup>2</sup> (matrix 128 × 128), slice thickness was 4 mm, number of

slices = 36 slices and flip angle = 90°. A SENSE P reduction factor of 2 was implemented and scans were acquired in an ascending interleaved fashion.

In addition to the fMRI scans, a high-resolution T1-weighted anatomical image of the brain was also acquired. This was taken using a magnetic preparation gradient-echo sequence (3D IR TFE). Imaging parameters were: TR = 7.5 ms, TE = 3.7 ms, in-plane resolution was 1 mm<sup>2</sup> (matrix 256 × 256), slice thickness was 1 mm, number of slices = 200 slices and flip angle = 8°.

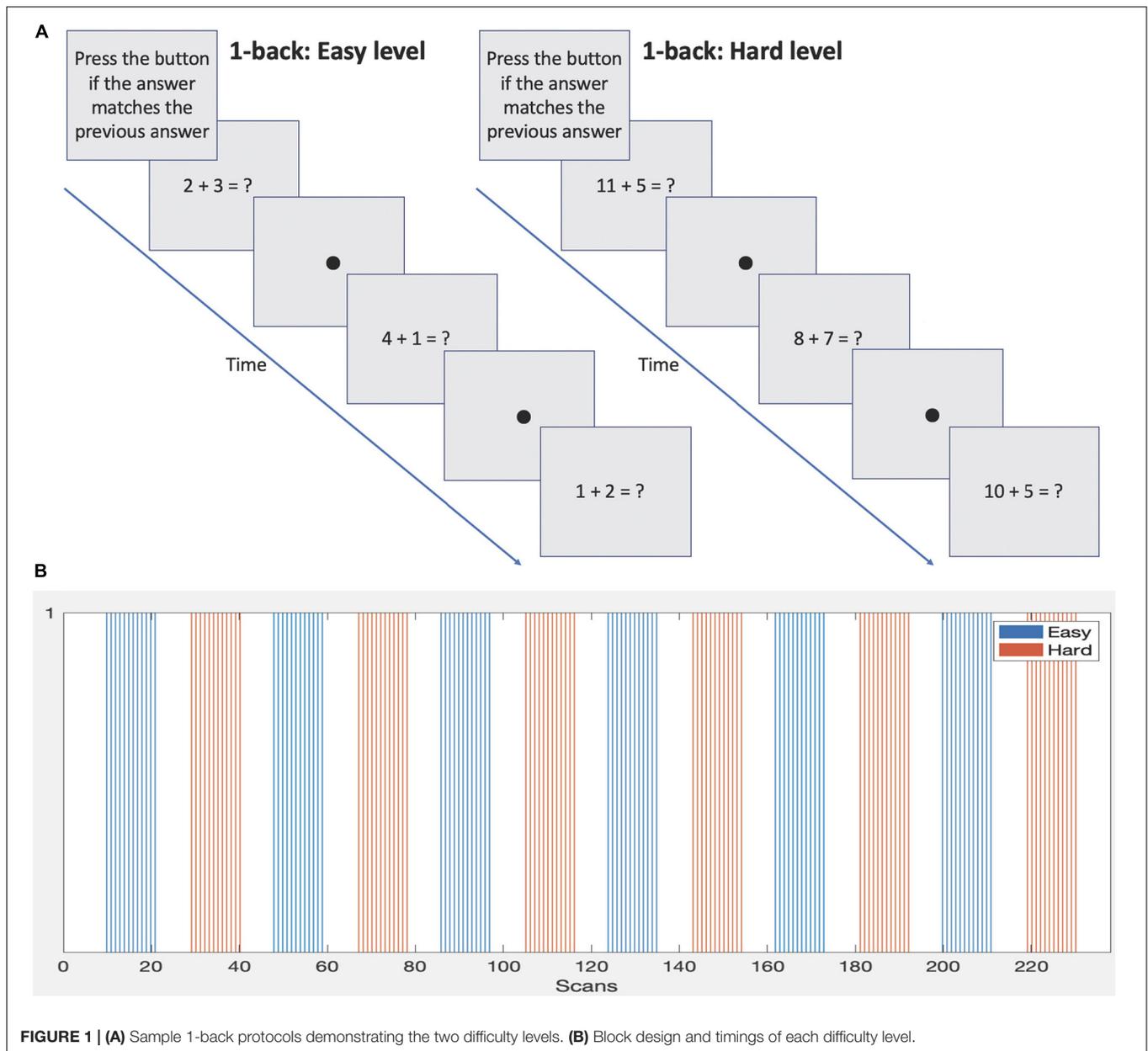
### Stimulus Protocols

During data acquisition subjects were presented with a mathematical version of the well-known 1-back memory task. It involved performing basic addition and subtraction calculations and required the answer to be remembered and compared to the next answer. Two difficulty levels were included. The protocol was developed by our lab as part of a battery to assess mathematical and working memory abilities in 14–17 year olds to evaluate the functional differences between those born EPT and those born at normal term. The stimulus was presented on an MRI compatible LCD monitor (manufactured by Cambridge Research Systems, Rochester, United Kingdom) positioned at the end of the bore and viewed via a mirror attached to the head coil. Equations were presented, for example, the subject may see “2 + 3 = ?.” The subject was required to work out the answer and then remember it while working out the next equation, for example “1 + 4 = ?.” If they thought the answers matched, then the subject pressed a button on a response box held in their right hand. If they thought the answers did not match, then they did nothing but remember the new answer to compare to the answer of the next equation. An example sequence is shown in **Figure 1A**.

The stimulus was presented in a block design, see **Figure 1B** and **Table 1**. Eight equations were presented per block. Each block lasted 36 s followed by 21 s of rest condition (fixation dot). Two difficulty levels were presented. The easier level consisted of single digit numbers to add or subtract and the answers were always a single digit. The harder level involved addition or subtraction of single or two-digit numbers and the answers were always two digits. Blocks of difficulty levels were alternated during the scan and a total of 6 blocks per level were presented. Note: although only 2 difficulty levels are used here, the setup is able to accommodate any number of difficulty levels. The full duration of the task was 238 scans or 11 min and 54 s. This was based on a moderate length of experimentation for a 1-back block design (e.g., see Daamen et al., 2015; Poudel et al., 2015; Di et al., 2020; Huang et al., 2020), allowing approximately 6 min for each difficulty level.

Two difficulty levels were included to investigate differences in neural responses associated with increasing task demand. As the brain is ‘pushed’ to solve more complex problems, differential networks may be apparent, and these may be different between normal term and EPT subjects. Additionally, increasing the difficulty level serves to maintain the subject’s attention and, generally, increases their effort. This can have the effect of increasing brain activation cluster sizes and magnitude as well as causing recruitment of additional areas, which is of

<sup>1</sup>www.proedinc.com



interest. Incorporating difficulty levels into protocols that can separately be terminated early demonstrates the flexibility of the proposed approach.

**TABLE 1 |** The scan number when each stimulus block is completed.

Block	Easy level	Hard level
1	21	40
2	59	78
3	97	116
4	135	154
5	173	192
6	211	230

## Real-Time Functional Magnetic Resonance Imaging Acquisition

The visual stimulus was presented using an in-house custom written program that was developed using the Python programming language (Python Software Foundation<sup>2</sup>) and libraries from PsychoPy - an open source visual presentation program (Peirce, 2007, 2009; Peirce and MacAskill, 2018). The program connected to a Cedrus Lumina controller to receive stimulus responses from the subject and trigger pulses from the MRI scanner (outputted every dynamic). The timing of the presentation of the visual stimulus was synchronized to the trigger pulses to ensure that stimulus images were displayed

<sup>2</sup><https://www.python.org/>

at the expected time. A Supervisor Window displayed on the experimenter's computer screen allowed the visual stimulus to be tracked throughout. It displayed the current block number being presented, how many remaining blocks there were and when the subject responded. The program was also able to terminate one or both of the difficulty levels if it received a signal indicating the relevant areas in the fMRI data were sufficiently classified across voxels. The program offered the flexibility to automatically terminate the presentation or the experimenter could override the termination instruction and continue stimulus presentation. The software is freely available from the Bitbucket repository<sup>3</sup>.

Real-time image transfer was achieved by XTC (eXternal Control). This is a program integrated into the Philips scanner software and enabled by a research clinical science key. XTC communicates with the reconstruction and scanner processes on the scanner computer and interfaces to a network client application using a minimalistic CORBA (Common Object Request Broker Architecture) (Schmidt et al., 1998) interface which uses TCP/IP as the transport layer. CORBA is platform independent, reliable, and has the ability to process large amounts of data with minimum overhead. Each CORBA message consisted of a hierarchical attribute collection identified with UUIDs (universally unique identifiers) (ITU-T, 2005). Messages carried reconstructed image data and meta-data containing details of scan protocols. Due to hospital network security protocols the reconstructed images were placed in a folder on the scanner computer and then pushed across to a Linux computer. To achieve necessary image transfer speeds to the scanner computer folder a modification to XTC was installed on the scanner to disable two-way communications as only one-way image transfer functionality was required. However, XTC does support two-way communication between the scanner and the Client.

The Linux computer was a custom-built server equipped with a solid state hard drive and two 8-core Intel Xeon E5-2687W processors running at 3.1 GHz and providing 40 MB L3 cache. It was installed with Centos 7.4 operating system. As the scans were received, custom written Python and Bash scripts implemented the analysis using core-based parallelization to preprocess the data and perform the SPRT statistical analysis. Preprocessing was performed using standard modules from AFNI (Analysis of Functional NeuroImages<sup>4</sup>) and FSL (FMRIB's Software Library<sup>5</sup>). The analysis sequence is detailed in the following section. The setup is shown in **Figure 2**.

## MRI Preprocessing

At the beginning of the scanning session a single fMRI scan (3 s) was acquired and used for coregistration (motion correction) purposes. In preparation, the skull was removed using FSL's Brain Extraction Tool (BET) (Smith, 2002) and a mask of the full brain was created. During the real-time adaptive fMRI scan session, new scans arrived every 3 s and were dumped in a folder on the Linux workstation where the following actions were applied to

each one. AFNI's 'dcm2niix\_afni' command was used to convert the .par/.rec files to NIFTI. Motion correction was performed using coregistration techniques. Every fMRI scan was realigned to the initial scan that was acquired before the task began, and AFNI's '3dvolreg' command was used. Spatial smoothing was also applied using an 8 mm kernel with AFNI's '3dmerge' command. The full brain mask created at the beginning of the session was applied using FSL's 'fslmaths' command to remove noisy voxels outside the brain (voxels of no interest). CSF and white matter regions were included in the full brain mask. This was due to time constraints at the beginning of the scanning session that did not allow for segmentation of the brain. However, it is feasible to create a ROI (full brain/segmented brain/particular structure) prior to the start of the scan session. If a pre-existing scan of the subject is not available for this purpose then the ROI can be created using a template brain (e.g., MNI) and morphed to the subject's initial single fMRI scan. This process was tested as a part of our study and is feasible to conduct, although not used here.

The resulting preprocessed images were then converted to ASCII text files for statistical analysis with SPRT producing one .txt file per fMRI scan. The text file contained the grayscale values for every voxel in a 3D image. The conversion was accomplished using 'fsl2ascii' command.

## Functional Magnetic Resonance Imaging Sequential Probability Ratio Test Analysis

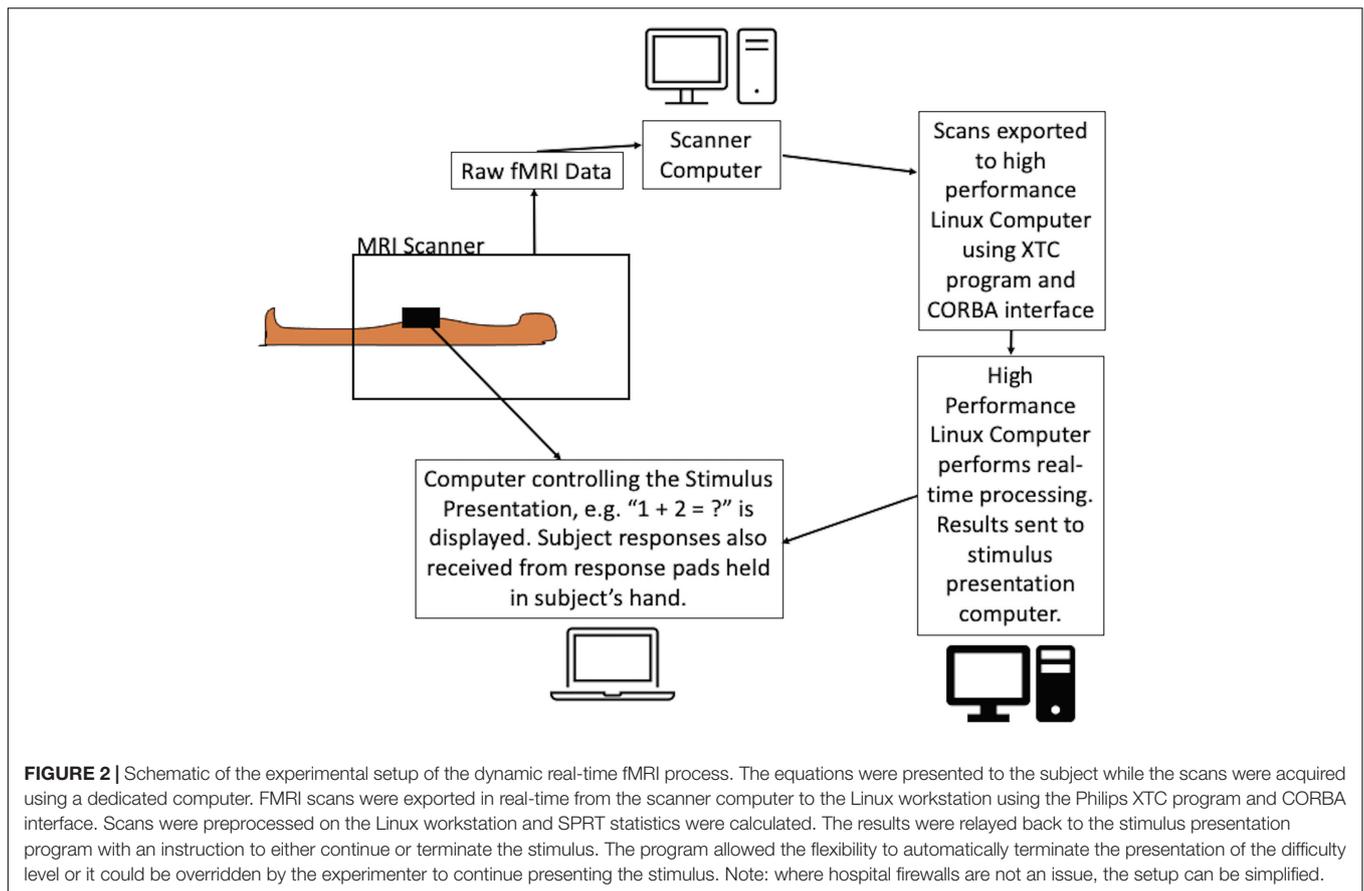
The SPRT analysis was applied using a highly optimized C++ program that used Intel Cilk Plus library for multicore and vector processing of data. BLAS routines from Intel MKL were used to enable instruction-based acceleration for matrix computation. They are available from the Bitbucket repository at <https://bitbucket.org/tatsuoka-lab>.

The design matrix was created prior to the scan session and contained 8 columns. One column of '1s' is for the intercept. Two columns represented the easy and hard stimuli. They were created using AFNI's '3dDeconvolve' command to model the HRF. Easy and hard levels were modeled separately and the contrasts applied post processing were easy or hard versus rest. We also included in the design matrix cosine functions of increasing periodicity to model major sources of noise in fMRI data. These include brain metabolism, physiology, and spontaneous fluctuations (Friston et al., 2000; Yan et al., 2009). For large periodicities, cosine functions are approximately linear for the time frame of scans we consider here, and hence are essentially collinear from a GLM modeling perspective. Five regressors were thus added to the design matrix and were formulated based on the information given in Kiebel and Holmes (2006). The frequencies were based on total/maximum scan duration and were: maximum scan duration\*2, maximum scan duration, maximum scan duration/1.5, maximum scan duration/2 and maximum scan duration/2.5. It is also possible to include the temporal derivatives of the HRF or other regressors in the design matrix where applicable in studies. Temporal derivatives were not included here due to the long durations of the block design used to present the task. Motion parameters are also frequently

<sup>3</sup><https://bitbucket.org/tatsuoka-lab/fmri-presentation>

<sup>4</sup><https://afni.nimh.nih.gov>

<sup>5</sup><https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>



used as regressors to remove correlated activations produced by movement. Here motion parameter regressors were not included with the estimation of the discrete cosine transforms due to the limitations of the computational resources.

Important design parameters for implementation were selected through analysis of a training sample. For training, each subject underwent the full duration of experimentation. We determined design parameters through the following criteria:

(1) First stage duration: It is desirable for the voxel-level error variance and beta parameter estimates to stabilize – we assessed this qualitatively by assessing plots from a sample of voxels. We tested 2 scenarios using either 2-blocks or 4-blocks of easy and hard stimuli first stage administration before allowing early stopping to occur. Where 2-blocks per difficulty level of stimulus administration were used before allowing early stopping, the first 78 scans were used for the first stage of experimentation. Where 4-blocks were used, 154 scans were used for the first stage. Recall, a full-length task protocol comprised 238 scans and lasted 11 min and 54 s.

(2) Activation threshold: We chose a  $z$ -score threshold of  $z_t = 3.10$  after the first stage since this is a standard threshold value for determining activation of a voxel. We then solved for  $\theta_1$ , as in (1.6). After this first stage, the null and alternative hypotheses at a given voxel are set as  $H_0: c\beta = 0$  and  $H_a: c\beta = \theta_1$ .

(3) Target Type I and Type II error levels: Typical values used in the literature were used to test stopping time performance,

with  $\alpha_E = 0.001$ ,  $\beta_E = 0.1$  (Banerjee et al., 2009; Cremers et al., 2017). We also considered  $\alpha_E = 0.0001$ ,  $\beta_E = 0.1$  and  $\alpha_E = 0.001$ ,  $\beta_E = 0.01$  combinations as well.

(4) Global stopping percentage: We investigate the scenarios of classifying 70, 80, and 90% of voxels. Ideally, early stopping occurs while there is still correspondence in activation patterns to full scan durations, and general concurrence with expected activations from neuroscientific literature.

The stopping rules were defined as described in Section “Sequential Probability Ratio Test”:

1. Continue sampling, if  $B < \Lambda_t < A$
2. Stop sampling and accept  $H_0$ , if  $\Lambda_t < B$
3. Stop sampling and accept  $H_a$ , if  $A < \Lambda_t$

where stopping boundaries are given in (1.4) and are based on target Type I and II error levels. The error levels are specified before testing begins.

Although our goal was to stop scanning early it was necessary to acquire first full length datasets to fully verify our analysis methods and check the feasibility of stopping early. In addition to the subjects reported in the results, we tested early stopping in 2 other subjects to confirm that our procedures worked. Subject details and data are not reported for them as full datasets are not available.

This 1-back arithmetic task involves not only number sense and mathematical calculations but also general cognitive skills involving working memory and sustained attention. The brain networks involved with each of these have been well characterized in the literature and lends itself to the evaluation of this real-time analysis method. There is a large amount of overlap for the active brain areas that control each of these functions and they appear as a frontoparietal network (Menon, 2016a,b; Chai et al., 2018). The areas of the brain we expect to activate in response to each contrast of easy versus rest and hard versus rest are: the intraparietal sulcus, supramarginal gyrus, premotor cortex, dorsal/ventral lateral prefrontal cortex, parietal lobe, Broca's area, occipital lobe, fusiform gyrus, precuneus, cingulate gyrus, anterior insula and frontal eye fields. Assessment of the location and extent of activations within this network will be used as additional criteria for judging appropriate stopping times, in addition to the statistical information determined through the SPRT analysis. This will include how well the cluster peaks coincide with the anatomical locations as well as their extent.

## Group Analysis

There are many possible applications in the research setting where individual level results may be the focus. A possible clinical application may be in clinical assessments for presurgical evaluation for brain surgery in patients with brain cancer or epilepsy. Still, group analyses are commonly conducted and form an essential aspect of fMRI analyses. The outputted results files from the SPRT analysis can be used directly to perform a group analysis using AFNI's 3dMEMA command (Mixed Effects Meta Analysis tool) (Chen et al., 2012), even with the variable scan durations. However, a group analysis was carried out using FSL which instead merges all subject data to conduct a combined mixed models analysis. We demonstrate that the data collected in real-time can still be used in a typical *post-hoc* analysis and can be processed with different parameters to those specified in SPRT. Raw data was preprocessed with FSL FEAT (Woolrich et al., 2001). Motion correction was performed using a rigid body transform, spatial smoothing with a full-width-at-half-maximum Gaussian kernel of 6 mm was applied, high pass temporal filtering of 90 s was carried out and coregistration to (MNI) standard space was done before performing a first level individual GLM analysis. The statistical output from these were used to perform the higher level group statistics using FLAME 1 [FMRIB's Local Analysis of Mixed Effects, (Woolrich et al., 2004)].

## RESULTS

### Individual Subject Results of Sequential Probability Ratio Test

The median control subject response time across both difficulty levels was 1.44 s (SD 0.51 s), and median task accuracy was 90.8% (SD 20.2%). When these are broken down by difficulty level, the easy level median task accuracy was 86.1% (SD 22.6%) with median response time of 1.28 s (SD 0.54 s); and the hard level median task accuracy was 90.0% (SD 18.4%) with median response time of 1.56 s (SD 0.51 s). EPT subjects had a slightly

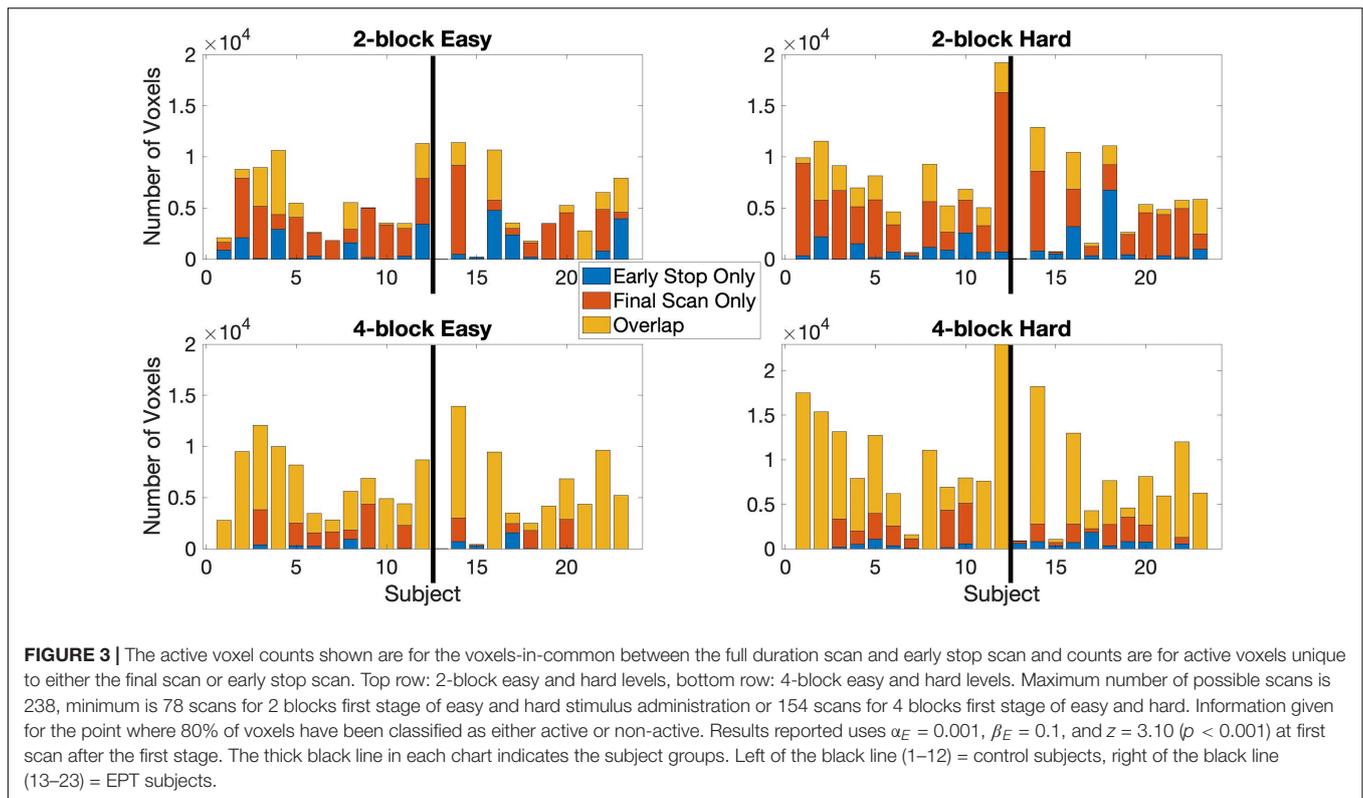
longer overall median response time of 1.91 s (SD 0.48 s) and overall median task accuracy was lower at 65.8% (SD 21.2%). For the easy level, the median accuracy was 72.2% (SD 24.2%) and median response time was 1.63 s (SD 0.49 s). For the hard level the median accuracy was 70.0% (SD 19.8%) with a median response time of 2.10 s (SD 0.54 s). Note that there are statistically significant differences in same subject differences in speed to completion by difficulty level (Wilcoxon signed rank test, two-sided  $p < 0.001$ ). Comparing correctness percentages per subject across birth status groups, there are significant differences with the hard level (Mann-Whitney  $U$ -test two-sided  $p = 0.037$ ), but not with the easy one (two-sided  $p = 0.401$ ). These results indicate that the difficulty levels have different psychometric properties, and affect the groups differently. We also see this in activation patterns, as discussed in Section "Group Analysis Results" and reflected in the group analysis results.

Real-time transfer speeds between the scanner and the single Linux workstation were consistently fast, with individual scan files taking less than 150 milliseconds to transfer. All subject scans were processed within the 3 s TR period. Offline testing showed

**TABLE 2 |** Subject early stop durations of the 1-back task using SPRT.

Subject	Scan when 80% Reached 2-Block Easy	Scan when 80% Reached 2-Block Hard	Scan when 80% Reached 4-Block Easy	Scan when 80% Reached 4-Block Hard
<b>Control</b>				
1	4E/3H	2E/2H	Not reached	Not reached
2	3E/2H	3E/3H	Not reached	Not reached
3	3E/2H	2E/2H	5E/4H	5E/4H
4	4E/3H	3E/2H	Not reached	5E/5H
5	3E/2H	2E/2H	5E/4H	5E/4H
6	2E/2H	3E/2H	5E/5H	6E/5H
7	2E/2H	3E/2H	5E/4H	4E/4H
8	3E/3H	3E/2H	5E/5H	Not reached
9	2E/2H	3E/2H	4E/4H	4E/4H
10	3E/3H	2E/2H	Not reached	4E/4H
11	3E/2H	3E/3H	5E/4H	Not reached
12	3E/3H	3E/2H	Not reached	Not reached
EPT	2E/2H	2E/2H	4E/4H	4E/4H
13	3E/3H	2E/2H	5E/5H	6E/6H
14	3E/3H	3E/2H	4E/4H	5E/4H
15	3E/3H	3E/2H	Not reached	5E/5H
16	2E/2H	2E/2H	5E/4H	6E/5H
17	2E/2H	4E/3H	4E/4H	6E/6H
18	3E/2H	2E/2H	Not reached	4E/4H
19	2E/2H	2E/2H	4E/4H	4E/4H
20	Not reached	2E/2H	Not reached	Not reached
21	3E/3H	3E/2H	Not reached	6E/6H
22	3E/3H	3E/2H	Not reached	Not reached
23	4E/3H	2E/2H	Not reached	Not reached

*Analysis reported here uses  $\alpha_E = 0.001$ ,  $\beta_E = 0.1$ , and  $z = 3.10$  threshold at first scan after the first stage. Maximum number of possible scans is 238, minimum is 78 scans for 2 blocks first stage of easy and hard stimulus administration or 154 scans for 4 blocks first stage of easy and hard. Information given for the point where 80% of voxels have been classified as either active or non-active. E = administration of easy level block, H = administration of hard level block.*



that the subject with the largest number of voxels (subject 21 with 135,379 voxels) required 145 s to conduct all SPRT analyses, starting from scan 79. Less than 1 s is required to analyze the first stage data. The subject with the fewest number of voxels (subject 14 with 77,359 voxels) required 79 s for SPRT analyses across the scans. Overall real-time processing times in this workflow were due in most part to pre-processing computations. For the subject with the largest number of voxels, the maximum time to process one scan in real-time would be about 1.6 s (0.54 s preprocessing time + 0.91 s SPRT calculations + 0.15 s transfer time). This would be even faster with more computational resources, faster transfer speeds or reducing the number of voxels being analyzed.

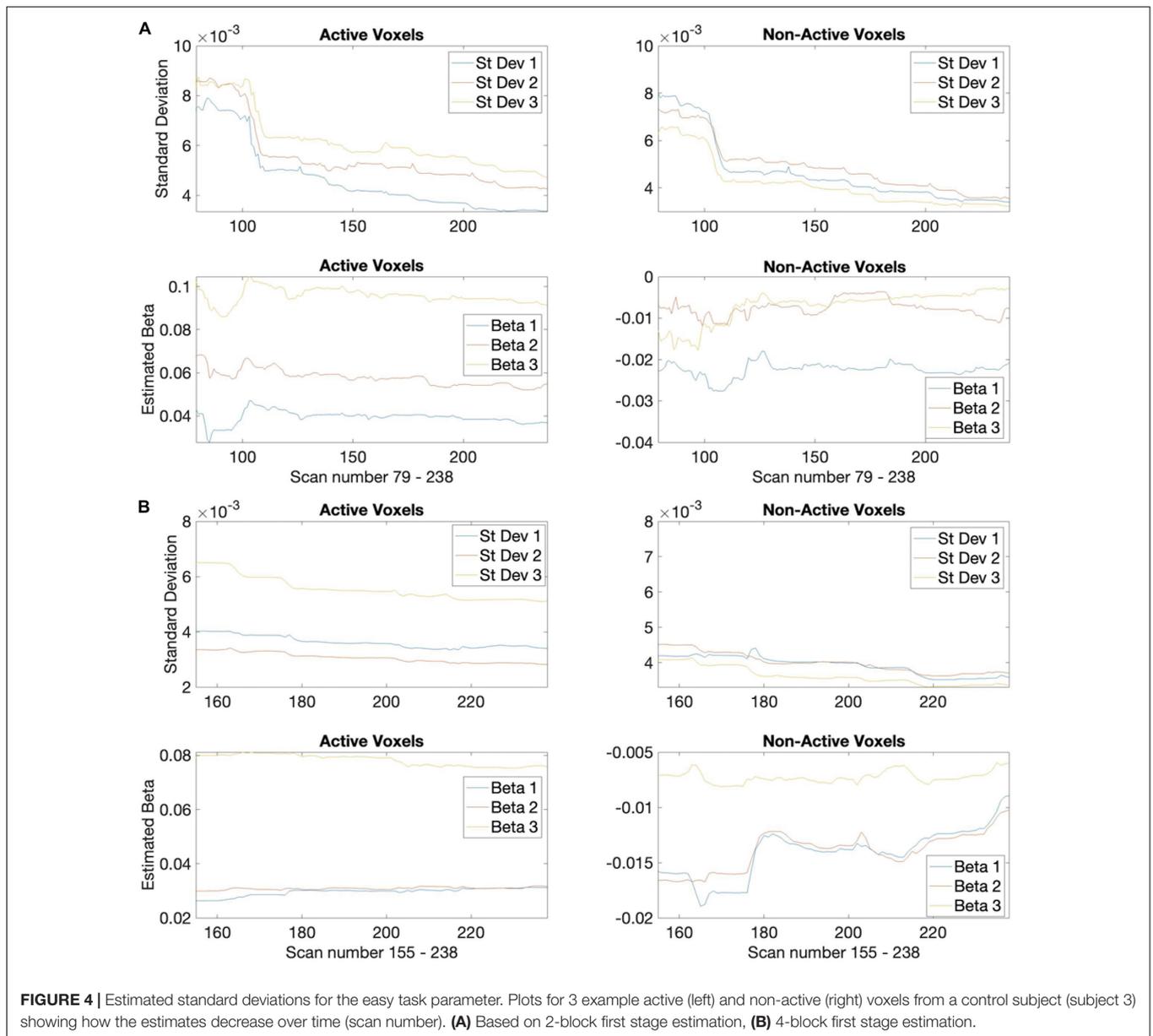
Inspection of the  $z$ -score maps for easy and hard levels for each subject showed that generally, across subjects, the largest activations were centered bilaterally around the inferior and superior parietal areas, taking in the intraparietal sulcus, a region highly associated with mathematical functioning. Further activations were seen in the cuneus. These are most likely correlated with the visual processing associated with the task. Additional activations were seen in the precuneus, bilateral areas in the medial frontal gyrus, anterior cingulate, insula and inferior frontal gyrus. These areas are often associated with attention and memory systems (Cavanna and Trimble, 2006; Menon, 2016a).

The stopping times for the 2- and 4-block first stage lengths for various criteria for each subject are given in Table 2. The spatial overlap between early stopping and full duration maps are explored in Figure 3 where we show the number of active voxels in common between the two durations. The median spatial overlap where early stopping occurs for control subjects was

20.0% (SD 27.0%) for the easy level, 2-blocks and 76.3% (SD 34.5%) for easy level, 4-blocks. For the hard level, there was 30.7% (SD 16.4%) and 77.2% (SD 26.5%) overlap for the 2-block and 4-block first stages, respectively. For EPT subjects the median overlap was 28.7% (SD 35.6%) and 82.7% (SD 34.5%) for the easy level, 2- and 4-block first stages, respectively. For the hard level, the median overlap values were 21.3% (SD 20.6%) and 82.5% (SD 26.5%) for 2-block and 4-block first stages. In Supplementary Table 1, maps of activations for instances where there is less than 50% overlap are presented.

The median spatial overlap for the 2-block first stage is less than 50% for both control and EPT groups. We see for instance in Figure 4, error variance and beta estimates are generally not stable after a 2-block first stage. It is important to “wait” until this happens, as these parameters play a central role in inference and on test statistic values. The 4-block first stage is more attractive in this way. Figure 5 indicates how early stopping is affected by the SPRT Type I error threshold values. Type II error levels seemed to have higher impact. For  $\beta_E = 0.01$ , although not shown, early stopping appears somewhat less often and closer to the end of the experiment.

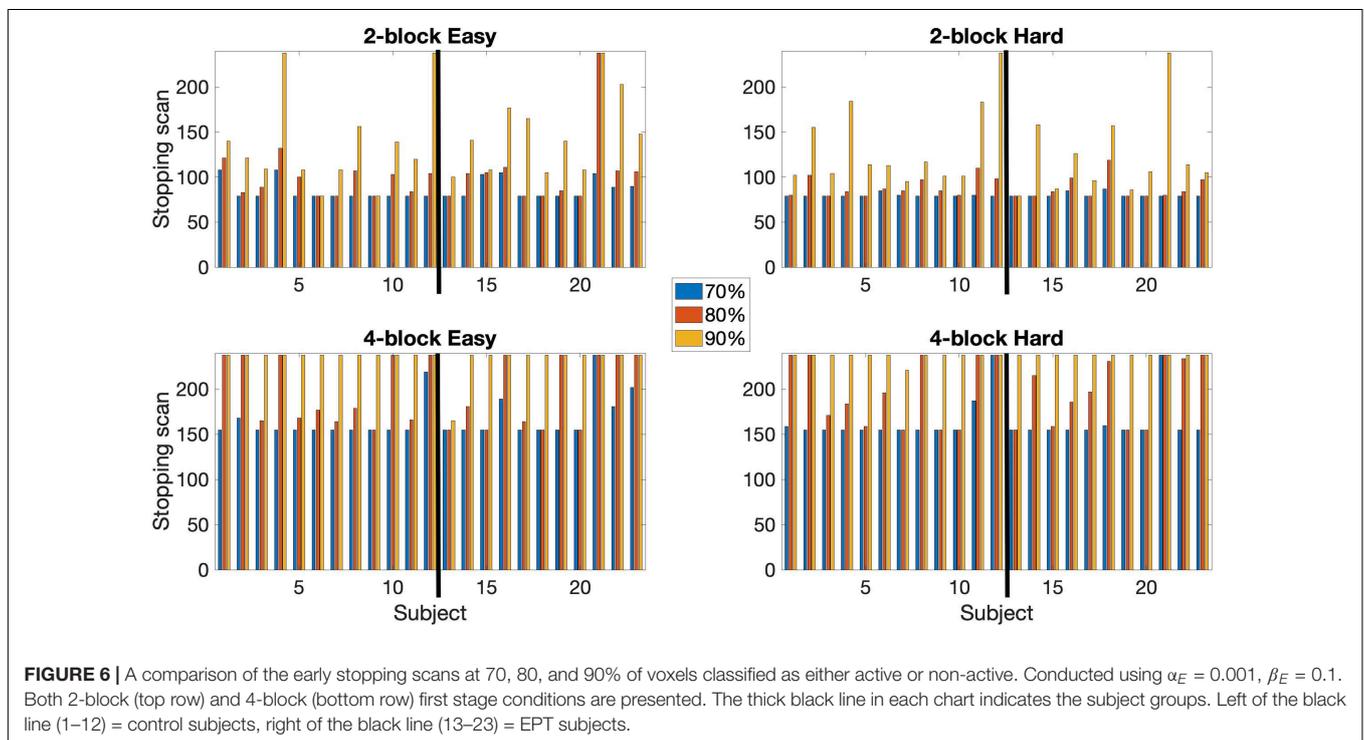
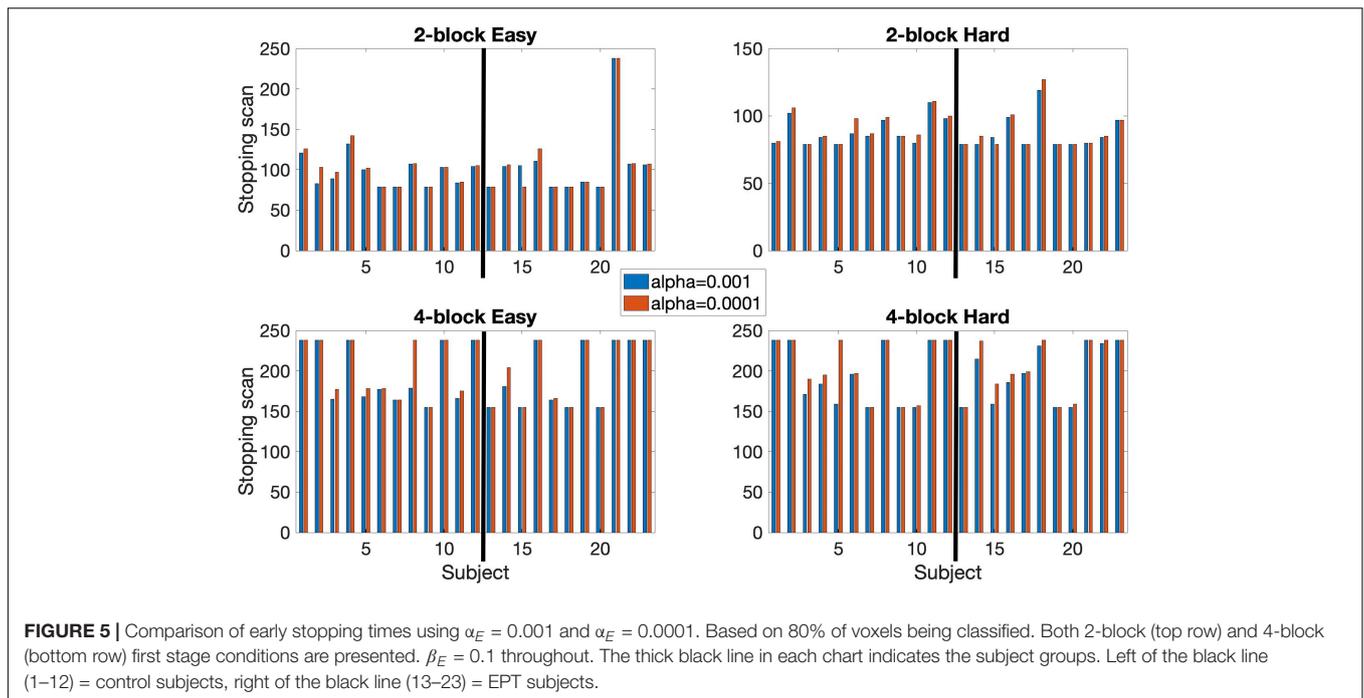
Stopping was reached at 80% of voxels classified as either active or non-active in all subjects with the exception of one instance (EPT subject, easy level) for 2-block first stage. For 4-block first stage, 7/12 control subjects stopped early for both the easy and hard levels. Among the EPT group, 6/11 subjects stopped early for the easy level and 9/11 subjects stopped early for the hard level. The median stopping duration for both difficulty levels for control subjects was 3 blocks of easy and 2 blocks of



hard stimulus administration for 2-blocks first stage. For 4-blocks first stage, the median stopping time was 5 easy, 4 hard for both difficulty levels. In EPT subjects, the median stopping time for 2-block first stage was 3 easy, 2 hard blocks of stimuli. For 4-blocks first stage, the median stopping time was 4 easy and 4 hard for the easy level and 5 easy, 5 hard for the hard level. Depending on the number of first stage blocks, time savings of 1/3 to 2/3 (4 to 8 min on a 12-min scan) can be achieved.

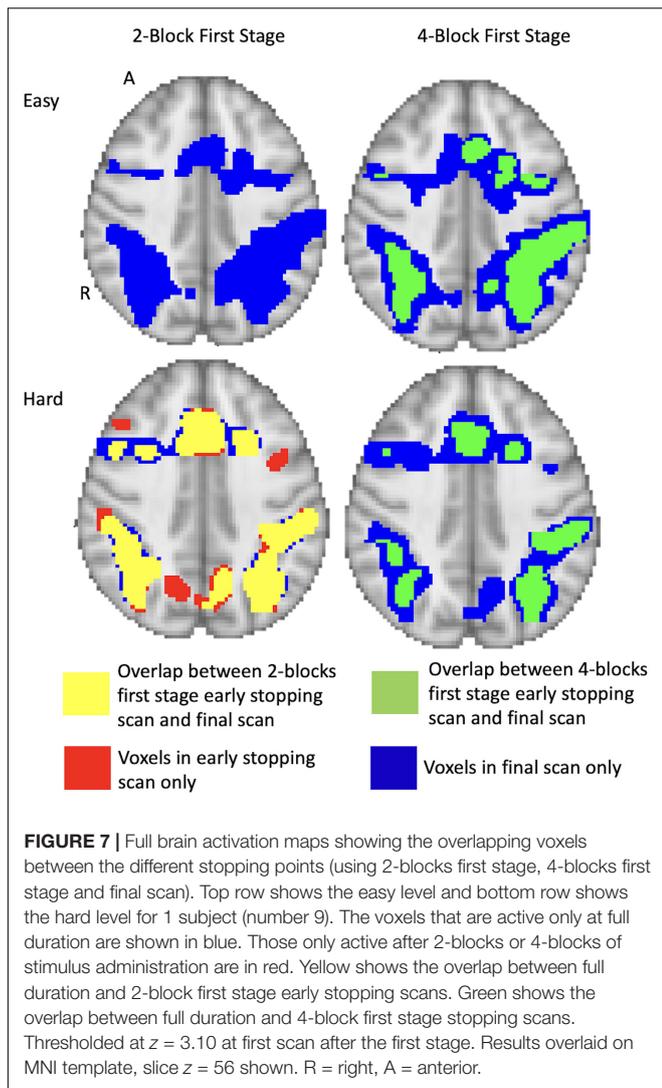
An early stopping rule based on a classification rate of at least 70 or 90% was also tested. Results reported in **Figure 6**. At 70% classification most subjects stopped early. For the 2-block first stage all subjects stopped early. Median stopping scan number was 79 for both the easy and hard levels for both groups. For the 4-block first stage – easy level, 1 EPT subject did not stop early, and 1 EPT and 1 control subject did not stop early for the

hard level. Median stopping scan was 155 for both difficulty levels for both groups. At 90% classification, most subjects stopped early with 2-block first stage but most did not stop early with 4-block first stage. For the 2-blocks first stage condition, 3 subjects did not stop early for the easy level and 2 subjects for the hard level. Only 1 subject stopped early under the 4-blocks first stage condition for the easy level and the hard level. When analyzing counts of voxels classified as active or non-active between these rules, the 80% thresholds lead to more non-active classifications, but the difference in active voxels is less systematic. Given that early stopping occurs almost invariably with the 70% rule, this value should also be considered. **Supplementary Table 2** provides plots of the percentage of voxels that are, respectively, classified as active and non-active over the course of the full scanning duration based on 80% of voxels being classified. A general trend



is that the percentage of non-active voxels gradually decreases while that of active voxels increases. Longer scan durations also allow for some adjustment in *post-hoc* analyses, and may have some potential advantages for group analysis, as discussed below. Hence, we present results for the more conservative 80% rule, which leads to relatively longer durations even when early stopping occurs.

In summary, these results indicate that early stopping is feasible but that the parameters need to be carefully chosen. The desire for earliest possible stopping must be balanced with selections of a first stage estimation duration and a voxel classification percentage which help insure that the number of active voxels will significantly overlap with those from a full scan, and that a sufficient amount of experimentation has been



conducted. Of the parameters we tested, our results indicate that useful results were obtained using 4-blocks first stage with 80% classification. We discuss this further in the Discussion section.

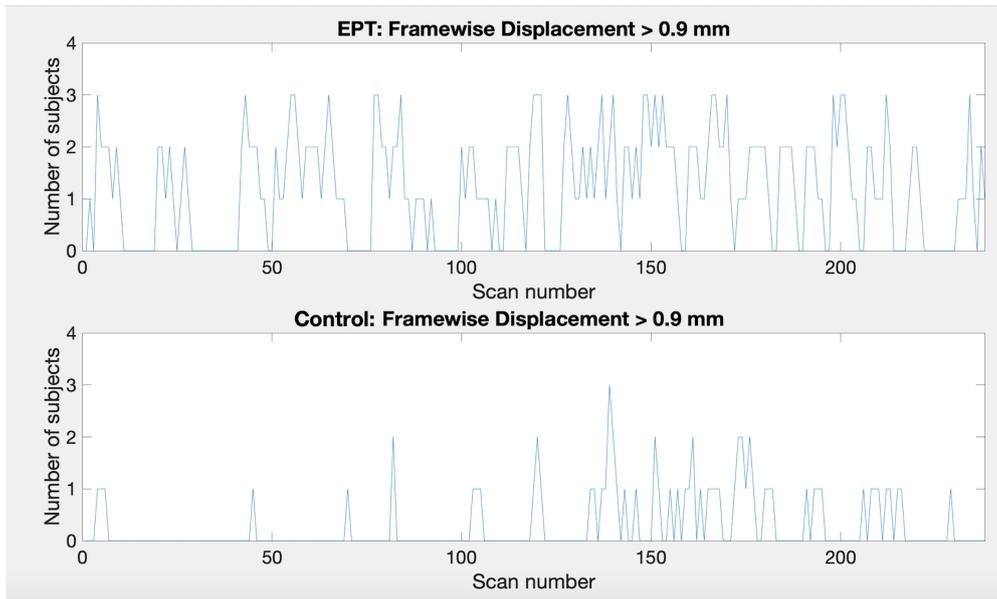
The activation maps under the different conditions are shown in **Figure 7** for a sample subject (subject 9). For the 2-block first stage, this subject terminated after 2 blocks of easy and hard administration for the easy level (scan 79) and after 3 blocks of easy and 2 blocks of hard administration for the hard level (scan 85). For the 4-block minimum, this subject terminated at scan 155, equal to 4 blocks of easy and hard stimulus administration, for both difficulty levels. The images show that at scan 79 there is very little activity present and the majority of the voxel classifications are non-active. By scan 155, there is much more activity which has a similar pattern to the final scan. The extent is not quite as large as the final scan, however, the foci of the clusters do overlap showing good correspondence between the two different time points. Other subject activation maps for instances where there is less than 50% overlap are shown in **Supplementary Table 1**.

By holding the alternative hypothesis threshold  $\theta_1$  constant after the first stage, note that the corresponding  $z_t$ -score associated with the  $\theta_1$  threshold actually increases as the scan durations increase. The decreasing variance of  $c\hat{\beta}_t$  is approximately proportional to  $c(X'X)^{-1}c'$  (as it would be under constant error variance and no serial correlation). Hence, for the easy level task, a  $z_t$ -score of 3.1 for 154 scans approximately corresponds to a  $z_t$ -score of 4.0 for the final scan (238 scans), holding  $\theta_1$  fixed. At 78 scans, a  $z_t$ -score of 3.1 corresponds to a  $z_t$ -score of approximately 8.37 at 238 scans, so there may be less overlap between early stopping and full duration for shorter first stage durations. See **Supplementary Table 1** for images resulting from different stop rules and first stage durations. In **Supplementary Table 2**, the trends in percentage of voxels classified as active and non-active reflect this phenomenon, at least for some of the subjects.

Importantly, the issue of whether early stopping or full duration provide better activation maps is best answered neuroscientifically, through the support of literature and hypotheses. We used this training sample to assess whether early stopping activation patterns were reasonable, or too early, and compared them with full duration results. In the Supplementary, we add plots for early stopping versus full duration for each subject for which early stopping was invoked. Comparing these plots, we see that in many instances that the cluster peaks are located in the expected anatomical locations.

In summary, although there are similar rates of early termination between the 2-block and 4-block first stage cases, the detected activation patterns suggest that using 4-blocks of stimulus administration is more suited to determining active voxels. In **Figure 4**, to illustrate the rate of decrease of the estimated  $\text{var}[c\hat{\beta}_t]$  values, we present a plot of  $\text{var}[c\hat{\beta}_t]$  values for one subject across a set of voxels over the duration of the experiment. These values give an indication that stopping based on the  $\theta_1$ -values at the end of the 2-block first stage may be too early for correspondence with full duration scans, as the estimated standard deviations are relatively larger and the corresponding  $\theta_1$ -values for  $z = 3.10$  in the alternative hypothesis can be larger as well, compared to a 4-block first stage.

Note that in a subset of subjects there was some volatility due to subject movement. This may in part explain why sometimes there is a lower number of active voxels for the easy task for full duration than when stopping early. In most subjects the active voxel count increases with scan duration. Subjects 13, 15 and 17 had the most pronounced decrease of active voxel counts. Subject 13 demonstrates very few active voxels at all and there is almost no consistency in location. Further investigation shows large relative framewise displacement occurs frequently throughout the scan and many of the responses have been missed or have relatively long response times, 60.5% correct overall and 2.08 s (SD 0.97 s) average response time (see plots for subject 13 in **Supplementary Table 3**). Taken together these suggest that either the task level may not have been aimed at the right level and/or the subject may have been uncomfortable and distracted in the scanner thereby attending to the task less than required for robust activations to occur. Subject 15 demonstrates consistent clusters but their size decreases over time. Framewise



**FIGURE 8** | Plots showing the number of subjects that pass the framewise displacement threshold of 0.9 mm for each scan. Top: EPT subjects, bottom: control subjects.

displacement shows very little motion, particularly from scan 180 onwards. The response plots (in **Supplementary Table 3**) show the subject is paying attention and responding appropriately. Subject 17 has a similar pattern of decreasing cluster sizes. The framewise displacement plots indicate a moderate amount of motion throughout. Although the subject has missed many of the task questions (65.8% correct), the pattern of responding indicates they are awake and attending to the task. In general, EPT subjects demonstrated more motion. The median number of scans with framewise displacement above a threshold of 0.9 mm [threshold determined from Siegel et al. (2014)] was 5 scans (SD 43 scans) for EPT subjects and 2.5 scans (SD 8 scans) for control subjects. One EPT subject passed the threshold a total of 124 scans out of 238 scans. In contrast, the control subject with the maximum number of threshold passes was 30/238 scans. This is further demonstrated in **Figure 8** where we show subject counts for each scan when the threshold has been passed. For both EPT and control subjects, it is clear that subjects are moving more frequently in the second half of the scans and supports stopping early to reduce motion artifacts and noise in the data. Formally, we see statistically significant differences when comparing counts of motion events with framewise displacement greater than 0.9 mm in the first versus second half of scanning ( $p = 0.003$ , two-sided signed rank test). EPT group also has significantly more movement in the first half of scanning ( $p = 0.035$ , two-sided Mann-Whitney  $U$ -test), indicating a group-level proclivity for more motion events.

## Group Analysis Results

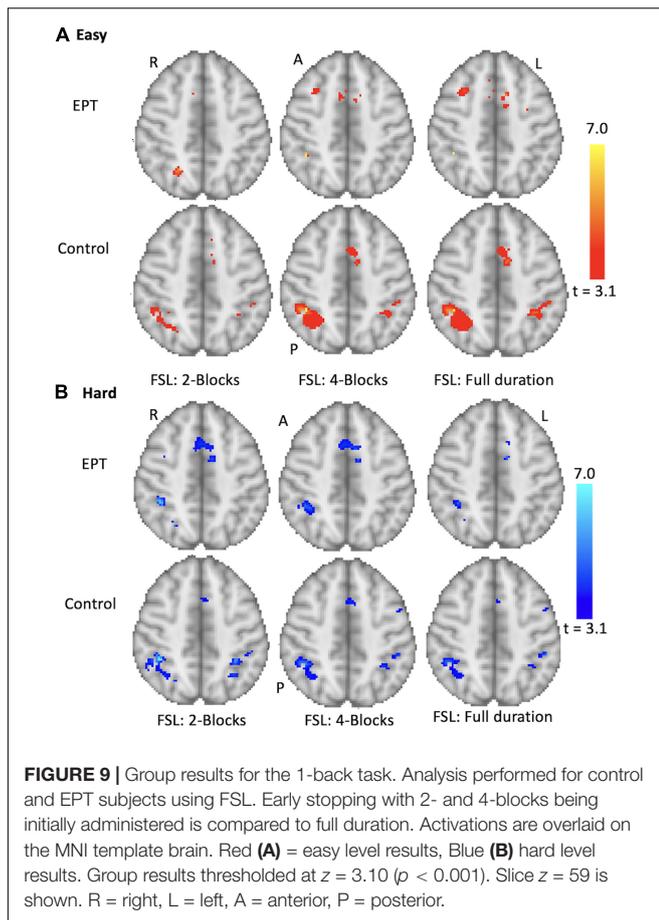
The results for the group 1-back easy and hard contrasts for the 2- and 4-block first stage conditions for EPT and control subjects are shown in **Figure 9**. Location of activity is listed

in **Supplementary Table 4**. The group results of full scan durations are compared to the group results using only the scans up to the early stopping point for each subject for each difficulty level and number of blocks completed before early stopping was allowed. We examined within-group differences as well as between group differences. The EPT > control (activity that is greater in the EPT group compared to the control group) and control > EPT (activity that is greater in the control group compared to the EPT group) contrasts. The focus for the results here is within-group for the easy and hard levels.

The control subjects show activations in the anterior cingulate and bilateral parietal regions, see **Supplementary Table 4**, and **Figure 9**. The cluster sizes and overlap are reported in **Table 3**. The easy and hard 4-block first stage scans for the control group appear similar to the final scans. There is less correspondence between the 2-block first stage scans and the final scans, reflecting the individual results reported above. The EPT group easy level scans are inconsistent across stages and there is more variability in the activations at both easy and hard levels. This is discussed below.

## DISCUSSION

Based on analysis of a training sample, we have presented a workflow for the implementation of an adaptive real-time fMRI system that allows for statistically driven dynamic adjustment of experimentation based on voxel-level SPRT. We show that this dynamic and adaptive statistical approach is under some circumstances comparable to corresponding fixed experimental designs in terms of detected activation.



**FIGURE 9 |** Group results for the 1-back task. Analysis performed for control and EPT subjects using FSL. Early stopping with 2- and 4-blocks being initially administered is compared to full duration. Activations are overlaid on the MNI template brain. Red (**A**) = easy level results, Blue (**B**) hard level results. Group results thresholded at  $z = 3.10$  ( $p < 0.001$ ). Slice  $z = 59$  is shown. R = right, L = left, A = anterior, P = posterior.

Our implementation illustrates the importance of first analyzing a training sample of full scan data, as we illustrate here. Experts can then review the activation thresholds, first stage duration lengths, stopping rule criteria, and the corresponding activation patterns that arise from early stopping rules, to see if they are plausible neuro-scientifically relative to those observed from longer scans. Certainly, our examples, particularly the group analyses are based on a small-scale neuroimaging study. In general, a larger training sample would provide a stronger basis for making these decisions.

We explored imposing two different first stage lengths before early stopping was considered using either 2- or 4-blocks each of easy and hard stimulus administration. The 4-block first stage is justified over the 2-block because of the comparative stability of the estimation of error variances and other GLM parameters. In contrast, for the 2-block first stage, parameter estimation can be more variable. Also, correspondence in early stop activation patterns to full scan duration involves larger  $z$ -score threshold changes over the scans from the initial 3.10-value that was specified, and activation magnitude thresholds for stopping that can be relatively higher. The 2-block first stage often led to most voxels being classified as non-active. See **Supplementary Table 2**. While the 4-block first stage provides less opportunity for efficiency gains, as the window for early stopping is narrower, but it is more prudent given the need for parameter estimates to

stabilize. It is possible that a 3-block initial stage could provide comparable results as the 4-block initial stage, but this was not explored here.

In the SPRT framework, other  $\alpha_E, \beta_E$  pairs were considered as well, to test how different combinations impact activity detection and early stopping. For instance, given selection of  $\alpha_E = 0.001$  and  $\beta_E = 0.01$ , early stopping was common for the 2-block first stage but arose less often for 4-block first stage and with much smaller savings in scan time. In this case, the more stringent choice of  $\beta_E$  makes it more difficult to cross either of the SPRT thresholds. We also saw that for either  $\alpha_E = 0.001$  or  $\alpha_E = 0.0001$  being paired with  $\beta_E = 0.1$ , early stopping occurred for both of the experimental conditions, with somewhat faster early stopping for the less stringent  $\alpha_E$ .

In terms of the global stop rule threshold, we observed that for the cases under consideration, stopping when 80% of voxels in the full brain (or smaller ROI), respectively, satisfy their SPRT-based stopping criterion generally leads to early stopping of stimulus administration, while also leading to comparable activation classification as with the full protocol. The stricter 90% criterion was infrequently satisfied, and did not often lead to early stopping of experimentation at 4-blocks first stage estimation. Recall that when GLM parameter values are “in-between” the null and alternative hypothesis values, SPRT-based stopping is less likely at the voxel level. A 100% stopping rule is thus not feasible, as are values relatively close to 100%. The 80% rule seems conservative in that not all participants are stopped early, but there are high levels of correspondence in individual and group level activation maps with full durations, particularly when the first stage is comprised of 4 blocks. The 70% rule is more aggressive, and early stopping is invoked more frequently. Given that the resultant images from early stopping in many cases appear similar across these two rules, the 70% rule should be considered as well.

The SPRT approach was effective at detecting brain activity at the individual level with early stopping in both the control and EPT groups. Note the individual variability among subjects in early stopping performance. Factors that can affect stopping times include the magnitudes of activation, variability in task performance, sustained attention levels, motion, and the noise levels in the BOLD signal. Those born EPT also can have structural abnormalities of the brain which can affect fMRI results and 2 subjects reported here had clear abnormalities that were obvious even in this low-resolution fMRI data ( $1.8 \text{ mm} \times 1.8 \text{ mm} \times 4.0 \text{ mm}$  voxel size). Subtle abnormalities may have been present in some of the other subjects.

The EPT group data demonstrated less consistent activity by comparison to control group data across all stopping points. In order to understand this result it is necessary to consider neuropsychological skills and structural and functional brain changes within the group. Working memory is a key skill required for both mathematics and this numerical 1-back task. Recall the lower accuracy and longer response times in the EPT group. fMRI studies on dyscalculia (difficulty in learning and performing mathematics) suggest that there is greater heterogeneity in activations with a more diffuse pattern being apparent (Berl et al., 2006; Kucian et al., 2006). Additionally,

**TABLE 3** | The number of active voxels that spatially overlap between early stopping and full duration group analyses are listed.

	No of Active Voxels	Voxels in Scan at Early Stopping but not Final Scan	Voxels in Final Scan but not in Scan at Early Stopping	% of Common Voxels with Final Scan	Standard Deviation Values
<b>EPT – Easy</b>					
2-Blocks	343	277	284	18.9	137.8
4-Blocks	439	234	145	58.8	92.6
Full duration	350				96.4
<b>EPT – Hard</b>					
2-Blocks	1,103	841	339	43.6	113.1
4-Blocks	955	491	137	77.2	98.5
Full duration	601				93.9
<b>Controls – Easy</b>					
2-Blocks	1,277	514	1147	39.9	122.3
4-Blocks	1,963	366	313	83.6	84.4
Full duration	1910				89.1
<b>Controls – Hard</b>					
2-Blocks	1,976	1111	1065	44.8	103.3
4-Blocks	1,672	326	584	69.7	90.3
Full duration	1930				84.5

Images thresholded at  $z = 3.10$  ( $p < 0.001$ ). The percentage of voxels-in-common is given relative to the total number of active voxels detected at full duration. Stopping based on 80% classification at the individual level. Standard deviation values are for the number of active voxels individually.

there is overlap in structural differences in white matter integrity, as measured from diffusion weighted imaging studies, between those born EPT and those with dyscalculia including inferior fronto-occipital fasciculus and the inferior and superior longitudinal fasciculi (Rykhlevskaia et al., 2009; Kucian et al., 2014; Young et al., 2017; Loe et al., 2019). These connect crucial areas associated with mathematics and working memory. A more diffuse and variable pattern of functional activity, perhaps partly due to structural differences, may confound a group analysis in this instance. The control group variances are relatively much lower throughout, as the extremely premature birth group was neurologically and cognitively more heterogeneous. If group-level analysis is a main objective, it is possible that groups could be treated differently in how early stopping is approached based on within-group heterogeneity and the need for more scan data to help overcome this. This issue needs further investigation. Certainly, relatively larger group sizes for patient groups also would be helpful in overcoming greater within-group variability.

Based on our analysis of head movements, early stopping of data collection could be particularly useful in populations where attention is problematic, such as with young children or adults with cognitive deficits. We see in our EPT population that head movement became more frequent and had a greater amplitude in the second half of the scans when compared to the control subjects. This may be an indication of attentional fatigue. In **Table 3**, it is seen for the easy level, the number of voxels classified as active actually decreases a bit at full scan duration compared with early stopping. This may be due to relative increases in variability from loss of attention. In general, we believe that there will be instances when early stopping can actually improve statistical power by reducing the possibility of fatigue.

In the future, it is possible that the first stage length can be tailored at the voxel level, once it is clear error variance and

other GLM parameter estimates are relatively stable, which is expected at some point due to the convergence properties of the estimators. This may facilitate earlier stopping. Alternatively, if local computational resources are limited, note that stopping can be assessed on an interval basis, and not necessarily after every scan. Although not considered here, these BOLD signal-based early stopping rules could also possibly be enriched by incorporating individual motion displacement patterns, as well as behavioral measures such as correctness rates in experimentation.

Here we demonstrated full brain analytics with parallelization using MKL Intel libraries for matrix computation with two Xeon E5-2687W 8-core processors. It is also feasible to consider only partial brain volumes where experiments demand more consideration of a particular area. Future directions for the study are to implement the SPRT and Bayesian sequential estimation methods using distributed computing approaches to increase processing speed and allow advance in stopping rule methods under shorter scan TR periods.

## CONCLUSION

We introduce a systematic, statistically based approach to dynamic experimentation with real-time fMRI. Saving in scan time and accurate voxel activation detection can be achieved, while redundant experimentation in block design is reduced. We investigate different aspects of how to determine early stopping rules. These analyses can be viewed as intended on a training sample to guide implementation of early stopping in future studies involving the same experiments and study populations. These methods can be particularly useful such as when investigating fragile patients and young children, as shorter scan requirements can enhance patient experience and

reduce fatigue effects. Another possible application is in real-time quality control, where it is important to know if sufficient testing has been conducted, or whether more experimentation is needed. These methods lay a foundation for future dynamic experimentation approaches and with real-time fMRI, including for resting state, neurofeedback and presurgical evaluation. Use of high-performance computing will enable the advent of more sophisticated real-time experimental designs and dynamically determined early stopping rules.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University Hospitals Cleveland Medical Center Institutional Review Board. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## REFERENCES

- Alegria, A. A., Wulff, M., Brinson, H., Barker, G. J., Norman, L. J., Brandeis, D., et al. (2017). Real-time fMRI neurofeedback in adolescents with attention deficit hyperactivity disorder. *Hum. Brain Mapp.* 38, 3190–3209. doi: 10.1002/hbm.23584
- Banerjee, A., Chitnis, U. B., Jadhav, S. L., Bhawalkar, J. S., and Chaudhury, S. (2009). Hypothesis testing, type I and type II errors. *Indust. Psychiatry J.* 18, 127–131.
- Berl, M. M., Vaidya, C. J., and Gaillard, W. D. (2006). Functional imaging of developmental and adaptive changes in neurocognition. *Neuroimage* 30, 679–691. doi: 10.1016/j.neuroimage.2005.10.007
- Carroll, R. J., Wang, S., Simpson, D. G., Stromberg, A. J., and Ruppert, D. (1998). *The Sandwich (Robust Covariance Matrix) Estimator*. Texas: Texas A&M University.
- Cavanna, A. E., and Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129, 564–583. doi: 10.1093/brain/awl004
- Chai, W. J., Abd Hamid, A. I., and Abdullah, J. M. (2018). Working memory from the psychological and neurosciences perspectives: a review. *Front. Psychol.* 9:401. doi: 10.3389/fpsyg.2018.00401
- Chen, G., Saad, Z. S., Nath, A. R., Beauchamp, M. S., and Cox, R. W. (2012). FMRI group analysis combining effect estimates and their variances. *Neuroimage* 60, 747–765. doi: 10.1016/j.neuroimage.2011.12.060
- Cox, D. R. (1963). Large sample sequential tests for composite hypotheses. *Indian J. Statist. Ser. A* 25, 5–12.
- Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173. doi: 10.1006/cbmr.1996.0014
- Cox, R. W., Jesmanowicz, A., and Hyde, J. S. (1995). Real-time functional magnetic-resonance-imaging. *Magn. Reson. Med.* 33, 230–236. doi: 10.1002/mrm.1910330213
- Creemers, H. R., Wager, T. D., and Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. *PLoS One* 12:e184923. doi: 10.1371/journal.pone.0184923
- Daamen, M., Bauml, J. G., Scheef, L., Sorg, C., Busch, B., Baumann, N., et al. (2015). Working memory in preterm-born adults: load-dependent compensatory activity of the posterior default mode network. *Hum. Brain Mapp.* 36, 1121–1137. doi: 10.1002/hbm.22691

## AUTHOR CONTRIBUTIONS

SC and CT contributed to study design, analysis and interpretation of data, and drafting of manuscript. WC contributed to study design, analysis and interpretation of data, software development, and drafting of manuscript. JF contributed to study design and software development. HF contributed to study design and technical support, and drafting of manuscript. JS-G contributed to software development and revision of manuscript. JZ contributed to analysis and interpretation of data, and revision of manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This study was supported by National Science Foundation (Award number: 1561716) and Phillips Healthcare.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2021.643740/full#supplementary-material>

- Debettencourt, M. T., Cohen, J. D., Lee, R. F., Norman, K. A., and Turk-Browne, N. B. (2015). Closed-loop training of attention with real-time brain imaging. *Nat. Neurosci.* 18:470. doi: 10.1038/nn.3940
- Di, X., Zhang, H. M., and Biswal, B. B. (2020). Anterior cingulate cortex differently modulates frontoparietal functional connectivity between resting-state and working memory tasks. *Hum. Brain Mapp.* 41, 1797–1805. doi: 10.1002/hbm.24912
- Feng, I. J., Jack, A. I., and Tatsuoka, C. (2015). Dynamic adjustment of stimuli in real time functional magnetic resonance imaging. *PLoS One* 10:e0117942. doi: 10.1371/journal.pone.0117942
- Friston, K. J., Josephs, O., Zarahn, E., Holmes, A. P., Rouquette, S., and Poline, J. B. (2000). To smooth or not to smooth? Bias and efficiency in fmri time-series analysis. *Neuroimage* 12, 196–208. doi: 10.1006/nimg.2000.0609
- Goebel, R. (2012). Brainvoyagerpast, present, future. *Neuroimage* 62, 748–756. doi: 10.1016/j.neuroimage.2012.01.083
- Hall, W. J. (1962). Some sequential analogs of steins 2-stage test. *Biometrika* 49:367. doi: 10.1093/biomet/49.3-4.367
- Hohenfeld, C., Nellesen, N., Dogan, I., Kuhn, H., Muller, C., Papa, F., et al. (2017). Cognitive improvement and brain changes after real-time functional MRI neurofeedback training in healthy elderly and prodromal Alzheimer's disease. *Front. Neurol.* 8:384. doi: 10.3389/fneur.2017.00384
- Huang, J. Z., Xie, L., Guo, R. W., Wang, J. H., Lin, J. Q., Sun, Z. B., et al. (2020). Abnormal brain activity patterns during spatial working memory task in patients with end-stage renal disease on maintenance hemodialysis: a fMRI study. *Brain Imaging Behavior* 15, 1898–1911. doi: 10.1007/s11682-020-00383-7
- ITU-T (2005). Series X: Data Networks and Open System Communications. Geneva, Switzerland.
- Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015
- Kauermann, G., and Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *J. Am. Statist. Assoc.* 96, 1387–1396. doi: 10.1198/016214501753382309
- Kiebel, S. J., and Holmes, A. P. (2006). "The general linear model," in *Statistical Parametric Mapping: the Analysis of Functional Brain Images*, ed. W. D. Penny (Elsevier Science & Technology).
- Koush, Y., Ashburner, J., Prilepin, E., Sladky, R., Zeidman, P., Bibikov, S., et al. (2017). OpenNFT: an open-source Python/Matlab framework for real-time fMRI neurofeedback training based on activity, connectivity and multivariate

- pattern analysis. *Neuroimage* 156, 489–503. doi: 10.1016/j.neuroimage.2017.06.039
- Kucian, K., Ashkenazi, S. S., Hanggi, J., Rotzer, S., Jancke, L., Martin, E., et al. (2014). Developmental dyscalculia: a dysconnection syndrome? *Brain Struct. Funct.* 219, 1721–1733. doi: 10.1007/s00429-013-0597-4
- Kucian, K., Loenneker, T., Dietrich, T., Dosch, M., Martin, E., and von Aster, M. (2006). Impaired neural networks for approximate calculation in dyscalculic children: a functional MRI study. *Behav. Brain Funct.* 2:31. doi: 10.1186/1744-9081-2-31
- Lawrence, E. J., Su, L., Barker, G. J., Medford, N., Dalton, J., Williams, S. C. R., et al. (2014). Self-regulation of the anterior insula: reinforcement learning using real-time fMRI neurofeedback. *Neuroimage* 88, 113–124. doi: 10.1016/j.neuroimage.2013.10.069
- Lazar, N. (2008). *The Statistical Analysis of Functional MRI Data*. New York: Springer-Verlag, doi: 10.1007/978-0-387-78191-4
- Li, J. X. (2010). *Sequential Probability Ratio Tests For Generalized Linear Mixed Models*. Riverside, CA: University of California.
- Lindquist, M. A., and Mejia, A. (2015). Zen and the art of multiple comparisons. *Psychosomatic Med.* 77, 114–125. doi: 10.1097/psy.0000000000000148
- Loe, I. M., Adams, J. N., and Feldman, H. M. (2019). Executive function in relation to white matter in preterm and full term children. *Front. Pediatr.* 6:418. doi: 10.3389/fped.2018.00418
- Lorenz, R., Monti, R. P., Violante, I. R., Anagnostopoulos, C., Faisal, A. A., Montana, G., et al. (2016). The automatic neuroscientist: a framework for optimizing experimental design with closed-loop real-time fMRI. *Neuroimage* 129, 320–334. doi: 10.1016/j.neuroimage.2016.01.032
- Lorenz, R., Violante, I. R., Monti, R. P., Montana, G., Hampshire, A., and Leech, R. (2018). Dissociating frontoparietal brain networks with neuroadaptive bayesian optimization. *Nat. Commun.* 9:1227. doi: 10.1038/s41467-018-03657-3
- Lui, S., Zhou, X. H. J., Sweeney, J. A., and Gong, Q. Y. (2016). Psychoradiology: the frontier of neuroimaging in psychiatry. *Radiology* 281, 357–372. doi: 10.1148/radiol.2016152149
- Menon, V. (2016a). Memory and cognitive control circuits in mathematical cognition and learning. *Mathe. Brain Across Lifespan* 227, 159–186. doi: 10.1016/bs.pbr.2016.04.026
- Menon, V. (2016b). Working memory in children's math learning and its disruption in dyscalculia. *Curr. Opin. Behav. Sci.* 10, 125–132. doi: 10.1016/j.cobeha.2016.05.014
- Peirce, J. W. (2007). PsychoPy psychophysics software in python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017
- Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinform.* 2:10. doi: 10.3389/neuro.11.010.2008
- Peirce, J., and MacAskill, M. (2018). *Building Experiments in Psychopy*. London: Sage Publications Ltd.
- Poudel, G. R., Stout, J. C., Dominguez, J. F., Gray, M. A., Salmon, L., Churchyard, A., et al. (2015). Functional changes during working memory in Huntington's disease: 30-month longitudinal data from the IMAGE-HD study. *Brain Structure Function* 220, 501–512. doi: 10.1007/s00429-013-0670-z
- Rykhlevskaia, E., Uddin, L. Q., Kondos, L., and Menon, V. (2009). Neuroanatomical correlates of developmental dyscalculia: combined evidence from morphometry and tractography. *Front. Hum. Neurosci.* 3:51. doi: 10.3389/neuro.09.051.2009
- Sato, J. R., Basilio, R., Paiva, F. F., Garrido, G. J., Bramati, I. E., Bado, P., et al. (2013). Real-time fMRI pattern decoding and neurofeedback using friend: an FSL-integrated BCI toolbox. *PLoS One* 8:e81658. doi: 10.1371/journal.pone.0081658
- Sawasd Tantaratana, J. B. T. (1977). Truncated sequential probability ratio test. *Inform. Sci.* 13, 283–300. doi: 10.1016/0020-0255(77)90050-0
- Schmidt, D. C., Levine, D. L., and Mungee, S. (1998). The design of the TAO real-time object request broker. *Comput. Commun.* 21, 294–324. doi: 10.1016/s0140-3664(97)00165-5
- Sherwood, M. S., Kane, J. H., Weisend, M. P., and Parker, J. G. (2016). Enhanced control of dorsolateral prefrontal cortex neurophysiology with real-time functional magnetic resonance imaging (rt-fMRI) neurofeedback training and working memory practice. *Neuroimage* 124, 214–223. doi: 10.1016/j.neuroimage.2015.08.074
- Sheslow, D., and Adams, W. (2003). *Wide Range Assessment Of Memory And Learning*. Lutz, FL: Psychological Assessment Resources, Inc.,
- Siegel, J. S., Power, J. D., Dubis, J. W., Vogel, A. C., Church, J. A., Schlaggar, B. L., et al. (2014). Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. *Hum. Brain Mapp.* 35, 1981–1996. doi: 10.1002/hbm.22307
- Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155. doi: 10.1002/hbm.10062
- Subramanian, L., Hindle, J. V., Johnston, S., Roberts, M. V., Husain, M., Goebel, R., et al. (2011). Real-time functional magnetic resonance imaging neurofeedback for treatment of Parkinson's disease. *J. Neurosci.* 31, 16309–16317. doi: 10.1523/jneurosci.3498-11.2011
- Subramanian, L., Morris, M. B., Brosnan, M., Turner, D. L., Morris, H. R., and Linden, D. E. J. (2016). Functional magnetic resonance imaging neurofeedback-guided motor imagery training and motor training for Parkinson's disease: randomized trial. *Front. Behav. Neurosci.* 10:111. doi: 10.3389/fnbeh.2016.00111
- Tantaratana, S., and Thomas, J. B. (1977). Truncated sequential probability ratio test. *Inform. Sci.* 13, 283–300.
- Vakamudi, K., Trapp, C., Talaat, K., Gao, K. X., Guimaraes, B. S. D., and Posse, S. (2020). Real-time resting-state functional magnetic resonance imaging using averaged sliding windows with partial correlations and regression of confounding signals. *Brain Connect.* 10, 448–463. doi: 10.1089/brain.2020.0758
- Wald, A. (1947). *Sequential Analysis*. Hoboken, NJ: John Wiley and Sons.
- Wald, A., and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* 19, 326–339. doi: 10.1214/aoms/1177730197
- Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2004). Multilevel linear modelling for fMRI group analysis using bayesian inference. *Neuroimage* 21, 1732–1747. doi: 10.1016/j.neuroimage.2003.12.023
- Woolrich, M. W., Ripley, B. D., Brady, M., and Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling Of fMRI data. *Neuroimage* 14, 1370–1386. doi: 10.1006/nimg.2001.0931
- Yan, C., Liu, D., He, Y., Zou, Q., Zhu, C., Zuo, X., et al. (2009). Spontaneous brain activity in the default mode network is sensitive to different resting-state conditions with limited cognitive load. *PLoS One* 4:e5743. doi: 10.1371/journal.pone.0005743
- Yoo, J. J., Hinds, O., Ofen, N., Thompson, T. W., Whitfield-Gabrieli, S., Triantafyllou, C., et al. (2012). When The brain is prepared to learn: enhancing human learning using real-time fMRI. *Neuroimage* 59, 846–852. doi: 10.1016/j.neuroimage.2011.07.063
- Young, J. M., Morgan, B. R., Whyte, H. E. A., Lee, W., Smith, M. L., Raybaud, C., et al. (2017). Longitudinal study of white matter development and outcomes in children born very preterm. *Cerebral Cortex* 27, 4094–4105. doi: 10.1093/cercor/bhw221

**Conflict of Interest:** HF was employed by the company Philips Healthcare, USA. JS-G was employed by the company Philips Healthcare, Spain.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declare that this study received funding from Philips Healthcare. The funder had the following involvement in the study: technical support and software development for the real-time output of fMRI files and revision of the manuscript.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Carr, Chen, Fondran, Friel, Sanchez-Gonzalez, Zhang and Tatsuoka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.