Improving the signal detection accuracy of functional Magnetic Resonance Imaging

Niels Janssen[a,b,*], Juan A. Hernández-Cabrera[a,c], Laura Ezama Foronda[a]

[a] *Psychology Department, Universidad de la Laguna, Tenerife, Spain*
[b] *Institute of Biomedical Technologies, Universidad de la Laguna, Tenerife, Spain*
[c] *Basque Center on Cognition, Brain and Language, San Sebastián, Spain*

Abstract

A major drawback of functional Magnetic Resonance Imaging (fMRI) concerns the lack of detection accuracy of the measured signal. Although this limitation stems in part from the neuro-vascular nature of the fMRI signal, it also reflects particular methodological decisions in the fMRI data analysis pathway. Here we show that the signal detection accuracy of fMRI is affected by the specific way in which whole-brain volumes are created from individually acquired brain slices, and by the method of statistically extracting signals from the sampled data. To address these limitations, we propose a new framework for fMRI data analysis. The new framework creates whole-brain volumes from individual brain slices that are all acquired at the same point in time relative to a presented stimulus. These whole-brain volumes contain minimal temporal distortions, and are available at a high temporal resolution. In addition, statistical signal extraction occurred on the basis of a non-standard time point-by-time point approach. We evaluated the detection accuracy of the extracted signal in the standard and new framework with simulated and real-world fMRI data. The new slice-based data-analytic framework yields greatly improved signal detection accuracy of fMRI signals.

Keywords: fMRI BOLD, detection accuracy, FIR basis functions, statistical modeling, Slice-Based fMRI

Brain function is frequently investigated using the Blood Oxygen Level Dependent (BOLD) signal in functional Magnetic Resonance Imaging (fMRI; Ogawa et al., 1990). Improving the accuracy of methods that detect the BOLD signal is of primary importance in many fMRI research contexts. One recent approach has relied on the implementation of advanced MRI pulse-sequences and updated hardware configurations to acquire whole-brain fMRI data with a high temporal resolution (e.g., Chang et al., 2013; Feinberg et al., 2010; Lin et al., 2006; Moeller et al., 2010; van der Zwaag et al., 2006). The higher temporal resolution enables a more precise sampling of the BOLD signal and leads to improved statistical detection and estimation of BOLD signal dynamics in task-based fMRI studies (Chen et al., 2015; Constable & Spencer, 2001; Dilharreguy et al., 2003; Sahib et al., 2016; Vu et al., 2016; Witt et al., 2016). In addition, a complimentary approach to improve BOLD signal detection has relied on specialized paradigm design and statistical techniques. For example, past studies have used jittered stimulus presentation with Finite Impulse Response (FIR) modeling to yield higher temporal resolution BOLD signals (e.g., Josephs et al., 1997; Lindquist et al., 2009; Maccotta et al., 2001; Miezin et al., 2000; Price et al., 1999; Serences, 2004; Toni et al., 1999). Here we attempted to further improve these latter data-analytic methods of BOLD signal detection by focusing on two specific issues that hamper the accuracy of BOLD signal extraction: 1) the volume-creation method, and 2) the statistical method.

The first reason why BOLD signal detection in the current fMRI data-analytical framework may be suboptimal is due to the specific method of volume creation. Volume creation refers to the way in which individually acquired brain slices are inserted into whole-brain volumes. A peculiar aspect of fMRI data acquisition is that instead of sampling the entire brain at once, spatially separate brain slices that cover the entire brain are sampled at different moments in time (Cohen & Weisskoff, 1991; Moeller et al., 2010). The current standard practice to create whole-brain volumes from such individually acquired brain slices is to simply time-shift spatially adjacent slices into whole-brain volumes (see

*Corresponding author
Email address: njanssen@ull.es (Niels Janssen)

Figure 1 and Appendix 1 for a formal treatment). Given typical whole-brain fMRI sampling parameters of 1 to 3 seconds, this means that time-varying signals sampled from spatially adjacent brain locations may be temporally shifted. Several studies have shown that such data yield BOLD signals that are detected with poor accuracy (Calhoun et al., 2000; Henson et al., 1999; Parker et al., 2017; Sladky et al., 2011). Consequently, these studies also show that signal detection can be improved by a procedure called Slice-Time Correction (STC). STC attempts to alleviate the temporal distortions by either interpolating signals between timepoints (Calhoun et al., 2000; Henson et al., 1999; Sladky et al., 2011), or by first low-pass filtering and then re-aligning signals in time (Parker et al., 2017). However, while these studies demonstrate that STC enhances BOLD signal extraction, it is also clear that STC is only required because of the temporal distortions introduced by the specific method of volume creation. It therefore remains to be seen whether signal extraction can be further improved by alternative methods of volume creation that crucially do not introduce such temporal distortions and hence do not require STC.

A second reason why BOLD signal extraction may be suboptimal is because of the statistical method of signal extraction. Specifically, within the current data-analytical framework, BOLD signal extraction is performed using so-called FIR basis functions (e.g., Josephs et al., 1997; Lindquist et al., 2009; Maccotta et al., 2001; Miezin et al., 2000; Ollinger et al., 2001; Price et al., 1999; Serences, 2004; Toni et al., 1999). The FIR basis functions represent parameters in a General Linear Model (GLM) that each capture a particular point in the progression of the BOLD signal generated by the presentation of stimuli in an imaging run. Formally, within this framework, for a given set of stimuli S, the design matrix X with m volumes (rows) and n basis functions (columns) is represented by:

$$X_{ij} = \begin{cases} 1, & \text{if } j = i - (S_p - 1) \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where S_p ranges over all possible volume-based stimulus onsets. The number of basis functions is typically determined by the ratio between the desired epoch length and the repetition time TR and represents the temporal resolution of the extracted signal. Additional basis functions and appropriate jittering of stimuli can be used to increase the temporal resolution (e.g., Josephs et al., 1997; Toni et al., 1999). Given the design matrix X determined above, modeling of fMRI time-series data Y for a given voxel is performed using the standard GLM function:

$$Y = X\beta_{0..n} + e, \tag{2}$$

where each β_j is a value that indexes the strength of the BOLD signal at a particular time point since the presentation of the stimulus. Importantly, given the design of matrix X, note that the number of datapoints to go into the estimation of each β_j value is equal to the number of stimuli in the imaging run (i.e., the number of 1s in each column of X). Approximate values for the β_js in this set of linear equations is typically obtained by the least-squares solution:

$$\beta_{0..n} = (X^T X)^{-1} X^T Y. \tag{3}$$

Obtaining an associated t-value with each beta coefficient first involves calculating the mean square error of this model:

$$\sigma^2 = \frac{(Y - Xb)(Y - Xb)^T}{n - m}, \tag{4}$$

where the numerator term $Y - Xb$ refers to the difference between the obtained and fitted data (i.e., the residuals), and $n - m$ to the available degrees of freedom. Next, the variance associated with each estimated beta-coefficient is given by:

$$var(\beta_{1..n}) = \sigma^2 (X^T X)^{-1}, \tag{5}$$

where the standard error for a given β_j is obtained by taking the square root of the diagonal values in this matrix. The final t-value can then simply be calculated as the ratio between a given β_j and its standard error. An appropriate ordering of the beta coefficients or t-values will then result in the statistically extracted BOLD signal.

There are at least three main problems with this FIR based approach that may hamper optimal detection of the BOLD signal. First, parameter estimation in the FIR modeling approach is optimal only if the stimulus induced BOLD signal is stationary across the imaging run (Donnet et al., 2006). Under such conditions, a given β_j corresponding to a particular timepoint in the BOLD signal progression is estimated from data that contains a minimal amount of distortion in time, and the estimates will therefore be optimal. However, previous studies have observed attentional and top-down influences on the trial-by-trial variability in BOLD signal onset and shape across an imaging run (e.g., Donnet et al., 2006; Duann et al., 2002; Grill-Spector et al., 2006), and therefore undermine the assumption of stationarity. The data from these studies raise the question of whether alternative methods exist that are better suited to address

the trial-by-trial variability in BOLD signal onset and shape.

Second, a particular statistical limitation of the GLM is that it precludes the modeling of random sources of variance such as those due to item variability (e.g., Bedny et al., 2007; Westfall et al., 2016). It is well-known that ignoring sources of variance in the data may introduce biases in parameter estimation. As before, this raises the question of whether BOLD signal detection may be improved by alternative modeling techniques in which the aforementioned trial-by-trial variability is brought under statistical control.

Finally, the FIR modeling approach ascribes a rather counterintuitive meaning to the standard errors associated with the $beta_j$s at each timepoint. Specifically, in the FIR modeling approach, data from the entire imaging run is used to estimate all the timepoints simultaneously. This means that the standard error that is associated with each β_j corresponding to a particular timepoint is not only determined by the quality of the model fit at that particular timepoint, but by the quality of the model fit at *all* timepoints (see Equation 5). In other words, the standard error at a particular timepoint does not reflect the quality of data fitting at that particular timepoint alone, but reflects the quality of data fitting at all other timepoints as well. A practical implication of this is that a noise event in the fMRI signal at one particular timepoint will increase the standard error at all extracted timepoints. Consequently, if BOLD signal extraction relied on t-values, this will affect the accuracy of BOLD signal extraction at all timepoints, even if the noise event affected only a single timepoint.[1] Thus, for these three reasons, the FIR based method of signal extraction may lead to a suboptimal detection of the BOLD signal from fMRI data.

To summarize, within the current framework of fMRI data analysis, BOLD signal extraction is hampered by the specific method of volume creation as well as by the specific method of statistical modeling. Here we proposed a new framework for the analysis of fMRI data. This framework incorporates a new method of volume creation, as well as a non-standard technique of statistical signal extraction. The framework places special importance on the slice acquisition times, that is the exact points in time when each slice in the fMRI data stream is acquired. Specifically, in the new method, whole-brain volumes are created out of slices that are all acquired at the same point in time *relative to a presented stimulus*. This is achieved by presenting stimuli in-phase with the slice acquisition times, and then calculating when each slice was acquired relative to a presented stimulus. These relative acquisition times for each slice can then be used to compose whole-brain volumes in which each slice was acquired at the same moment in time relative to a stimulus. (see Figure 2 and Appendix 1 for a formal treatment). Importantly, this method of whole-brain volume construction does not rely on time-shifting slices as in the standard method. This means that no temporal distortion is introduced in the data and hence, no STC is required.

In addition, in this new fMRI data format, the BOLD signal is extracted using a non-standard Timepoint by Timepoint approach. Although this statistical approach to signal extraction is commonly used in EEG/MEG research (Janssen et al., 2014; Lage-Castellanos et al., 2010; Smith & Kutas, 2015), it is only rarely applied to fMRI data (but see Cohen et al., 1997; Leung et al., 2000). In the Timepoint by Timepoint approach, the raw, sliced-based fMRI signal is first epoched into time periods where the BOLD response is likely to occur (i.e., stimulus-locked), and then signal intensities from a baseline period (e.g., time points prior to stimulus onset) are compared to signal intensities obtained at later time points in the epoch. Similar to previous studies (e.g., Josephs et al., 1997), because stimuli are presented in-phase with the slice acquisition times, the number of timepoints in an epoch and therefore the maximum temporal resolution with which the BOLD signal can be extracted is determined by $\frac{TR}{num\_slices}$, and may be on the order of tens of milliseconds. Crucially, the Timepoint by Timepoint approach may be less affected by variability in the BOLD signal onset and shape because model coefficients depend on the direct comparison of intensity values between the timepoints in the epoch and the baseline, and leading to more accurate parameter estimation. In addition, parameter estimation in this approach is performed using Linear Mixed Effect (LME) modeling (Bates, 2005; Pinheiro & Bates, 2000; Westfall et al., 2016). This modern statistical modeling approach permits the inclusion of multiple sources of random variance (see Appendix 2). Finally, because separate models are fitted at each timepoint instead of fitting all timepoints simultaneously, standard errors are less sensitive to potential noise events at other timepoints. Given the central role of slices in this method, we will refer to this framework as *Slice-Based* fMRI.

The current paper reports on tests that evaluated the accuracy of BOLD signal detection in the new Slice-Based method versus the standard FIR based models with STC and without STC. Given that the Slice-Based

[1] This may suggest that only beta-values should be used. However, ignoring the standard error introduces new complications in the modeling efforts.

method contains both a new method of volume creation and a different method of statistical signal extraction, a fourth, intermediate model was considered that relied on a standard method of volume creation with STC, but used the Timepoint by Timepoint method of statistical signal extraction. We will refer to this latter model as the Timepoint by Timepoint with STC method. The comparison of these four models allowed for an evaluation of both the new volume creation method as well as the new Timepoint by Timepoint technique on the accuracy of BOLD signal extraction from fMRI data. Specifically, a contrast of the FIR with STC model with the Timepoint by Timepoint with STC model uses the same volume creation method yet uses a different statistical technique and therefore allowed for the evaluation of the new statistical method of signal extraction. In addition, the comparison of the Timepoint by Timepoint with STC and the Slice-Based model uses the same statistical method but relies on different methods of volume creation and therefore allowed for the evaluation of the new volume creation technique.

These four models were evaluated in the context of three simulations and one real-world experiment. The simulations were not designed to examine signal extraction under ideal circumstances, but instead, provided an evaluation of the four models under relatively realistic conditions in an fMRI experiment. In Simulation 1, we examined the impact of trial-by-trial variability in the onset of the BOLD response in consecutive stimulus presentations in an imaging run. In Simulation 2 we examined the impact of trial-by-trial variability in BOLD shape, and in Simulation 3 we examined the impact of a single noise event in the imaging run (a signal intensity spike). Each method's performance was examined in the context of a slow event-related imaging run with 36 stimuli. The data were sampled from 3 slices containing only a single voxel. To examine the impact of increasing the sampling frequency the simulations were repeated with TRs of 3 and 1 second. BOLD signal extraction was performed using t-values. Performance was evaluated in terms of two measures: (i) the Pearson correlation between the ground-truth signal and the extracted signal, and (ii) the mean absolute difference between the ground-truth signal and the extracted signal. Given the arguments presented above we expected superior performance of the Slice-Based method compared to all other methods.

Finally, the four methods were evaluated in the context of in-vivo fMRI data collected from 30 participants performing a picture naming task. This task was chosen because of its various cognitive components (visual identification, name retrieval from memory, and motor output) which may yield complex BOLD signal dynamics across different areas of the brain. The question was which of the four methods were best suited to detect activity under such conditions. We first evaluated the basic signal detection capabilities of the Slice-Based method by comparing group-level activation maps obtained using this method to the standard GLM and Timepoint by Timepoint methods using Pearson and Dice indices. In addition, we compared BOLD signal extraction using the aforementioned methods from three adjacent slices covering left motor cortex. BOLD signal extraction was compared in terms of four measures: (i) the mean inter-slice correlation, (ii) the mean number of unique peaks (UP), (iii) the mean Time To Peak (TTP), and (iv) the mean maximum t-value (MAXT). Given the reduced impact of temporal distortions on volume creation and the more sensitive statistical method, we expected better performance for the Slice-Based method.

Methods

Simulation 1 - variability in BOLD onset

Simulations were performed in the software R (v3.4.0) using the neuRosim package (v0.2-12; Welvaert et al., 2011). To simulate an fMRI imaging run, 36 stimuli presented at long 18 s intervals induced a series of hemodynamic responses that were modeled with a double gamma function with default parameters (a1=6, a2=12, b1=0.9, b2=0.9, c=0.35). This signal was generated at a very high temporal resolution (accuracy = 0.1 s). The precise onsets of the stimuli were constructed to be in-phase with the slice acquisition times determined by the fMRI sampling parameters described below. Variability in the onset of the BOLD response was modeled by a stochastic process that for each BOLD response either shifted the onset by +0.5 s in time or did not shift onset (P=0.5). This means that for a given simulation, about 18 out of 36 stimuli yielded a BOLD onset that was 0.5 s off a (stimulus-induced) stationary onset. If such shifts in onset yield commensurate delays in behavioral response times, then they would yield a standard deviation in response time across all stimuli of around 250 ms. This value is well within the range observed in many behavioral tasks such as picture naming and therefore justifies our choice of realistic parameters for this simulation (e.g., Szekely et al., 2004).

Next, the hemodynamic signal was sampled by three slices in a simple bottom up sequential fashion. Each slice had only a single voxel, meaning that only a single time course was obtained for a given slice. The signal was sampled at two different sampling frequencies. At

the TR of 3 s with 3 slices this meant that every slice sampled the signal at a 1 second interval; At the TR of 1 s with 3 slices, the signal was sampled at a 0.33 s interval. Thus, each slice sampled the exact same hemodynamic response, although as mentioned before, the sequential nature of this serial sampling procedure introduces temporal shifts. In the last step of data sampling white noise with sigma=0.15 was added to the generated time series. Although fMRI data is known to contain other sources of noise (i.e., machine noise, physiological noise), in order to facilitate interpretation it was decided to only add white noise.

Next, three data sets of whole-brain volumes were created from the raw fMRI data. First, a standard volume creation method was used to create a time series of 362 whole-brain volumes in which it was assumed that all three slices within a volume were acquired at the same point in time (see Figure 1). A second data set was created by applying AFNI's *3dTshift* STC function to the first data set. Importantly, signals were aligned to the first slice in the volume meaning that no adjustments to the design matrix were required. Interpolation was based on the default Fourier method which is assumed to be the most accurate. This therefore yielded a slice time corrected dataset. Finally, the Slice-Based method of volume creation was applied to the raw fMRI data to create a third data set in which all slices within a volume were acquired at the same moment in time relative to a stimulus (see Figure 2). As mentioned before, this was achieved by combining slices with identical relative acquisition times acquired during the presentation of different stimuli into the same volume. At the TR = 3 s, this epoch had 18 timepoints (i.e., 1 s temporal resolution), whereas at TR = 1 s the epoch had 54 time points (i.e., 0.33 s temporal resolution). Importantly, these three data sets created by different volume creation methods were always based on the same raw fMRI data.

Statistical extraction of the BOLD signal by the FIR, Timepoint by Timepoint, and Slice-Based methods was performed on these data. For the FIR methods, we constructed a design matrix with $\frac{epoch\_length}{TR}$ basis functions (e.g., 6 basis functions for an epoch length of 18 s and a TR of 3 s; See Equation 1). To obtain a temporal resolution higher than the TR and equal to the resolution obtained using the Slice-Based method, two additional sets of $\frac{epoch\_length}{TR}$ basis functions were added and corresponded to (jittered) stimulus onsets close to multiples of 0.33 and 0.67 * TR (e.g., Dale, 1999; Josephs et al., 1997; Price et al., 1999; Toni et al., 1999). This led to a design matrix with a number of parameters that depended on the TR. Specifically, at TR = 3 s there were 18 parameters in the design matrix, whereas for TR = 1 s, there were 54 parameters in the design matrix. Note that all basis functions were orthogonal, and that although the number of parameters is high, it remained well below the total number of available datapoints, thereby avoiding overfitting risks. No temporal derivatives were used. This same design matrix was used for the FIR without STC and the FIR with STC methods, where the FIR without STC used the standard dataset for signal extraction, and the FIR with STC used the slice time corrected data set. All statistical modeling was done using the linear modeling (*lm*) function of R.

For the Timepoint by Timepoint method, epochs were extracted from the standard volume creation dataset with slice-time correction. It was assumed that each volume in the dataset was acquired at the onset of the TR. Next, for each stimulus onset, a set of volumes corresponding to the epoch length were chosen and for each volume in the epoch the relative time since stimulus onset was calculated. BOLD signal extraction took place on the basis of comparing signal intensities at baseline (define as timepoint 0) with those of subsequent timepoints in the epoch. No averaging of data was performed. Model fitting took place using the R package lme4 (v1.1_13) (Bates, 2005). Specifically, the formula used was `lmer(Intensity~Time+(1|epoch))`, where `Time` was a fixed-effect factor with two levels (the baseline and the relevant timepoint), and `epoch` was random-effect variable referring to the item number. Finally, the Slice-Based method used the same signal extraction method as the Timepoint by Timepoint method, except that the volume creation method was slice-based and not volume-based. This difference in volume creation method may lead to more accurate signal extraction in the Slice-Based method for two reasons: First, given that no STC is required, and hence no data is interpolated, extraction of a more veridical signal is expected than in the Timepoint by Timepoint with STC method. Second, given that in the Slice-Based method the onset and offsets of epochs are determined by the precise slice-acquisition times and not by the TR-based volume acquisition times, extracted epochs correspond more closely to actual stimulus onsets and offsets and therefore result in a more precise allocation of datapoints to timepoints in the epoch than in the Timepoint by Timepoint method. This improved alignment may then result in a more accurate extraction of the BOLD signal (see Discussion and Supplementary Materials for further discussion of this point).

Performance of each model was evaluated by the comparison to a ground-truth signal. Because the original signal was specified in different units than the statis-

tically extracted signal, no direct comparisons were possible. Instead, the ground-truth signal was set to have a maximum t-value amplitude of 25. This amplitude of the ground-truth signal was found to be sufficiently high such that the simulations performed with the particular noise levels did not reach this value. The ground-truth signal was then calculated with this maximum effect-size parameter using the double gamma function that formed the basis of the original fMRI data. Importantly, the same ground-truth signal was used across all simulations and was the same for all four evaluated methods. The accuracy of BOLD signal detection was determined using two measures: First, accuracy was determined by the Pearson correlation between the ground-truth signal and the signal at a particular slice. The mean Pearson correlation (denoted $\bar{r}_1$) was then computed as the mean correlation across all slices. In addition, the accuracy was also determined by the mean absolute distance between the ground truth and the signal at a particular slice:

$$d = \frac{\sum_{i=1}^{n} |a_i - b_i|}{n} \quad (6)$$

where n is the number of timepoints in the epoch, a is the ground truth signal and b is the extracted BOLD signal at a given slice. The value $\bar{d}$ was then calculated as the mean d value across all slices. The main advantage of this distance measure over a Pearson correlation is that the distance measure takes into account the amplitude of the response and therefore provides a more precise indication of the degree to which the extracted BOLD signal approximated the ground-truth signal. Note that lower $\bar{d}$ values indicate a more closely extracted signal. In total 100 simulations were performed at each TR.

Simulation 2 - variability in BOLD shape

In Simulation 2, the impact of variability in the BOLD shape across an imaging run on BOLD signal extraction by the four methods was examined. Variability in the BOLD shape was modeled by changing the parameter values of the double gamma function that was used to generate the baseline BOLD signal. Specifically, for half the stimuli in this simulation experiment, the BOLD response was generated by a double gamma function with adjusted values (a1=6, a2=12, b1=0.7, b2=0.7, c=0.25), while the other half had default parameter values (see above). Note that the b1 parameter controls the dispersion of the response, the b2 parameter controls the dispersion of the undershoot, and that the c parameter controls the scale of the undershoot. With respect to the default settings in the gamma function, these parameters were reduced to yield a BOLD response that was slightly more narrow. All other aspects of Simulation 2 were identical to Simulation 1.

Simulation 3 - impact of single spike

In Simulation 3, the impact of a single intensity spike on BOLD signal extraction by the four methods was investigated. This spike was modeled by changing a single intensity value in the fMRI simulated time series of slice 1 at a timepoint that was sampled at the end of an epoch (i.e., during the BOLD undershoot). This particular intensity value at this timepoint was set to 5 times the maximum BOLD signal (i.e., the maximum BOLD signal was 1, the value was set to 5). In other words, the fMRI time series of slice 1 consisted of 362 time points, and the intensity value at a single timepoint that was located at the end of a stimulus induced BOLD signal was set to 5 times the maximum BOLD signal. Intensity values at all other 361 timepoints for slice 1 remained unchanged. Note that such spikes in the signal are a frequent occurrence in fMRI data and are thought to be the result of head motion and the resulting spin-history artifacts (e.g., Friston et al., 1996).

In-vivo data - Picture Naming

Participants

Thirty native speakers of Spanish took part in the experiment (20 females, 10 males, mean age 22 yrs). Participants were students at the University of La Laguna, and received course credit or were paid 10 Euro. Twenty-nine participants were right-handed. The study was conducted in compliance with the declaration of Helsinki, and all participants provided informed consent in accordance with the protocol established by the Ethics Commission for Research of the university of La Laguna (Comit de tica de la Investigacin y Bienestar Animal).

Experimental setup and procedure

Two stimuli were used in the task: First, an image which participants were asked to name aloud, and second, a fixation cross ('+') which indicated rest (see Figure 3 for an overview). Twenty-seven pictures were selected from an image database that contained standardized line-drawings that were normed on various aspects (Szekely et al., 2004). Only those images were selected that had names that were consistently produced across participants in the norming study (i.e., those with ¿ 90% name-agreement).

Stimuli were presented in a slow event-related design, where a stimulus was presented for 0.5 s followed by an ISI blank screen for 12 s plus an additional jitter period.

The duration of the jitter period was randomly chosen without replacement from a uniform distribution of 36 times from 0 to 1855 ms in steps of 53 ms. This method of stimuli presentation resulted in the optimal jittering of stimuli for the Slice-Based method (see Figure 2 for further details). Stimulus presentation was directly synchronized with the MRI machine.

The Experiment involved three consecutive runs. In each run, 36 stimuli were presented, of which half were pictures and half were rest (i.e, fixation cross). In each run, nine different pictures were randomly selected and which were presented twice. Different pictures were selected for each run, and all twenty-seven pictures were presented in the experiment. For each run, the order of the stimuli was fully randomized on a by-participant basis. Stimulus presentation was controlled by Neurobs Presentations (v14). Participants in the scanner viewed the stimuli with MRI compatible goggles made by VisuaStim. These goggles provided an image resolution of 800 by 600 pixels at 60 Hz.

MRI acquisition parameters

MR-images were acquired using a 3T Signa Excite scanner (General Electric, Milwaukee, WI, USA) using a standard transmit/receive 8 channel gradient head coil. Head movement was strenuously avoided by fixating each participant's head with spongepads inside the coil. T2*-weighted images were obtained using standard Gradient Echo, Echo Planar Imaging (EPI) sequences.

Each run started with 10 dummy volumes that allowed for steady-state tissue magnetization. Each volume contained 36 slices that were acquired top-down, axially and interleaved. Slice thickness was 3.7 mm with 0.3 mm gap. The FOV was 256 x 256 mm, matrix size 64 x 64, resulting in 4 x 4 x 4 mm isometric voxels. TR was 1908 ms, echo time (TE) 21.6 ms, and the flip angle 75°. This unusual TR was chosen because it was the fastest TR possible in the context of the other parameter settings and therefore would generate the maximum amount of data. In addition, 1908 is a multiple of 36 and this simplifies determining the slice acquisition times and stimulus presentation times. In each run 255 volumes were collected and lasted 8 minutes and 6 seconds.

Separate high resolution T1-weighted images were acquired using the 3D FSPGR sequence: TI/TR/TE: 650/6.8/1.4 ms, flip angle = 12°, 196 slices, slice thickness 1 mm, matrix 256 x 256, voxel size = 1 x 1 x 1 mm.

Pre-processing

Only minimal data pre-processing was applied: Low frequency drifts were removed using a high pass filter at 0.01 Hz (Smith et al., 2004), and the data were motion corrected using FSL MCLFLIRT (Jenkinson et al., 2002). Note that in the context of the Slice-Based method, motion correction poses a certain challenge. This is because the Slice-Based method not only assumes that a given voxel samples the same physical brain area throughout the imaging run (as in all fMRI methods), but also assumes that this voxel is sampled at regular well-known time intervals. Motion correction may lead to the translation of a physical brain area across slices and therefore impact the time interval at which this brain area was sampled. To address this issue we implemented a method for motion correction that allowed for standard spatial motion correction and provided an additional temporal correction that updates the time intervals of the physical areas underlying the voxels that were sampled (see Janssen et al. (submitted) for further details). However, it should be pointed out that our data set did not include a sufficiently large amount of motion to accurately verify the efficacy of this motion correction method. Finally, a second data set was created that was slice-time corrected using the AFNI *3dTshift* using the first slice as a reference and the standard Fourier interpolation method. The STC function was applied before the motion correction. Spatial smoothing was not used in any of the data sets.

Comparison of activation maps

The activation map for the standard method was obtained using the GLM method implemented in FSL Feat (Jenkinson et al., 2012) and used the STC dataset described above. Precise picture naming onsets were extracted from the participant-specific Presentation logfiles. The rest periods were not explicitly modeled (Pernet, 2014). The expected HRF was modeled as a double gamma function with default parameters, pre-whitening was applied, and the temporal derivative was included in the GLM model. The analysis only included the first run of each participant. A final group level map was generated by performing a one-sample t-test on the standard space transformed beta-coefficient maps of each participant using FSL randomise (Winkler et al., 2014).

The activation maps for the Timepoint by Timepoint with STC and Slice-Based method were constructed by comparing signal intensities at a baseline timepoint between one TRs before and one TR after stimulus presentation to signal intensities at timepoints between one and 5 TRs following stimulus presentation (i.e., capturing the peak of the BOLD signal). As in the simulation

data, signal intensities were modeled using LME with the lme4 package in R. Signal intensities were modeled as a function of a fixed effect variable Timepoint (signal intensities at baseline versus timepoint 1). In addition, the variations that arise due to items were brought into the model by including a random intercept for Item (i.e., the formula was `lmer(Intensity~Time+(1|Item))`. This model therefore provides for each participant a single activation map that indicates how signal intensity increased or decreased at timepoint 1 relative to baseline. The group level maps for the Timepoint by Timepoint and Slice-Based methods were generated in the exact same way as for the standard method. Note that as we mentioned earlier, the Timepoint by Timepoint method is volume-based and therefore may introduce uncertainties in the signal extraction compared to the Slice-Based method. Note also that the same transformation matrices to normalize each participant's scanner space image to standard space were used between the three methods.

BOLD signal extraction from left motor cortex

BOLD signals were extracted from three adjacent slices in the left motor cortex that all showed strong involvement in the task. To identify the active voxels on adjacent slices in left motor cortex we first created a mask of each participant's left precentral gyrus using the lateralized Harvard-Oxford probabilistic atlas (Desikan et al., 2006). Any voxels included in the mask on the medial surface of the left hemisphere were removed as those regions are unlikely to be involved in primary motor cortex control of speech. Next, the voxel with the maximum t-value in the masked GLM signal detection map (see above) was identified for each participant, which corresponds to the voxel with the maximum t-value in the left motor cortex. This resulted in three times series from adjacent slices in left motor cortex that were strongly involved in the task. This set of three time series for every participant formed the input to the four techniques.

Specifically, the BOLD signals extracted from these three slices in left motor cortex were examined with the FIR without STC, FIR with STC, Timepoint by Timepoint with STC, and Slice-Based methods. The extraction was performed exactly as described above using the simulated data. Note no temporal derivates were included in the model. We examined the mean inter-slice correlation, the mean UP, the mean TTP, and the mean MAXT in the BOLD signal across the three slices for all participants. The TTP was calculated in the standard way as the mean timepoint at which the extracted BOLD signal at each slice reached its maximum value. The UP indicated the number of unique peaks found for each slice and was calculated by counting the number of unique TTPs found across slices. The Maximum t-value was the maximum t-value across all slices. We extracted the BOLD signal in the thee slices in left motor cortex at two temporal resolutions, the TR (1908 ms), and TR/2 (954 ms). We performed statistical comparisons of these values on a by-participant basis.

Results

Simulation - variability in BOLD onset

A graphical presentation of a single representative result from this simulation experiment is presented in Figure 4. A visual impression of this result suggested that the Slice-Based method yielded the BOLD signals in closest correspondence with the ground-truth signal (dashed line). Given the high value of the mean correlation with the ground-truth signal ($\bar{r}_1$) for all methods (suggesting ceiling effects), the analyses were focused on the mean absolute difference between the ground-truth signal and the mean extracted signal across all slices ($\bar{d}$). An overview of the means and statistics of the $\bar{d}$ value for each method across all simulations in presented in Figure 7.

For the simulation on the variability in the onset of the BOLD signal, the analyses revealed that for TR = 3, the lowest $\bar{d}$ values were observed for the Slice-Based method. Specifically, the Slice-Based method yielded a lower $\bar{d}$ value compared to FIR without STC ($F(1,198) = 3555.5$, $p < 0.0001$), FIR with STC ($F(1,198) = 1589.8$, $p < 0.0001$), and Timepoint by Timepoint with STC methods ($F(1,198) = 8.7$, $p < 0.004$). In addition, the Timepoint by Timepoint with STC differed from the FIR without STC ($F(1,198) = 4170.3$, $p < 0.0001$), and the FIR with STC methods ($F(1,198) = 1761.6$, $p < 0.0001$). Finally, the FIR with STC differed from the FIR without STC method ($F(1,198) = 747.1$, $p < 0.0001$).

Likewise, for TR = 1, the Slice-Based method again had the lowest $\bar{d}$ values. In particular, the Slice-Based method had lower $\bar{d}$ values than FIR without STC ($F(1,198) = 350.4$, $p < 0.0001$), FIR with STC ($F(1,198) = 67.4$, $p < 0.0001$), and Timepoint by Timepoint with STC methods ($F(1,198) = 5.3$, $p < 0.03$). And as before, the Timepoint by Timepoint with STC differed from the FIR without STC ($F(1,198) = 258.0$, $p < 0.0001$), and the FIR with STC methods ($F(1,198) = 27.9$, $p < 0.0001$). Finally, the FIR with STC differed from the FIR without STC method ($F(1,198) = 482.6$, $p < 0.0001$).

Simulation 2 - variability in BOLD dispersion

As with the previous simulation, a visual impression of Figure 5 suggested that the Slice-Based method was also superior under conditions of variability in peak BOLD signal dispersion. The statistical analyses revealed that for TR = 3, the Slice-Based method yielded a lower $\bar{d}$ value compared to FIR without STC ($F(1,198) = 2158.3$, $p < 0.0001$), FIR with STC ($F(1,198) = 269.8$, $p < 0.0001$), and Timepoint by Timepoint with STC methods ($F(1,198) = 35.1$, $p < 0.0001$). In addition, the Timepoint by Timepoint with STC differed from the FIR without STC ($F(1,198) = 1780.8$, $p < 0.0001$), and the FIR with STC methods ($F(1,198) = 94.3$, $p < 0.0001$). Finally, the FIR with STC differed from the FIR without STC method ($F(1,198) = 3480.4$, $p < 0.0001$).

Likewise, for TR = 1, the Slice-Based method also had lower $\bar{d}$ values than FIR without STC ($F(1,198) = 163.8$, $p < 0.0001$), FIR with STC ($F(1,198) = 74.2$, $p < 0.0001$), and Timepoint by Timepoint with STC methods ($F(1,198) = 26.5$, $p < 0.0001$). And as before, the Timepoint by Timepoint with STC differed from the FIR without STC ($F(1,198) = 43.7$, $p < 0.0001$), and marginally from the FIR with STC methods ($F(1,198) = 4.1$, $p = 0.05$). Finally, the FIR with STC differed from the fir without STC method ($F(1,198) = 189.7$, $p < 0.0001$).

Simulation 3 - impact of single spike

In line with previous simulations, Figure 6 suggested that the Slice-Based method extracted a more accurate BOLD signal when a single spike was present in the data for a single voxel. The statistical analyses revealed that regarding the $\bar{d}$ value, for TR = 3, the Slice-Based method differed from FIR without STC ($F(1,198) = 3941.4$, $p < 0.0001$), FIR with STC ($F(1,198) = 595.2$, $p < 0.0001$), but not from Timepoint by Timepoint with STC methods ($F(1,198) = 2.7$, $p = 0.10$). In addition, the Timepoint by Timepoint with STC differed from the FIR without STC ($F(1,198) = 3629.7$, $p < 0.0001$), and the FIR with STC methods ($F(1,198) = 489.3$, $p < 0.0001$). Finally, the FIR with STC differed from the FIR without STC method ($F(1,198) = 7673.5$, $p < 0.0001$).

Likewise, for TR = 1, the Slice-Based method also had lower $\bar{d}$ values than FIR without STC ($F(1,198) = 210.2$, $p < 0.0001$), FIR with STC ($F(1,198) = 30.2$, $p < 0.0001$), but not Timepoint by Timepoint with STC methods ($F(1,198) = 0.22$, $p = 0.63$). As previously, the Timepoint by Timepoint with STC differed from the FIR without STC ($F(1,198) = 121.3$, $p < 0.0001$), and from the FIR with STC methods ($F(1,198) = 14.5$, $p < 0.0002$). Finally, the FIR with STC differed from the fir without STC method ($F(1,198) = 883.4$, $p < 0.0001$).

In-vivo Results

Comparison of activation maps

Figure 8 presents the comparison of the whole-brain group analysis of overt picture naming using the standard GLM with STC obtained with FSL Feat (panel A), a map obtained using the Timepoint by Timepoint with STC method (panel B), and that of the Slice-Based method (panel C). Inspection of the difference maps (Figure 8, panels D, E, F, and G), revealed that the Slice-Based method yielded substantially higher t-values in some areas of the brain such as the medial frontal cortex compared to the other two methods (see Figure 8 panels D and E). Quantitative comparisons confirmed these visual impressions. The Pearson's correlation coefficient between the non-thresholded and vectorized versions of the Slice-Based and GLM maps was high at 0.95 ($p < 0.001$) and between the Slice-Based and Timepoint by Timepoint map was 0.94 ($p < .001$), suggesting comparable activation patterns. However, the Dice index (see black line Figure 9), revealed that at higher t-value thresholds, the similarity between the Slice-Based and GLM map (Figure 9 panel A) and between the Slice-Based and Timepoint by Timepoint map (Figure 9 panel B) decreased. Further examination revealed that this decrease was due to the presence of more active voxels in the Slice-Based map at higher t-value thresholds (see gray line Figure 9), indicating improved signal detection for the Slice-Based method.

BOLD signal extraction from left motor cortex

Figure 10 provides an overview of the extracted BOLD signals for the four methods from the three adjacent slices in left motor cortex for three representative participants from a single imaging run with 18 stimuli at the TR temporal resolution of 1.908 s. Figure 11 relied on the same data but BOLD signals were extracted at twice the temporal resolution (0.954 s). Overall the visual impression is that the Slice-Based method extracted the BOLD signal with increased t-values. A graphical overview of the means and statistical differences for each method can be seen in Figure 12. Statistical analyses confirmed these visual impressions. Specifically, for the TR = 1.908 s, the interslice correlation was highest for the FIR with STC (mean = 0.76) and differed from the FIR without STC (0.69; $t(29) = 4.2$, $p < 0.0003$) and Slice-Based methods (0.73; $t(29) = 3.1$, $p < 0.004$). For the TR = 0.954 s, Slice-Based and FIR with STC had

comparable interslice correlations. In addition, the interslice correlation of the FIR with STC model differed from the Timepoint by Timepoint (0.62; t(29) = 2.4, p ¡ 0.03) and FIR without STC (0.59; t(29) = 4.7, p ¡ 0.0001) methods.

Regarding the mean unique number of peaks across each slice, for TR = 0.954 s, the Slice-Based method had the highest number of unique peaks (2.6), and differed from the Timepoint by Timepoint STC (2.3; t(29) = 2.1, p ¡ 0.05), the FIR with STC (2.1; t(29) = 3.3, p ¡ 0.003), and the FIR without STC (2.3; t(29) = 2.6, p ¡ 0.02) methods.

For the mean time to peak, at TR = 0.954 s, the Timepoint by Timepoint model had the earliest time to peak (4.0), and differed from the FIR without STC method (4.4; t(29) = 2.2, p ¡ 0.04).

For the mean maximum t-value observed across all slices, at TR = 1.908 s, the Slice-Based method had the higest mean maximum t-value (9.2), and differed from the Timepoint by Timepoint with STC (8.6; t(29) = 2.7, p ¡ 0.02), the FIR with STC (8.4; t(29) = 2.2, p ¡ 0.04), and FIR without STC (7.9; t(29) = 3.7, p ¡ 0.001) methods. In addition, the FIR without STC had a lower maximum t-value than the FIR with STC (t(29) = 4.4, p ¡ 0.0002). Similarly, at TR = 0.954 s, the Slice-Based method again had the higest maximum t-value (8.7), and differed from the Timepoint by Timepoint with STC (7.8; t(29) = 3.1, p ¡ 0.005), the FIR with STC (6.7; t(29) = 5.6, p ¡ 0.0001), and without STC (6.4; t(29) = 6.4, p ¡ 0.0001) methods. In addition, the Timepoint by Timepoint method differed from both FIR with STC (t(29) = 3.4, p ¡ 0.003) and without STC (t(29) = 4.2, p ¡ 0.0003), and the FIR with STC differed from the FIR without STC (t(29) = 3.5, p ¡ 0.002) methods.

Finally, the above results give the impression that the Slice-Based method was more robust against increased in temporal resolution compared to the FIR based methods. Further analyses confirmed this impression: For the FIR with STC method, the statistical difference in maximum t-value between TR and TR/2 was larger (t(59) = -9.11, p ¡ 0.0001) than for the Slice-Based method (t(59) = -2.55, p ¡ 0.02), suggesting that the Slice-Based method indeed provides better BOLD signal extraction accuracy under conditions of increased temporal resolution than the FIR based methods.

Discussion

The current study evaluated four methods that differed in how the BOLD signal can be extracted from fMRI data. These four methods differed in how volumes were created from individually acquired brain slices, and in how the statistical extraction of the BOLD signal takes place. Specifically, whereas the FIR without STC, FIR with STC, and Timepoint by Timepoint with STC methods all relied on the same standard volume creation technique (see Figure 1), the Slice-Based method relied on a new volume creation technique that does not require STC (see Figure 2). In addition, statistical extraction of the BOLD signal relied on the standard FIR with GLM approach (i.e., the FIR without and with STC methods), or a non-standard Timepoint by Timepoint approach using LME modeling (i.e., the Timepoint by Timepoint with STC, and Slice-Based methods). Results from three simulation experiments revealed that the best performance was observed for the Slice-Based method (mean increased accuracy in terms of $\bar{d}$ of Slice-Based vs FIR with STC for TR 3 was 32%, for TR 1 it was 13%). Similarly, analyses of real-world data revealed that the Slice-Based method yielded dramatically improved signal detection in group level maps (at a t-value threshold of 8, the Slice-Based map contained about 3 times more active voxels compared to the standard GLM map, and 8 times more active voxels compared to the Timepoint by Timepoint map), as well as higher maximum t-values of the extracted BOLD signal in the motor cortex (mean increased accuracy in terms of maximum t-value of Slice-Based vs FIR with STC for TR resolution was 9%, for TR/2 it was 24%). Overall, the Slice-Based method significantly improved BOLD signal extraction accuracy compared to the standard methods.

Before discussing the main implications, two aspects of these data deserve further scrutiny. First, in the real-world data, the FIR with STC method yielded a higher inter-slice correlation ($\bar{r}_2$) at the TR resolution and a lower number of unique peaks (UP) at the TR/2 resolution than the Slice-Based method (see Figure 12). This may suggest that the FIR with STC method yielded more coherency in the BOLD signals extracted from different slices than the Slice-Based method. However, note that this increase in similarity between signals on adjacent slices was directly caused by the STC function that effectively smoothed the data, which led to more similar signals on adjacent slices (e.g., Parker et al., 2017; Sladky et al., 2011). However, it is not obvious that these more similar signals reflect the veridical underlying signal. Indeed, although the simulation studies revealed that STC increased the accuracy with which the ground-truth signal was extracted, these studies also revealed that this accuracy of signal extraction in the FIR with STC method was relatively poor when compared to the accuracy of the Slice-Based method (see Figure 7). Thus, whereas the FIR with STC method will pro-

duce signals that appear similar across adjacent slices, this does not necessarily imply more accurate signal detection, and the simulation results indicated BOLD signals were extracted more accurately with the Slice-Based method.

Another aspect of the data that deserves further mention is that whereas signal detection accuracy greatly improved due to the Timepoint by Timepoint statistical method, additional improvements due to the slice-based method of volume creation were seemingly more modest (comparing FIR with STC, Timepoint by Timepoint with STC to Slice-Based methods, see Figure 7D). Given that the Slice-Based technique comprises both the Timepoint by Timepoint technique and the new volume creation method, one possible conclusion from these results is therefore that the main contribution of the Slice-Based framework lies primarily in the statistical modeling technique and not in the new method of volume creation. However, this conclusion would be premature for the following three reasons: First, note that in the simulation experiments, although improvements between the Timepoint by Timepoint and Slice-Based methods were modest, they were statistically significant (Figure 7D). In addition, in the real-world data, there were relatively large differences between the Timepoint by Timepoint and the Slice-Based methods in terms of the maximum extracted t-value (especially at 954 ms temporal resolution, see Figure 12D). Finally, as mentioned earlier, epoching of data is more precise in the Slice-Based framework, leading to more accurate allocation of datapoints to timepoints in the Slice-Based method compared to the Timepoint by Timepoint method. Additional simulations presented in the Supplementary Materials (see Figure S2 and S3) further highlight conditions in which the Timepoint by Timepoint with STC method yielded inaccurate signal detection (and may explain the differences observed in Figure 12D). Overall, these observations suggest that the new method of Slice-Based volume creation confers additional advantages of BOLD signal detection over and above the advantages already accrued by the Timepoint by Timepoint method of statistical signal extraction.

How did the new volume creation technique improve BOLD signal extraction accuracy? There are two major differences between the new and the standard method of volume creation. First, as mentioned in the Introduction, in the new method, whole-brain volumes contain slices with signals that are temporally aligned, meaning that there are no within volume temporal distortions of signals across adjacent slices. By contrast, in the standard method these signals are not temporally aligned, meaning that there are within volume temporal distortions of signals across slices (cf., Figures 1 and 2). The important implication of the presence of temporally aligned signals across slices within consecutive volumes is that the STC procedure is no longer required. Removing the STC procedure from the fMRI processing pipeline means that less data transformations are required, and this means that in the new method a more accurate and veridical BOLD signal can be extracted. Second, the new volume creation method leads to a reorganization of the fMRI data that fundamentally differs from the standard method of volume creation. Specifically, the new method reorganizes the data into stimulus-locked epochs where each datapoint in this epoch is labeled with timing information that is determined by the exact slice acquisition times (on the order of tens of milliseconds). By contrast, in the standard method, datapoints are labeled with timepoints determined by the TR (on the order of hundreds to thousands of milliseconds). This means that in the Slice-Based volume creation technique, timing information is much more precise. In our data, the availability of this highly precise timing information enabled statistical modeling approaches that improved the accuracy of BOLD signal extraction. Thus, the new volume creation technique improved the accuracy of BOLD signal extraction because it removes the need for STC and it includes more precise timing information in the data.

Besides the new volume creation technique, the current study also introduced a new method of statistical BOLD signal extraction. Specifically, the modeling approach used in the current study relied on a Timepoint by Timepoint technique. While the method of comparing timepoints in an epoch to a baseline timepoint has been used in previous fMRI studies (e.g., Leung et al., 2000), the specific implementation that uses LME modeling proposed here is new. This modeling approach improved BOLD signal extraction over the standard FIR based GLM approach in three ways. First, in the Timepoint by Timepoint technique, model coefficients are estimated by comparing signal intensities at all timepoints in an epoch to a baseline timepoint (e.g., data collected prior to stimulus onset). The direct comparison of signal intensities between a baseline and a timepoint means that more data is available for coefficient estimation compared to a no baseline FIR model (as used here), avoids problems associated with subtracting model coefficients in FIR models that use a baseline, and avoids the need for explicit baseline (rest) periods in the experimental design. This leads to more stable model estimates in BOLD signal extraction. In addition, the specific statistical modeling technique used in the current study was LME modeling implemented in

R (Bates, 2005; Pinheiro & Bates, 2000). A major advantage of LME modeling over the standard GLM concerns the use of more complex random effect structures (Westfall et al., 2016). For example, the current study included random-intercepts for the effect of item in the statistical model. The inclusion of known sources of variance in the model leads to more stable model estimates, and improved signal extraction. Finally, because in the Timepoint by Timepoint approach separate models are fitted at each timepoint, the standard errors are less sensitive to sudden noise events at other timepoints in the epoch (Friston et al., 1996), and hence, lead to more stable estimates. In short, the Timepoint by Timepoint approach for statistical signal extraction improved the accuracy of BOLD signal extraction over the standard GLM based methods.

As mentioned in the Introduction, current advances in MRI pulse sequence design and hardware coils have enabled the acquisition of fMRI data at very short TRs. For example, several studies have shown that with Simultaneous Multi-Slice acquisition methods (SMS), whole-brain volumes can be acquired at TRs of 500 ms and less (Chen et al., 2015; Feinberg & Setsompop, 2013; Vu et al., 2016). Interestingly, although reductions in TR are generally associated with lower SNR, studies have demonstrated that the TR reductions yielded increased statistical power, suggesting that the additional data offset the reductions in SNR (Constable & Spencer, 2001; Sahib et al., 2016). The Slice-Based method proposed here may be combined with these existing SMS techniques to further improve BOLD signal detection. In practical terms, this combination is straightforward since all that the Slice-Based method requires is a list of the specific timepoints at which each slice was acquired. Although the impact of temporal distortions produced by time-shifting slices is expected to be reduced with short TRs in SMS techniques (Sahib et al., 2016), further minimizing the impact of such distortions can be achieved with the Slice-Based method. In addition, the Slice-Based method has the advantage of not requiring state-of-the-art MRI pulse sequences or hardware, but can extract BOLD signals with high temporal resolution and accuracy at long TRs with good SNR using run-of-the-mill pulse sequences. In short, the Slice-Based method could improve signal detection for existing cutting-edge techniques, as well as provide a method for BOLD signal extraction with high temporal resolution and accuracy that is not restricted by specific hardware configurations.

The current study has several limitations. First, as explained in the Method section, although head motion is considered a problem for all fMRI techniques, it poses special challenges to Slice-Based fMRI. Although the results presented here relied on the standard procedures implemented in FSL to alleviate head motion artefacts (i.e., MCFlirt), they were accompanied by specialized Slice-Based fMRI procedures developed to alleviate any temporal distortions induced by head motion (see Janssen et al., submitted for details). Although these specialized Slice-Based motion correction procedures improved signal detection (Janssen et al., submitted), further validation of these procedures is required in dedicated future studies. Second, the current study relied on real-world fMRI data collected from a picture naming task. We are currently planning studies that use different tasks in which expectations about BOLD signal dynamics are more easily verified, such as, for example, inducing delayed BOLD responses between left and right motor cortex (e.g., Menon et al., 1998; Miezin et al., 2000). Third, although the Slice-Based technique works with both slow event-related and blocked designs, it currently does not work for fast event-related designs. Note, however, that fast event-related designs pose challenges to all linear modeling approaches (e.g., Vazquez & Noll, 1998). Finally, the Slice-Based framework requires more statistical tests than the standard FIR based approaches. In the Slice-Based framework, a statistical model is fitted to each voxel in the brain *at each timepoint*, and therefore raises issues about multiple comparisons. In our data we dealt with this issue by using high t-value thresholds. In the future, this issue may be addressed by implementing 4D cluster correction techniques such as those available for EEG and MEG data (Guthrie & Buchwald, 1991; Maris & Oostenveld, 2007). To facilitate further development on all these issues an example data set and tutorial script for analyzing an fMRI data set within the Slice-Based framework is available at `https://github.com/iamnielsjanssen/slice-based`. In brief, we think that despite these limitations the Slice-Based technique is a feasible alternative for the analysis of fMRI data.

To conclude, current methods for extracting BOLD signals rely on volumes constructed by time-shifting slices, STC, and FIR-based statistical modeling techniques (Josephs et al., 1997). Here we show that such an approach leads to a relatively poor detection of the true underlying BOLD signal. By contrast, the Slice-Based framework that was proposed here relies on a fundamentally different method of volume construction, does not require STC, and uses non-standard Timepoint by Timepoint modeling for signal extraction. The results from both simulated and real-world data showed that this new fMRI data-analytic method led to im-

proved BOLD signal detection accuracy. Although the Slice-Based framework places more strict constraints on paradigm design, the new framework is a relatively simple data-analytic method that does not require specialized MRI hardware, is more robust against trial-by-trial variability in BOLD onset and shape, is more robust against sudden noise events, allows for an easy application of modern LME modeling techniques, and can be applied to a wide variety of fMRI research contexts that rely on both event-related and blocked designs. Ultimately we think that further improvements in BOLD signal detection will come from a combination of the Slice-Based framework with newer hardware oriented techniques.

Bates, D. M. (2005). Fitting linear mixed models in r. *R News*, *5*, 27–30.

Bedny, M., Aguirre, G. K., & Thompson-Schill, S. L. (2007). Item analysis in functional magnetic resonance imaging. *Neuroimage*, *35*, 1093–1102.

Calhoun, V., Golay, X., & Pearlson, G. (2000). Improved fmri slice timing correction: interpolation errors and wrap around effects. In *Proceedings, ISMRM, 9th annual meeting, Denver* (p. 810).

Chang, W.-T., Nummenmaa, A., Witzel, T., Ahveninen, J., Huang, S., Tsai, K. W.-K., Chu, Y.-H., Polimeni, J. R., Belliveau, J. W., & Lin, F.-H. (2013). Whole-head rapid fmri acquisition using echo-shifted magnetic resonance inverse imaging. *Neuroimage*, *78*, 325–338.

Chen, L., Vu, A., Xu, J., Moeller, S., Ugurbil, K., Yacoub, E., & Feinberg, D. (2015). Evaluation of highly accelerated simultaneous multi-slice epi for fmri. *Neuroimage*, *104*, 452–459.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, *386*, 604.

Cohen, M. S., & Weisskoff, R. M. (1991). Ultra-fast imaging. *Magnetic resonance imaging*, *9*, 1–37.

Constable, R. T., & Spencer, D. D. (2001). Repetition time in echo planar functional mri. *Magnetic resonance in medicine*, *46*, 748–755.

Dale, A. M. (1999). Optimal experimental design for event-related fmri. *Human brain mapping*, *8*, 109–114.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T. et al. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, *31*, 968–980.

Dilharreguy, B., Jones, R. A., & Moonen, C. T. (2003). Influence of fmri data sampling on the temporal characterization of the hemodynamic response. *Neuroimage*, *19*, 1820–1828.

Donnet, S., Lavielle, M., & Poline, J.-B. (2006). Are fmri event-related response constant in time? a model selection answer. *Neuroimage*, *31*, 1169–1176.

Duann, J.-R., Jung, T.-P., Kuo, W.-J., Yeh, T.-C., Makeig, S., Hsieh, J.-C., & Sejnowski, T. J. (2002). Single-trial variability in event-related bold signals. *Neuroimage*, *15*, 823–835.

Efron, B., & Morris, C. (1977). Stein's paradox in statistics. *Scientific American*, *236*, 119–127.

Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Glasser, M. F., Miller, K. L., Ugurbil, K., & Yacoub, E. (2010). Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging. *PloS one*, *5*, e15710.

Feinberg, D. A., & Setsompop, K. (2013). Ultra-fast mri of the human brain with simultaneous multi-slice imaging. *Journal of magnetic resonance*, *229*, 90–100.

Friston, K., Williams, S., Howard, R., Frackowiak, R., & Turner, R. (1996). Movement-related effects in fmri time-series. *Magnetic resonance in medicine*, *35*, 346–355.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in cognitive sciences*, *10*, 14–23.

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology*, *28*, 240–244.

Henson, R., Buechel, C., Josephs, O., & Friston, K. (1999). The slice-timing problem in event-related fmri. *NeuroImage*, *9*, 125.

Janssen, N., Hernández-Cabrera, J. A., & Ezama Foronda, L. (submitted). Data on motion correction in slice-based functional magnetic resonance imaging. *Data in Brief*, .

Janssen, N., Hernández-Cabrera, J. A., van der Meij, M., & Barber, H. A. (2014). Tracking the time course of competition during word production: Evidence for a post-retrieval mechanism of conflict resolution. *Cerebral Cortex*, (p. bhu092).

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*, 825–841.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *Neuroimage*, *62*, 782–790.

Josephs, O., Turner, R., & Friston, K. (1997). Event-related fmri. *Human brain mapping*, *5*, 243–248.

Lage-Castellanos, A., Martínez-Montes, E., Hernández-Cabrera, J. A., & Galán, L. (2010). False discovery rate and permutation test: an evaluation in erp data analysis. *Statistics in medicine*, *29*, 63–74.

Leung, H.-C., Skudlarski, P., Gatenby, J. C., Peterson, B. S., & Gore, J. C. (2000). An event-related functional mri study of the stroop color word interference task. *Cerebral cortex*, *10*, 552–560.

Lin, F.-H., Wald, L. L., Ahlfors, S. P., Hämäläinen, M. S., Kwong, K. K., & Belliveau, J. W. (2006). Dynamic magnetic resonance inverse imaging of human brain function. *Magnetic resonance in medicine*, *56*, 787–802.

Lindquist, M. A., Loh, J. M., Atlas, L. Y., & Wager, T. D. (2009). Modeling the hemodynamic response function in fmri: efficiency, bias and mis-modeling. *Neuroimage*, *45*, S187–S198.

Maccotta, L., Zacks, J. M., & Buckner, R. L. (2001). Rapid self-paced event-related functional mri: feasibility and implications of stimulus-versus response-locked timing. *Neuroimage*, *14*, 1105–1121.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of eeg-and meg-data. *Journal of neuroscience methods*, *164*, 177–190.

Menon, R. S., Luknowsky, D. C., & Gati, J. S. (1998). Mental chronometry using latency-resolved functional mri. *Proceedings of the National Academy of Sciences*, *95*, 10902–10907.

Miezin, F. M., Maccotta, L., Ollinger, J., Petersen, S., & Buckner, R. (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.

Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E., Strupp, J., Harel, N., & Uğurbil, K. (2010). Multiband multislice ge-epi at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fmri. *Magnetic Resonance in Medicine*, *63*, 1144–1153.

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*, *87*, 9868–9872.

Ollinger, J., Shulman, G. L., & Corbetta, M. (2001). Separating processes within a trial in event-related functional mri: I. the method. *Neuroimage*, *13*, 210–217.

Parker, D., Liu, X., & Razlighi, Q. R. (2017). Optimal slice timing correction and its interaction with fmri parameters and artifacts. *Medical image analysis*, *35*, 434–445.

Pernet, C. R. (2014). Misconceptions in the use of the general linear model applied to functional mri: a tutorial for junior neuroimagers. *Frontiers in neuroscience*, *8*.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed effects models in S and S-PLUS*. Springer.

Price, C. J., Veltman, D., Ashburner, J., Josephs, O., & Friston, K. (1999). The critical relationship between the timing of stimulus presentation and data acquisition in blocked designs with fmri. *NeuroImage*, *10*, 36–44.

Sahib, A., Mathiak, K., Erb, M., Elshahabi, A., Klamer, S., Scheffler, K., Focke, N., & Ethofer, T. (2016). Effect of temporal resolution and serial autocorrelations in event-related functional mri. *Magnetic resonance in medicine*, *76*, 1805.

Serences, J. T. (2004). A comparison of methods for characterizing the event-related bold timeseries in rapid fmri. *Neuroimage*, *21*,

1690–1700.
Sladky, R., Friston, K. J., Tröstl, J., Cunnington, R., Moser, E., & Windischberger, C. (2011). Slice-timing effects and their correction in functional mri. *Neuroimage*, *58*, 588–594.
Smith, N. J., & Kutas, M. (2015). Regression-based estimation of erp waveforms: Ii. nonlinear effects, overlap correction, and practical considerations. *Psychophysiology*, *52*, 169–181.
Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E. et al. (2004). Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, *23*, S208–S219.
Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., Lu, C. C., Pechmann, T., Pléh, C., Wicha, N. et al. (2004). A new on-line resource for psycholinguistic studies. *Journal of memory and language*, *51*, 247–250.
Toni, I., Schluter, N. D., Josephs, O., Friston, K., & Passingham, R. E. (1999). Signal-, set-and movement-related activity in the human brain: an event-related fmri study. *Cerebral cortex*, *9*, 35–49.
Vazquez, A. L., & Noll, D. C. (1998). Nonlinear aspects of the bold response in functional mri. *Neuroimage*, *7*, 108–118.
Vu, A. T., Phillips, J. S., Kay, K., Phillips, M. E., Johnson, M. R., Shinkareva, S. V., Tubridy, S., Millin, R., Grossman, M., Gureckis, T. et al. (2016). Using precise word timing information improves decoding accuracy in a multiband-accelerated multimodal reading experiment. *Cognitive Neuropsychology*, *33*, 265–275.
Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G., & Rosseel, Y. (2011). neurosim: An r package for generating fmri data. *Journal of Statistical Software*, *44*, 1–18.
Westfall, J., Nichols, T., & Yarkoni, T. (2016). Fixing the stimulus-as-fixed-effect fallacy in task fmri. *bioRxiv*, (p. 077131).
Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *Neuroimage*, *92*, 381–397.
Witt, S. T., Warntjes, M., & Engström, M. (2016). Increased fmri sensitivity at equal data burden using averaged shifted echo acquisition. *Frontiers in Neuroscience*, *10*.
van der Zwaag, W., Francis, S., & Bowtell, R. (2006). Improved echo volumar imaging (evi) for functional mri. *Magnetic resonance in medicine*, *56*, 1320–1327.

Appendix 1

Formally, fMRI data D can be represented as a set of m slices S that are repeatedly sampled n times:

$$D = [S_{1,1}, ..., S_{m,n}], \tag{7}$$

where each S is itself a two dimensional matrix of acquired fMRI signal intensities (not shown here). This data matrix of slices D is accompanied by a similar size $m \times n$ matrix of slice acquisition times DT.

$$DT = [t_{1,1}, ..., t_{m,n}]. \tag{8}$$

Under the assumption of a standard sequential slice acquisition scheme, each specific time point $t(a, b)$ in this matrix can be determined by the following function:

$$t(a, b) = \frac{TR}{m} \times (a + (b - 1) \times m), \tag{9}$$

where a and b index the specific slice and acquisition number.

In the standard way to create whole-brain volumes, raw fMRI data D is transformed from $m \times n$ individual slices, to a n size vector D' of whole-brain volumes $V_1, ..., V_n$:

$$[S_{1,1}, ..., S_{m,n}] \rightarrow [V_1, ..., V_n]. \tag{10}$$

In this new formulation of the data D', it is simply assumed that all slices within a given volume are acquired at the same point in time given by:

$$D'T = [tv_1, ..., tv_n], \tag{11}$$

where each volume acquisition time $tv(v)$ is determined by the function:

$$tv(v) = TR \times v, \tag{12}$$

where v ranges from 1 to n.

In the Slice-Based method, volume-creation requires a set P of m stimuli $[p_1, ..., p_m]$, whose corresponding stimulus presentation times PT coincide precisely with the slice acquisition times determined by equation 8:

$$PT = [pt_{1,a}, ..., pt_{m,y}] = [t_{1,a}, ..., t_{m,y}], \tag{13}$$

Next, we create m epochs $E_1, ..., E_m$ corresponding to each stimulus presentation. Each epoch has length Δt. A given epoch E_j then contains raw fMRI signal intensities as defined as the set of slices:

$$E_j = [S_{j,a}, ..., S_{k,d}], \tag{14}$$

where j correspond to the slice acquired during stimulus presentation, and k to the slice acquired at the end of an epoch. The corresponding set of slice acquisition times for an epoch is:

$$ET_j = [t_{j,a}, ..., t_{k,d}], \tag{15}$$

where each specific time point in this set is determined by Equation 8. Next, for each given epoch E_j we compute the relative time difference RET_j between the exact presentation time of the stimulus $pt(j, a)$ and each time point in the epoch:

$$RET_j = [t_{j,a} - pt_{j,a}, ..., t_{k,d} - pt_{j,a}] = [rt_{j,a}, ..., rt_{k,d}]. \tag{16}$$

Following this step, we create a single epoch L with r whole-brain volumes

$$L = [V_1, ..., V_r], \tag{17}$$

where r is determined by the ratio between the epoch length Δt and the slice sampling frequency $\frac{TR}{m}$. The corresponding vector of volume acquisition times LT is determined by

$$LT = [lt_1, ..., lt_r], \tag{18}$$

where each lt is determined by the function:

$$lt(v) = \frac{TR}{m} \times v. \tag{19}$$

Each volume in the epoch L contains slices that are acquired at the same time point relative to the onset of the stimulus. This is achieved by combining slices from different epochs $E_1, ..., E_m$ on the basis of their RET values. Specifically slices $1, ..., m$ can be combined into a whole-brain volume if their corresponding relative times rt match. For a given volume:

$$V_e = [S_{1,a}, S_{2,d}, ..., S_{m,y}] \iff rt_{1,a} = rt_{2,d} = ... = rt_{m,y}. \tag{20}$$

This then leads to an epoch of whole-brain volumes that do not contain any temporal distortions, and where volumes are available at a temporal resolution equal to the sampling frequency. Finally, note that binning across timepoints may be used to improve the SNR. In this case, the temporal resolution is determined by the ratio between the epoch length and the number of bins.

Appendix 2

The particular statistical modeling technique that is used in the Slice-Based framework is called Linear Mixed Effect modeling (LME; Pinheiro & Bates, 2000). LME modeling is best seen as an extension of the GLM, meaning that both techniques are multiple linear regression techniques. However, LME modeling offers a number of advantages over the GLM technique. First, the LME technique is a multilevel technique which can handle complex random effect structures. This means that, for example, fMRI data can be modeled while taking into account random variability between participants and items, instead of just between participants as is currently the standard. Second, the LME technique can handle unbalanced data sets. This means experimental conditions can have missing values, and model estimation will still work. Finally, the LME parameter estimation takes into account a statistical phenomenon known as shrinkage (Efron & Morris, 1977). This means that parameter estimates are adjusted in the context of what the model knows about the other estimates, thereby preventing overfitting of the model. Overall, these features improve model estimation and reduce the probability of biased parameter estimates.

One way to formally present a LME model is as follows:

$$\begin{aligned} Y &= X\beta + Z\gamma + \epsilon, \\ \gamma &\sim N_r(0, G), \\ \epsilon &\sim N_N(0, R), \end{aligned} \tag{21}$$

where Y is a $N \times 1$ vector of fMRI signal intensities at a particular voxel p, at a particular time point t, X is the design matrix of size $(N \times (1+k))$, β are the estimates of the fixed effects $((1+k)\times 1)$, Z is the design matrix of the random effect predictors with g clusterings (e.g., items) and r random effects $(N \times (r \times g))$, γ the random effect estimates $(r \times g) \times 1$, and ϵ the residual errors. The γ parameter is assumed to be a random variable chosen from a normal distribution with a mean of zero and a covariance matrix G, while the residual error ϵ is assumed to be normally distributed with mean of zero and covariance matrix R. The unknown model parameters for β, G, and R can be estimated using Maximum Likelihood or Restricted Maximum Likelihood methods. The parameter γ is not a parameter of the model but its values are simply derived once the other parameters have been discovered.

Acknowledgements

See `https://github.com/iamnielsjanssen/slice-based` for a full analysis script using the Slice-Based method. This work was supported by The Spanish Ministry of Economy and Competitiveness (RYC2011-08433 and PSI2013-46334 to NJ). The authors would like to thank Cristin Modroo (SEGAI, Universidad de La Laguna) and Jos Mara Prez Gonzlez for technical support, and Sara Duque Gonzlez, Rebeca de Luis Sosa, and Alba Rodrguez Gonzlez for help with data collection. Correspondence and requests for materials should be addressed to NJ (`njanssen@ull.es`).

Figures

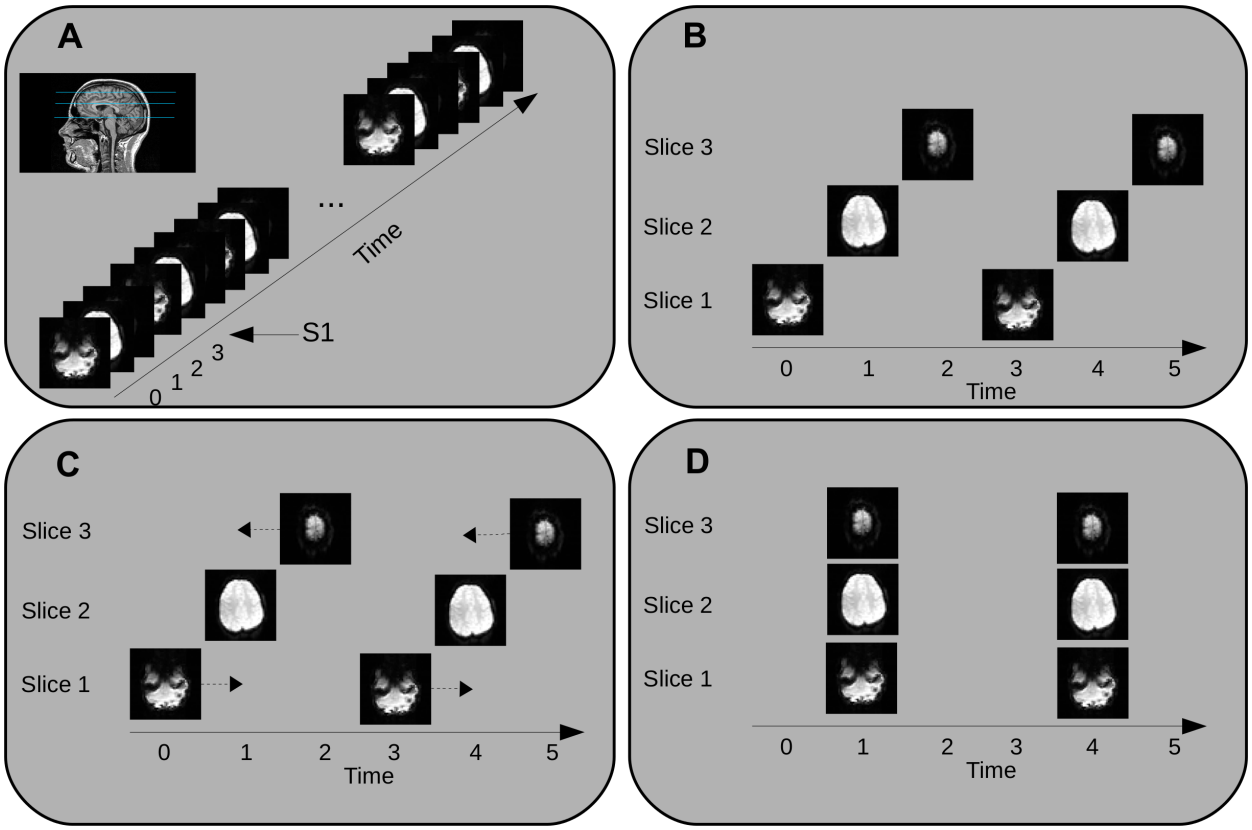

Figure 1: Current standard method for creating whole-brain volumes from raw fMRI data. Panel A shows an imaging run where a set of three slices are sequentially sampled at well defined points in time. Panel B reveals the same data, reorganized to illustrate that at no sampled time point information from the whole-brain is available, requiring data transformation. Panel C shows the standard solution, where slices are time-shifted to new positions in time (arrows indicate shift direction), using the middle slice as an arbitrary reference. Panel D shows the final transformed data, where whole-brain volumes are available every TR. Note how the final volumes contain slices acquired at different points in time, and how time points where data was sampled are no longer used.

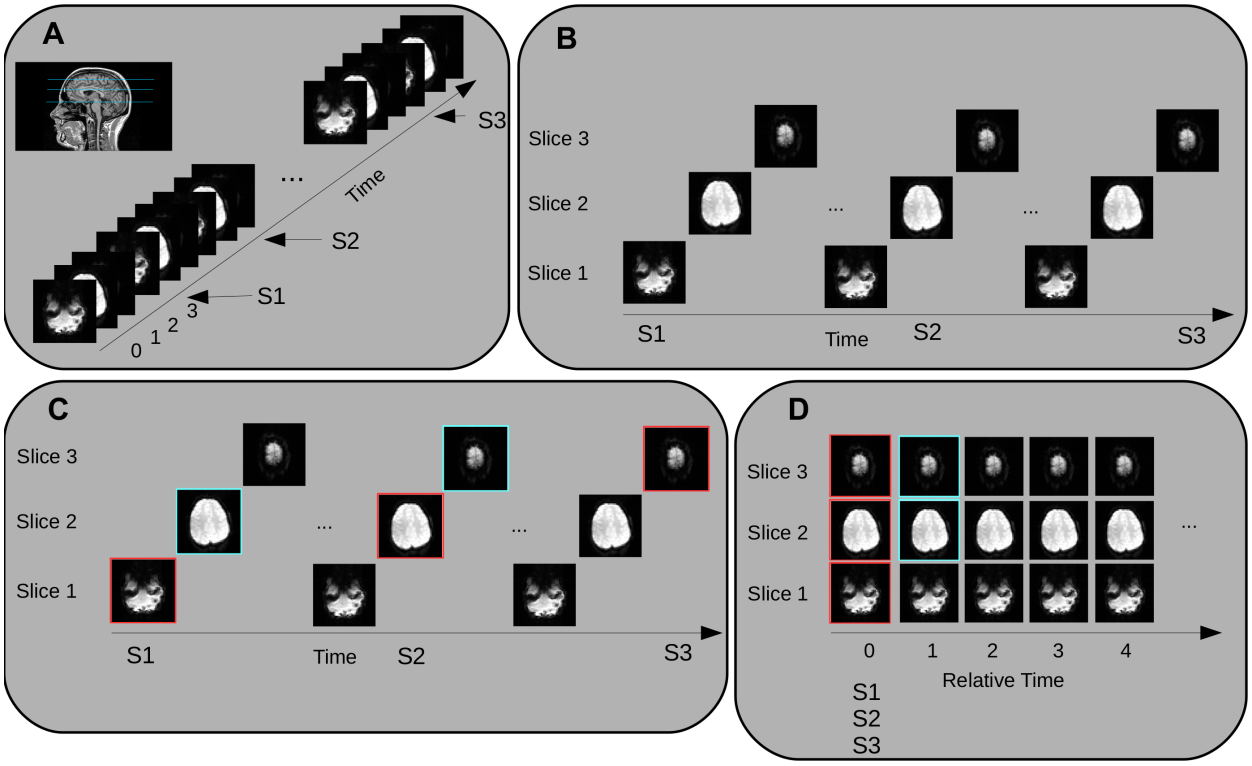

Figure 2: Slice-based method for creating whole-brain volumes from raw fMRI data. Panel A shows an imaging run where again three slices are sampled sequentially. Three stimuli S1, S2, and S3 of the same experimental class are presented during the run. Panel B shows that these stimuli are presented in-phase with slice acquisitions: S1 is presented in-phase with acquisition of slice 1, S2 with slice 2, and S3 with slice 3. Panel C shows how whole-brain volumes are created. Slices acquired at the same point in time relative to the onset of a stimulus can be combined (e.g., those highlighted in red and magenta). Panel D shows the final transformed data, where whole-brain volumes are available that only contain slices that are acquired at the same moment in time relative to a presented stimulus, and where whole-brain volumes are available at the sampling frequency (here TR/3).

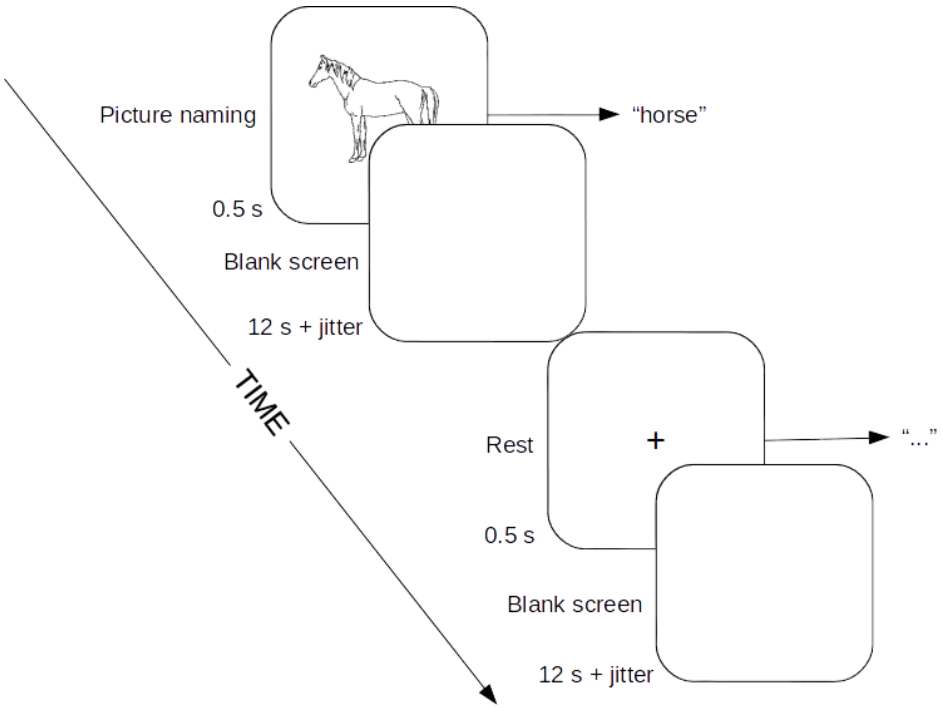

Figure 3: Temporal structure of the picture naming task used in the experiment. Stimuli consisted of either a picture or a fixation point that was presented for 0.5 s. Each stimulus presentation was followed by a blank screen that lasted for 12 s plus an additional jitter period. The jitter period was randomly selected without replacement from a uniform distribution of times that coincided with the slice acquisition times and ranged from 0 to 1855 ms in steps of 53 ms (see text for further details). Participants were instructed to name aloud presented pictures and remain quiet (i.e., rest) for presented fixation points. The order of stimuli presentation was fully randomized, and was different for every participant.

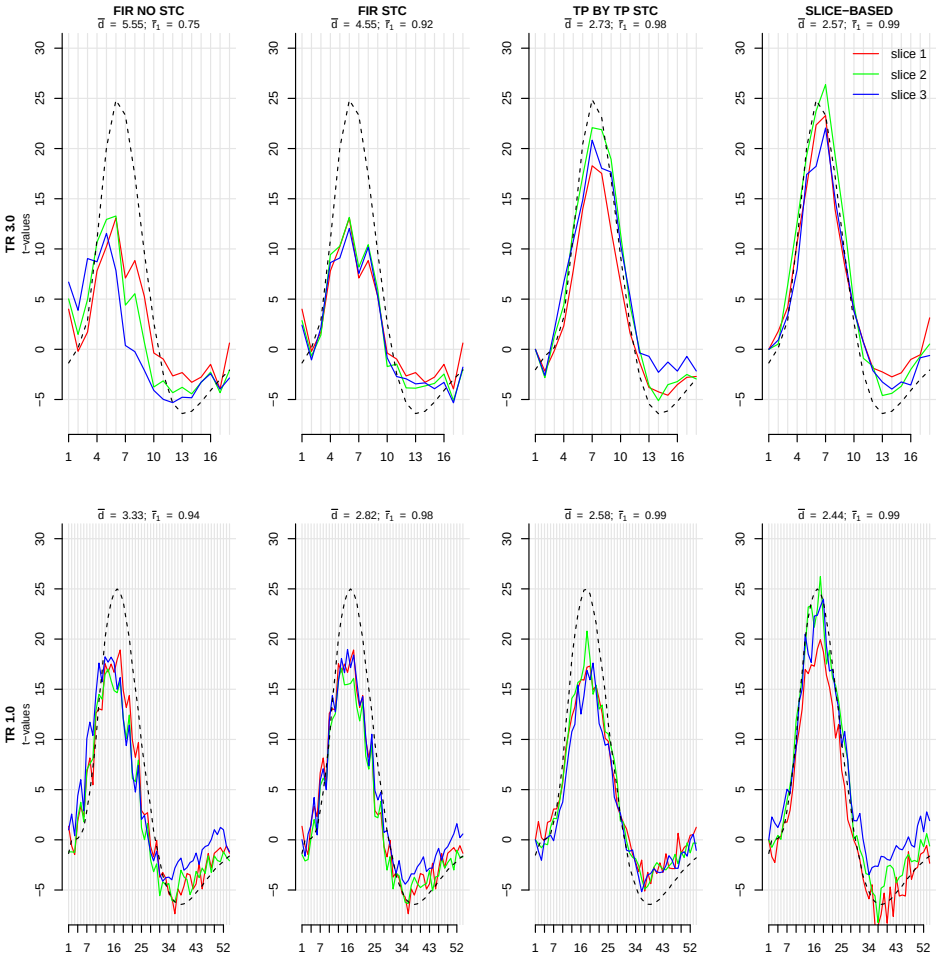

Figure 4: Differences in the detection accuracy of the BOLD signal due to trial-by-trail variability in the onset of the BOLD signal in a simulated fMRI experiment. Each column in the figure represents a different method (first column = FIR without STC; second = FIR with STC; third = Timepoint by Timepoint with STC; fourth = Slice-Based), and each row represents a different TR (TR = 3 top row; TR = 1, bottom row). BOLD response variability was modeled by randomly delaying its onset by 0.5 seconds for half the stimuli. In all simulations, white noise was modeled with $\sigma = 0.15$. Shown are the extracted signals from a single representative simulation. Figure titles list the mean absolute difference between the ground-truth signal and the extracted signals across slices ($\bar{d}$), and the mean correlation between the ground-truth signal and the signal from each slice ($\bar{r}_1$). Note the overall high $\bar{r}_1$ values, and that the lowest $\bar{d}$ values are found in the Slice-Based method.

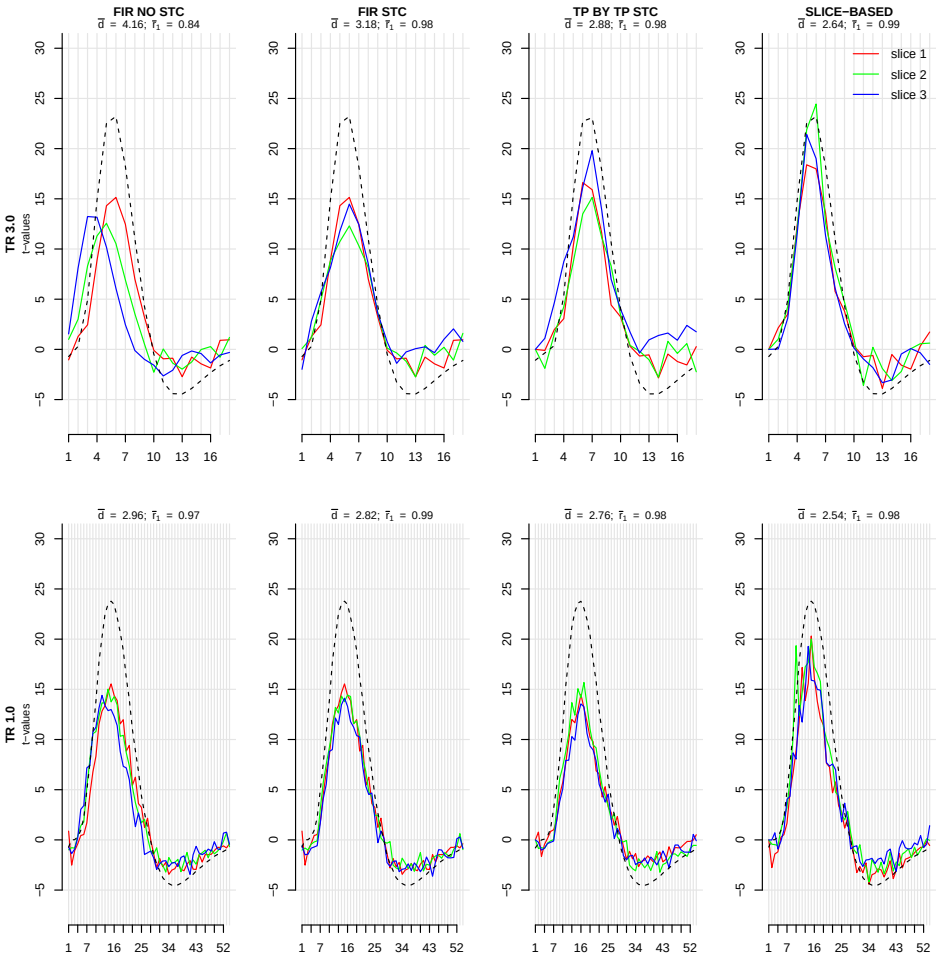

Figure 5: Differences in detection accuracy due to trial-by-trial variability in the shape of the BOLD signal in a simulated fMRI experiment. Half of the stimuli evoked a BOLD response with standard parameters, while the other half yielded a BOLD response with alternative parameters that indicated reduced dispersion of the main peak (see text for details). Note again that the Slice-Based method yielded the lowest $\bar{d}$ values, suggesting this method extracted the most similar ground-truth signal.

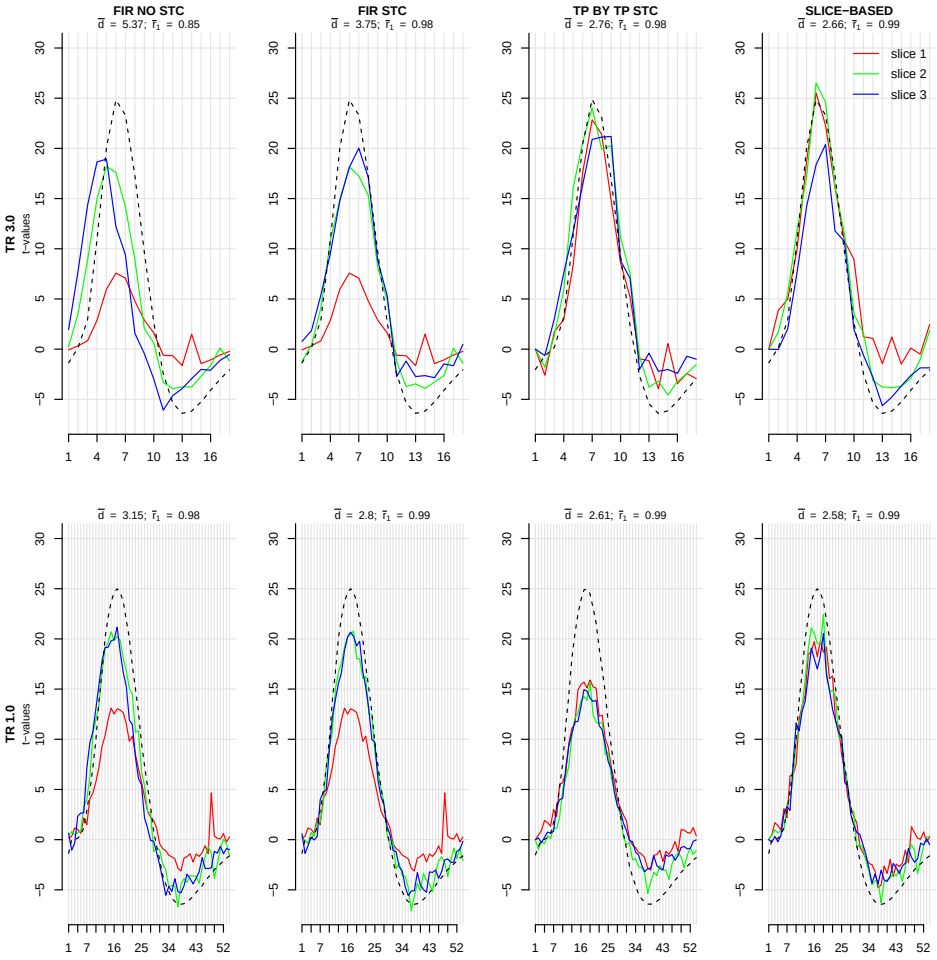

Figure 6: Differences in detection accuracy due to a single signal intensity spike in a simulated fMRI experiment. The single spike was modeled by changing a single intensity value in the time series sampled at the voxel on slice 1 (red line) to 5 times the maximum BOLD signal. The statistical impact of this single spike can be seen in the small peak in the undershoot of the extracted BOLD signal on slice 1. Note that this single spike strongly affected detection accuracy for the FIR based methods at all timepoints (note the overall reduced t-values for the red line), whereas detection accuracy was largely unaffected for the timepoint by timepoint methods.

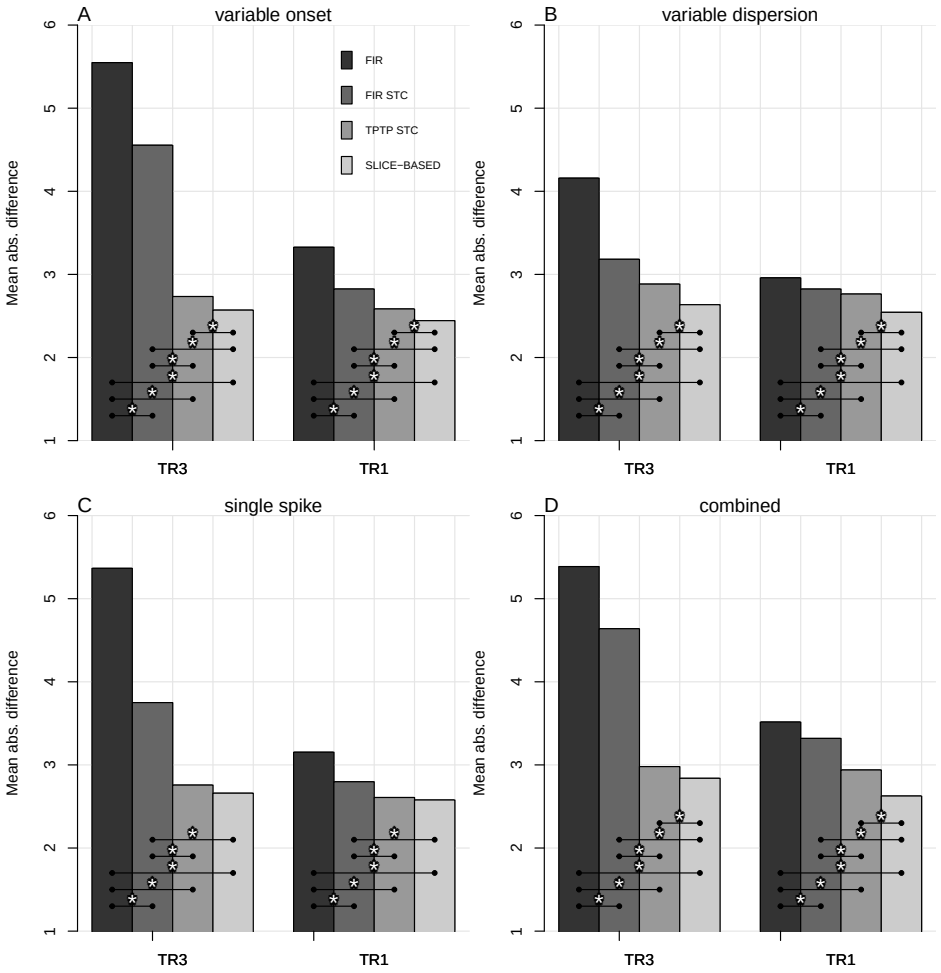

Figure 7: Graphical overview of the means and statistics of the three simulation experiments (panels A-C) and an additional simulation experiment combining all three previous simulations (panel D). Each bar represents the mean absolute difference between the ground-truth signal and the signal at each slice ($\bar{d}$). Note that for the Slice-Based method had the lowest $\bar{d}$ values, suggesting that the extracted signal more closely resembled the ground-truth signal. (*) denotes significant at p ¡ 0.001, see text for details.

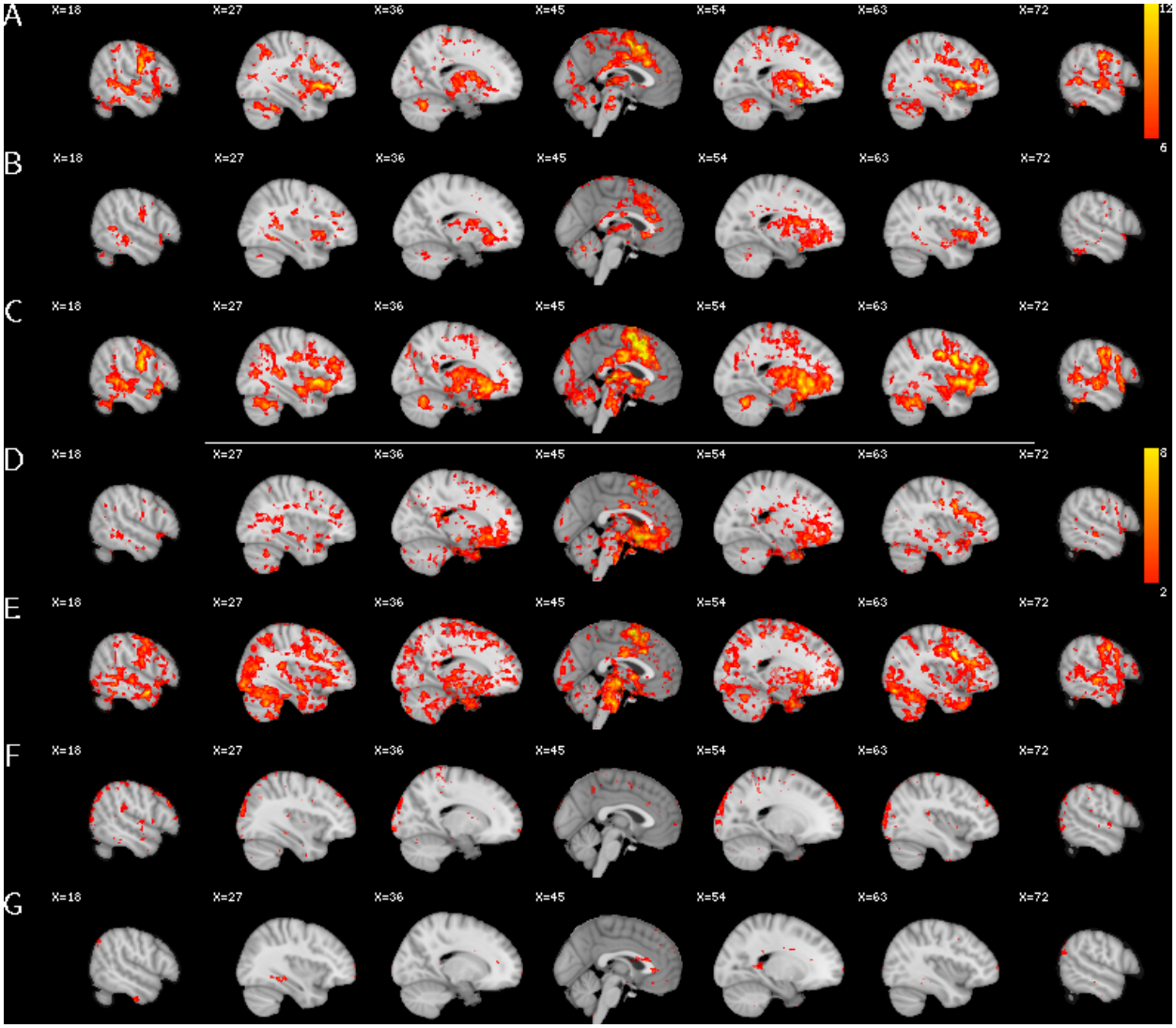

Figure 8: Comparison of standard GLM (panel A), Timepoint by Timepoint with STC (panel B) and Slice-Based (panel C) methods in basic signal detection during picture naming at the group-level with a threshold t ¿ 6.0. The same minimally preprocessed data was used for all three analyses (see text for details). Panels D-G reveal subtractions between unthresholded maps: Slice-Based minus GLM (panel D); Slice-Based minus Timepoint by Timepoint (panel E); GLM minus Slice-Based (panel F); Timepoint by Timepoint minus Slice-Based (panel G). Presented are saggital slices, slice number in upper left corner. Note that although all three methods yielded overall similar pattens of activity, the Slice-Based method has improved signal detection (most notably in medial frontal cortex, panels D and E).

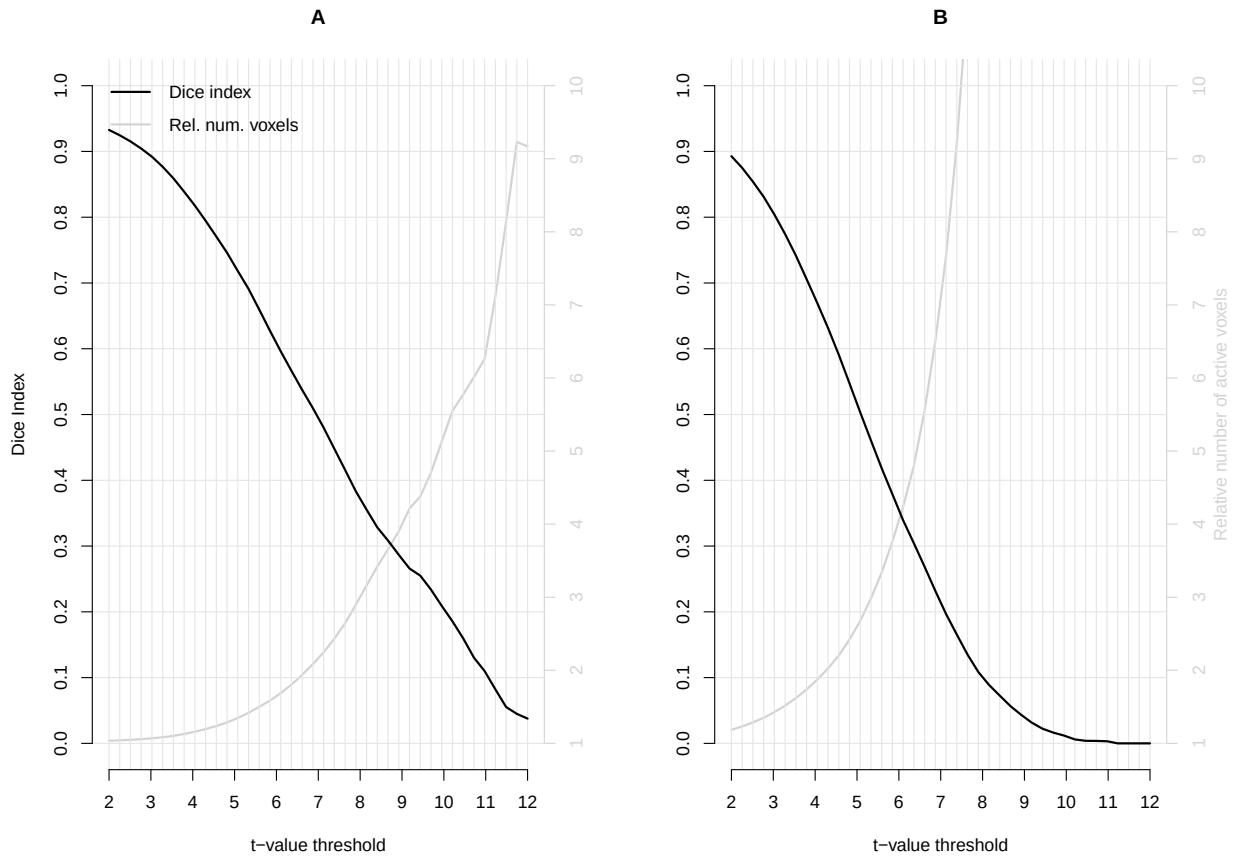

Figure 9: Similarity between the t-value maps of the standard Slice-Based method vs GLM (panel A) and Slice-Based vs Timepoint by Timepoint STC (panel B) at different t-value thresholds (x-axis). The lefthand y-axis shows the Dice index, an index of similarity between two statistical maps. The righthand y-axis shows the relative number of active voxels ($\frac{Slice\ Based}{GLM}$ and $\frac{Slice\ Based}{TPTP}$). Note that the Dice index (black line) revealed decreased similarity between maps at higher thresholds ($t > 4.0$). Furthermore, this decreased similarity at higher thresholds is caused by a dramatic increase in active voxels in the Slice-Based map relative to the GLM and TPTP maps (grey line), suggesting improved signal detection for the Slice-Based method.

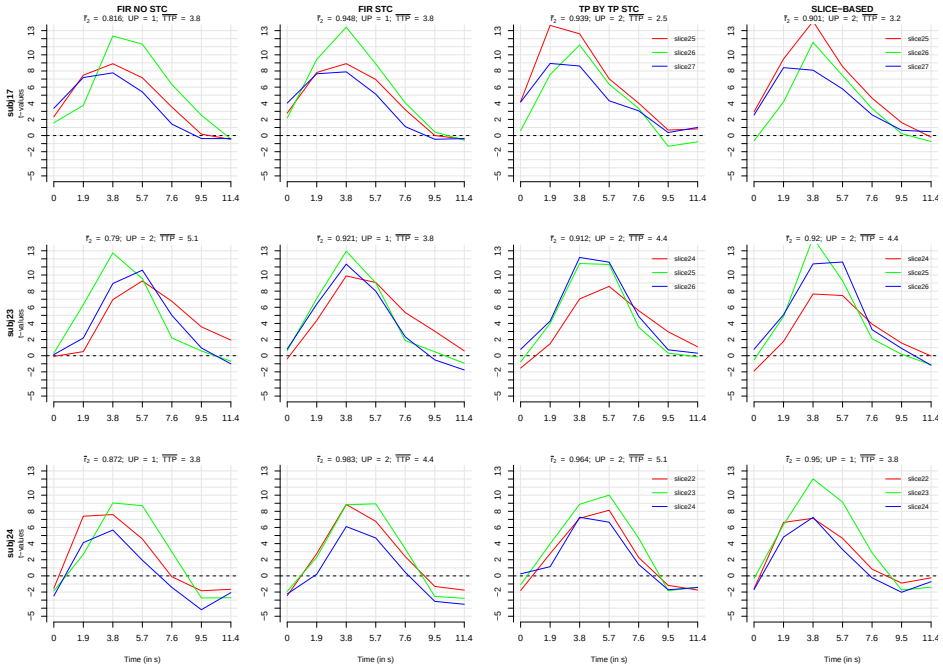

Figure 10: Method comparison using real data from left motor cortex activity obtained using the picture naming task. Each column in the figure represents a different method (first column = FIR without STC; second = FIR with STC; third = Timepoint by Timepoint with STC; fourth = Slice-Based). Signals are extracted from three voxels that appear on adjacent slices (see legend) in the left motor cortex in three representative subjects (top, middle, and bottom row for subjects 17, 23, and 24, respectively). Figure titles list the interslice correlation ($\bar{r}_2$), the mean number of Unique Peaks (UP), and the mean Time To Peak (TTP) for the extracted signals in the graph. Note how the STC methods yielded smoother signals due to signal interpolation but had lower t-values than the Slice-Based method.

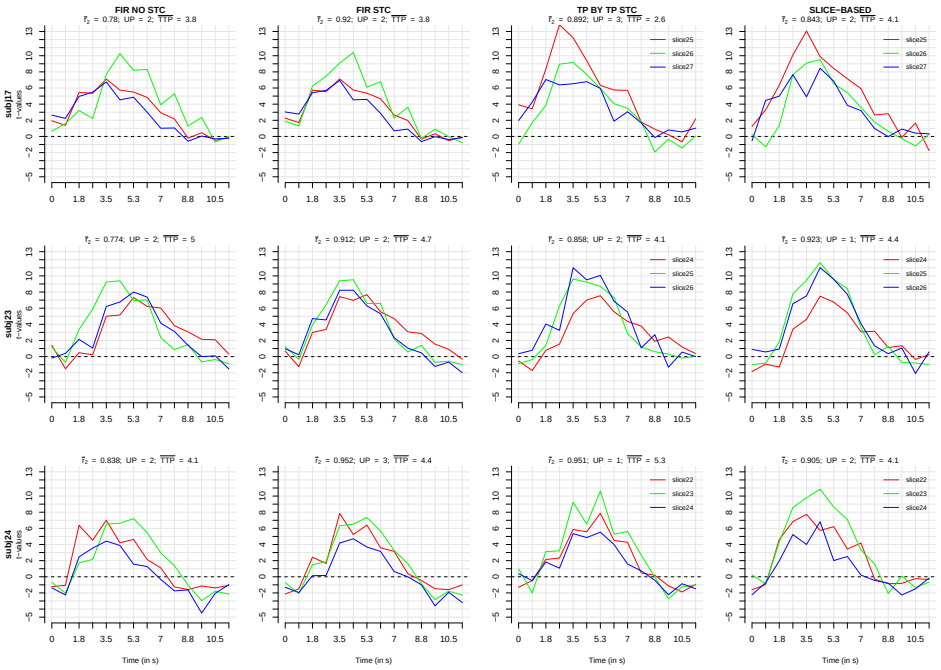

Figure 11: Method comparison using real data from left motor cortex activity obtained using the picture naming task. BOLD signal extracted using the four aforementioned methods at a fixed temporal resolution of 1/2 TR (954 ms). Other aspects identical to those used to obtain Figure 10. Not again how the Slice-Based method detected higher t-value signals despite the increase in temporal resolution.

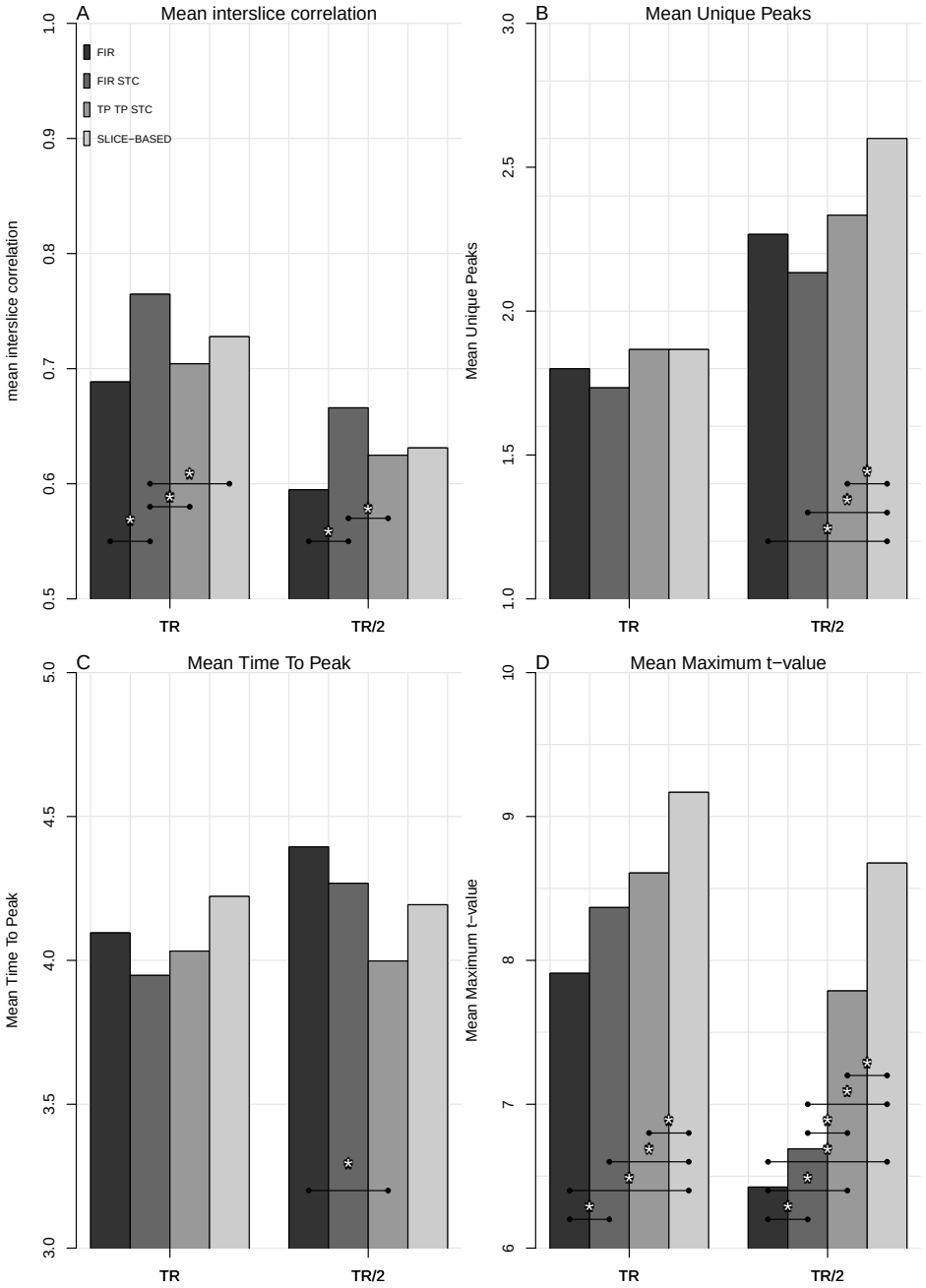

Figure 12: Mean interslice Pearson correlation (A), mean number of Unique Peaks (B), mean Time To Peak (C), and mean Max t-value (D) for the four methods at TR (1908 ms) and TR/2 (954 ms) temporal resolutions. Values obtained from three adjacent slices covering left motor cortex in 30 participants performing the picture naming task. (*) denotes significant at p ¡ 0.05. The slice-based method yielded increased t-values and more stable performance at higher temporal resolution.