fMRI protocol optimization for simultaneously studying small subcortical and cortical areas at 7 T

Miletić, S.; Bazin, P.-L.; Weiskopf, N.; van der Zwaag, W.; Forstmann, B.U.; Trampel, R.

DOI
10.1016/j.neuroimage.2020.116992

Publication date
2020

Document Version
Final published version

Published in
NeuroImage

License
CC BY

Citation for published version (APA):
fMRI protocol optimization for simultaneously studying small subcortical and cortical areas at 7 T

Steven Miletić\textsuperscript{a,}* , Pierre-Louis Bazin\textsuperscript{a,}b , Nikolaus Weiskopf\textsuperscript{b,c} , Wietske van der Zwaag\textsuperscript{d} , Birte U. Forstmann\textsuperscript{a} , Robert Trampel\textsuperscript{b}

\textsuperscript{a} University of Amsterdam, Department of Psychology, Nieuwe Achtergracht 129B, Amsterdam, the Netherlands
\textsuperscript{b} Max Planck Institute for Human Cognitive and Brain Sciences, Department of Neurophysics, Stephansträße 1A, Leipzig, Germany
\textsuperscript{c} Felix Bloch Institute for Solid State Physics, Faculty of Physics and Earth Sciences, Leipzig University, Linnesträße 5, Leipzig, Germany
\textsuperscript{d} Spinoza Center for Neuroimaging, Meibergdreef 75, Amsterdam, the Netherlands

\section*{1. Introduction}

Most fundamental cognitive processes rely on brain networks that include both cortical and subcortical structures. Studying such networks using functional magnetic resonance imaging (fMRI) requires a data acquisition protocol that provides blood-oxygenation-level dependent (BOLD) sensitivity across the entire brain. However, when using standard single echo, echo planar imaging protocols, researchers face a tradeoff between BOLD-sensitivity in cortex and in subcortical areas. Multi echo protocols avoid this tradeoff and can be used to optimize BOLD-sensitivity across the entire brain, at the cost of an increased repetition time. Here, we empirically compare the BOLD-sensitivity of a single echo protocol to a multi echo protocol. Both protocols were designed to meet the specific requirements for studying small, iron rich subcortical structures (including a relatively high spatial resolution and short echo times), while retaining coverage and BOLD-sensitivity in cortical areas. The results indicate that both sequences lead to similar BOLD-sensitivity across the brain at 7 T.
pulse using a single-shot multi echo EPI protocol (Gowland and Bowtell, 2007; Poser et al., 2006; Posse et al., 1999; Speck and Hennig, 1998; Weiskopf et al., 2005). For subsequent statistical analyses, the multiple echoes can then be recombined into a single timeseries. Depending on Weiskopf et al., 2005). For subsequent statistical analyses, the multiple echoes can then be recombined into a single timeseries. Depending on Weiskopf et al., 2005). For subsequent statistical analyses, the multiple echoes can then be recombined into a single timeseries. Depending on Weiskopf et al., 2005). For subsequent statistical analyses, the multiple echoes can then be recombined into a single timeseries. Depending on Weiskopf et al., 2005). For subsequent statistical analyses, the multiple echoes can then be recombined into a single timeseries. Depending on Weiskopf et al., 2005). For subsequent statistical analyses, the multiple echoes can then be recombined into a single timeseries. Depending on Weiskopf et al., 2005). For subsequent statistical analyses, the multiple echoes can then be recombined into a single timeseries. Depending on Weiskopf et al., 2005). For subsequent statistical analyses, the multiple echoes can then be recombined into a single timeseries. Depending on Weiskopf et al., 2005). For subsequent statistical analyses, the multiple echoes can then be recombined into a single timeseries. Depending on Weiskopf et al., 2005). For subsequent statistical analyses, the multiple echoes can then be recombined into a single timeseries. Depending on Weiskopf et al., 2005). For subsequent statistical analyses, the multiple echoes can then be recombined into a single timeseries. Depending on Weiskopf et al., 2005). For subsequent statistical analyses, the multiple echoes can then be recombined into a single timeseries. Depending on Weiskopf et al., 2005). For subsequent statistical analyses, the multiple echoes can then be recombined into a single timeserie...
\[ CNR(TE) = \frac{S_0}{\sigma} T_{\text{Eexp}} \left( -\frac{TE}{T_2^*} \right) \]  

(1)

where \( S_0 \) is the signal at \( TE = 0 \) ms. The expected CNR of the multi echo protocol depends on the method for combining the echoes (Gowland and Bowtell, 2007; Kettinger et al., 2016; Posser et al., 2006; Posser and Norris, 2009; Posse et al., 1999; Weiskopf et al., 2005). When using the optimal combination (OC) method (Posse et al., 1999), the weighting factor \( w \) of echo \( n \) is determined by the echo time \( TE \) and the estimated \( T_2^* \) of the voxel, according to:

\[ w_n = \frac{T_{E\text{exp}} \left( -\frac{T_n}{T_2^*} \right)}{\sum_n T_{E\text{exp}} \left( -\frac{T_n}{T_2^*} \right)} \]  

(2)

Assuming that noise is uncorrelated across echoes, and TE-independent, the corresponding CNR is then given by Posser et al. (2006):

\[ \text{CNR}_{\text{oc}} = \frac{\sum Nw_n^* T_{E\text{exp}}^* S_n}{\sqrt{\sum Nw_n^2}} \]  

(3)

When calculating the expected CNR of both protocols for an iron-rich area with a \( T_2^* \) value of 15 ms, and assuming that the SNR of the multi echo protocol is 0.57 times the SNR of the single echo protocol, we find that \( \text{CNR}_{\text{oc}} \approx 0.78 \). For cortical areas with \( T_2^* \) values of 33 ms (as is typical for cortical grey matter at 7 T, Van der Zwaag et al., 2016), and assuming that the SNR of the single echo protocol is 0.62 times the SNR of the single echo protocol, \( \text{CNR}_{\text{oc}} \approx 1.19 \). Thus, these calculations suggest that the single echo protocol should have higher CNR in iron-rich subcortical areas, whereas the multi echo protocol should have higher CNR in cortical areas.

Several cautionary notes apply to these calculations, which could cause discrepancies between the theoretical estimations above and empirically observed data. Firstly, the calculations depend on a range of assumptions. The CNR estimations assume mono-exponential decay of transverse magnetization, and in the case of the multi echo protocol, uncorrelated noise across echoes and TE-independence of noise. Furthermore, it does not take into account susceptibility-gradient echo time shifts (see Volz et al., 2019, for an overview of these effects), which may further impact CNR.

Secondly, the calculations ignore the potential additional effects of multiband acceleration on SNR for the multi echo protocol (more on this in the Discussion), and the effects of ramp sampling. The latter may cause two effects which are not necessary or practically impossible, respectively, to take into account in these calculations. For one, a certain part of the outer \( k \)-space is acquired with a varying effective bandwidth due to the sampling with varying readout gradient amplitudes modifying the effective SNR and noise correlations in this area of \( k \)-space. However, since only the outer \( k \)-space is affected, which typically contributes little to the SNR compared to the \( k \)-space center, this would only have a small effect on our SNR and CNR metrics. Furthermore, ramp sampling can cause increased Nyquist (N/2) ghosting, since readout polarity dependent gradient imperfections are pronounced at points of dynamic gradient changes (e.g., at the transition between the ramp and flat top of the readout gradient). The ghosting can significantly degrade image quality. However, the dependence of the ghosting artifact on the extent of ramp sampling is non-trivial and cannot be predicted from the ramp sampling percentage based on established models, but depends on the details of the EPI implementation and scanner hardware. Therefore, we performed a visual inspection of the EPI time series in order to avoid severe Nyquist ghosting in order to minimize the second effect of ramp sampling.

Thirdly, the estimation cannot take parameters into account which depend on the specific subject. Certain scanning techniques, however, for instance parallel imaging or ramp sampling, may especially suffer from subject motion and/or breathing. Therefore, the multi echo protocol may depend more on subject behavior than the single-echo protocol. This results in a small effect for highly compliant volunteers but could be a significant cause for image degradation when scanning MR-naive subjects or patient groups.

2.4. Behavioral paradigm

The task used was an auditory stop-signal response task (Logan et al., 1984; Verbruggen et al., 2019). This task was chosen because it is known to elicit BOLD-responses both in cortical areas and in small, subcortical nuclei such as the subthalamic nucleus (e.g., Aron and Poldrack, 2006; De Hollander et al., 2017). In this task, the participant is presented with an arrow pointing left or right, and is instructed to respond as fast as possible without making a mistake, by pressing a button with the corresponding hand. On a subset of the trials (25%), an auditory tone is presented shortly after the arrow. On these trials, the participant is instructed to withhold responding. The time between the presentation of the arrow and the auditory stop signal (the stop-signal delay; SSD) is adjusted to the participant’s performance on the task using a staircase procedure to obtain a successful inhibition rate of 50% (Verbruggen et al., 2019). Response times and response directions are recorded on each trial.

Trials are categorized as go trials (no stop signal presented), failed stop (a stop signal was presented, but the participant was not able to inhibit their response), or successful stop (a stop signal was presented, and the participant successfully inhibited their response). Due to the simplicity of the go trials, incorrect go responses (e.g., a left button press in response to an arrow pointing rightwards) are infrequently observed and are typically excluded from statistical analyses. In fMRI research using the stop-signal response task (e.g., Aron and Poldrack, 2006; De Hollander et al., 2017; Li et al., 2008), three contrasts are of main interest to study which brain areas are involved in stopping an initiated response: failed stop > go, successful stop > go, and failed stop > successful stop.

2.5. Procedure and exclusions

Before each session, the participant received instructions and practiced the task for approximately 10 min. Each session started with a short localizer scan, followed by two functional runs, a field map, the third functional run, and the anatomical scan. Each functional run took 17:09 min (343 vol). Each trial took 9 s (3 vol), and consisted of a fixation cross (with a jittered duration of 750 ms, 1500 ms, 2250 ms, or 3000 ms to decorrelate trial onsets), the stimulus (1 s), followed by a blank screen for the remainder of the trial duration. The total duration of a full session was approximately 1 h.

After the second run in the multi echo session, one participant reported not being able to hear the auditory stop signal. This participant was excluded from all further analyses. Further, one other participant did not perform the task significantly above chance level (50% correct answers in go trials) in the third run during both sessions due to fatigue. These runs were excluded from analyses. For one other participant, the third run was aborted after 12 min due to technical issues during the multi echo session. Of this participant, the data from the third runs of both sessions were excluded from analyses.

For five participants, there was no time left at the end of one of the sessions to finish the MP2RAGE scan. It was ensured that an MP2RAGE scan was always acquired in at least one of the two sessions. Finally, for one participant, the flip angle in the multi echo sequence had to be lowered to 60° to remain below the specific absorption rate (SAR) limit.

2.6. MRI data preprocessing

All functional data were preprocessed using fMRIPrep (version 1.2.6; Esteban et al., 2019). The pipeline included slice timing correction using 3dTshift from AFNI 20160207 (Cox and Hyde, 1997), head motion
correction using mcfilt (FSL 5.0.9, Jenkinson et al., 2002), susceptibility distortion correction using a custom workflow, realignment using bbregister (FreeSurfer; Greve and Fischl, 2009), and registration to the ICBM 152 Nonlinear Asymmetrical template version 2009c (henceforth “MN12009c”, Fonov et al., 2009) using antsRegistration (ANTS 2.2.0, Avants et al., 2008). Full pipeline details are included in Appendix A.

Prior to running fMRIPrep, background noise in the MP2RAGE T1-weighted images was reduced using the robust MP2RAGE combination method (O’Brien et al., 2014). This was required to ensure that the skull-stripping routine included in fMRIPrep can process the data. For the multi echo data, fMRIPrep was modified to process each echo separately. For the second and third echoes, however, head motion correction and realignment were based on the parameters obtained from processing the first echo. This was done because the first echo has highest signal intensity and should lead to best estimation of head motion correction and realignment parameters (Speck and Henning, 2001).

After running fMRIPrep, all data were temporally filtered using a highpass filter (cut-off 1/128 Hz) to remove slow drifts. The multi echo data were then recombined into a single timeseries using a weighted sum, using the optimal combination (OC) method (Posse et al., 1999). In this method, the weighting factor \( w \) of echo \( n \) is determined by the echo time \( TE \) and the estimated \( T_2^* \) of the voxel, according to Equation (1). Voxelwise \( T_2^* \)-values were estimated by fitting a mono-exponential decay curve through the mean signal (across time) of the three echoes. Initial explorations showed that the results and conclusions would be highly similar using the parallel acquired inhomogeneity desensitized (PAID; Poser et al., 2006) method of combination (see also Kettinger et al., 2016).

### 2.7. Behavioral analyses

Trials in which the participants responded faster than 150 ms or slower than the stimulus duration of 1 s were excluded from analyses (0.20% of all trials). Then, for both sessions, we calculated the median RT for each trial on go trials and on failed stop trials, response accuracy, mean stop-signal delay (SSD), and proportion of successful stops. Further, for each participant, we estimated the amount of time it took to stop an initiated response: The stop-signal response time (SSRT). SSRTs were estimated using the mean method, which assumes that go and stop processes race independently from each other to a common threshold in a so-called horse race (Aron and Poldrack, 2006; Logan et al., 1984). Under this assumption, the SSRT can be calculated by finding the percentile of go RTs that corresponds to the proportion of successful stops (approximately 50%), and subtracting the mean SSD from this value.

### 2.8. Regions of interest (ROIs) analyses

Alongside the whole-brain analyses reported below, we are interested in several ROIs of the response inhibition network. We specifically study IFG, M1, preSMA, STN, GPe, and GPI. For all ROIs except preSMA, masks from probabilistic atlases were used. The masks of STN, GPe, GPI, and STR were obtained from the ATAG atlas (Young Adults; Keuken et al., 2014). The mask of primary motor cortex was obtained from the Brainnetome atlas (“upper limb area”, Fan et al., 2016). The mask of IFG was extracted from the Harvard-Oxford cortical atlas (Desikan et al., 2006), identified as the union of the pars opercularis and pars triangularis. Finally, a discrete preSMA mask was based on coordinates provided by Johansen-Berg et al. (2004).

All masks were registered to MN12009c space (and down sampled to 1.6 mm isotropic) using tools embedded in nighres (Huntenburg et al., 2018). For brevity, ROI analyses reported below are based on the right hemisphere only, as it has been suggested that the stopping network is right lateralized (Aron and Poldrack, 2006). Explorations of the left hemisphere ROIs suggest, however, similar patterns of BOLD-activity as their contralateral counterparts (see also Fig. 5 below).

### 2.9. Temporal signal to noise ratios (tSNR) and contrast to noise ratios (CNR)

For both protocols, as well as for the individual echoes of the multi echo protocol, we calculated the tSNR as the voxelwise ratio of the mean and the standard deviation of the signal across the timeseries. tSNR values for each ROI are a weighted mean of each voxels’ tSNR, with the weighting factor equal to the voxels’ probability of belonging to the ROI. For each subject and session, tSNR-values were first calculated per run, after which the tSNR-values were averaged across runs. When using the standard deviation of the residuals of the GLM analyses reported below as a noise measure (e.g., Weiskopf et al., 2005), the results are qualitatively similar and lead to the same conclusions.

Secondly, we estimated the expected relative BOLD contrast to noise ratio (CNR), where BOLD contrast is defined as the amount of signal change per unit change in \( T_2^* \) due to fluctuations in tissue oxygenation (Poser et al., 2006). For a single echo sequence, the CNR per unit change in \( T_2^* \) as a function of TE is given by Equation (1). Assuming mono-exponential decay, the measured signal decays with:

\[
S(TE) = S_0 \exp \left( -\frac{TE}{T_2^*} \right)
\]  

From (1) and (4), and under the assumption that noise is independent from TE, it follows that CNR can be rewritten as \( CNR = \frac{S_0}{T_2^*} \), which implies that CNR can be calculated as the voxelwise product of the tSNR and TE of the sequence (Poser et al., 2006). We note that we did not adjust for susceptibility-gradient related echo time shifts (Volz et al., 2019), which may further impact the BOLD sensitivity.

For a combined multi echo image, the CNR depends on the echo combination method (Gowland and Bowtell, 2007; Kettinger et al., 2016; Poser et al., 2006; Poser and Norris, 2009; Posse et al., 1999; Weiskopf et al., 2005). When using a weighted sum, the contrast can be calculated by taking the weighted sum of the contrast in all echoes (estimated in the same way as the contrast in a single echo image). The noise term can be estimated by taking the standard deviation of time series of the combined signal (Poser et al., 2006).

Note that these analyses depend on a range of simplifying assumptions (e.g., signal decay is mono-exponential, no susceptibility-gradient related echo time shifts, noise is Gaussian), which likely do not all hold. As a consequence, the BOLD-sensitivity estimates from the GLM analyses reported below can deviate from these analyses.

### 2.10. Signal deconvolution

To estimate the shape and size of the signal responses to each trial type in the experimental paradigm, we performed signal deconvolution analyses (Gitelman et al., 2003; Glover, 1999; Poldrack et al., 2011). The signals used were weighted mean timeseries within ROIs (without spatially smoothing the data), where voxel contributions to timeseries were weighted by their probability of belonging to the region according to the probabilistic masks. ROI time series were rescaled to percentage signal change by dividing each timepoint’s signal value by the mean signal value (as baseline), multiplying by 100, and subtracting 100. Data from the three runs (per session per participant) were concatenated. Identical analyses were performed for the single echo data and the optimally combined multi echo data, as well as each of the individual echoes of the multi echo protocol.

For deconvolution, Fourier basis sets were used (Bullmore et al., 1996; Gitelman et al., 2003). Fourier basis sets provide a smoothness constraint, which decreases variance in the parameter estimates compared to finite impulse response (FIR) basis sets, at the cost of minimal parameter estimation bias (Liu et al., 2017; Poldrack et al., 2011). Per trial type, five basis sets were used (an intercept, two sines, and two
cosines), totaling 16 regressors (three trial types plus an overall intercept) in the design matrix. The model was fit to the data using ordinary least squares. We also fit a second deconvolution model with seven additional motion regressors (rotation and translation in three dimensions, and framewise displacement), but as the results were highly similar, we report only the results without these additional motion regressors. Deconvolution analyses were performed using nideconv (https://github.com/VU-Cog-Sci/nideconv).

We visualize signal time courses (in seconds after trial onset) expressed as percentage signal change. Due to the TE-dependence of percentage signal change (Kundu et al., 2012; Menon et al., 1993), it is expected that percentage signal change is higher in the optimally combined multi echo data (which has, on average, higher TEs) than in the single echo data. However, since the amount of noise in the data is also expected to increase with TE, the percentage signal change itself is not a pure estimate of statistical detection power. To quantify the noise component, we calculated $\sigma^2$ (the sum of squared errors divided by the degrees of freedom) of the models. Formal statistical comparisons between the two protocols are performed using GLMs, which we turn to in the next section.

### 2.11. General linear models (GLMs)

To statistically assess differences between protocols, we first fit a set of GLMs using the canonical double-gamma hemodynamic response function (HRF) and temporal derivatives (Friston et al., 1998, 1995; 1994; Glover, 1999). In the design matrix, the three trial types of the experimental paradigm were included as regressors of interest. Further, six motion parameters (translation and rotation in three dimensions) obtained from head motion correction and the framewise displacement were included in the design matrix as confounding variables. Including the intercept, this leads to a total number of 13 regressors. The main contrasts of interest were failed stop > go, successful stop > go, and failed stop > successful stop.

GLM analyses were done both whole-brain and ROI-based (using the weighted mean of the signal timeseries within each ROI). Whole-brain GLMs were performed using FSL FEAT (Jenkinson et al., 2012; Woolrich et al., 2001) using the FILM method to account for autocorrelated residuals. Prior to fitting the GLMs (for the whole-brain analyses only) the data were spatially smoothing using a Gaussian smoothing kernel (FWHM = 5 mm). This spatial smoothing decreases the effective resolution considerably, but it allows for comparison of the overall whole-brain GLM with previous studies (notably, De Hollander et al., 2017). Fixed-effects GLMs were used to combine data from different runs per subject and session. FLAME1 (Woolrich et al., 2004) was subsequently used to analyze the group-level results per session separately. Statistical parametric maps (SPMs) were subsequently created to visualize the whole-brain results of the group-level models. All $p$-values were
corrected for the false discovery rate (FDR; Yekutieli and Benjamini, 1999), and we used a critical value of \( q < 0.05 \) to determine significance.

ROI-based GLMs were done by first concatenating all runs within a session per participant (after rescaling each run’s timeseries to percent signal change). Unlike the whole-brain GLMs, these data were not spatially smoothed to ensure that signal within each ROI is not contaminated with signal from outside the ROI. Then, we fit the GLM using an AR(1)-model as implemented in Python package statsmodels (Seabold and Perktold, 2010) to account for temporal autocorrelation in the timeseries. The whitened residuals were inspected to ensure there were no between-protocol differences in the remaining autocorrelation, which could bias \( t \)-values. GLMs were fit to data from both sessions separately.

Statistical analyses comparing the two protocols were based on the analyses presented in Puckett et al. (2018). From the fitted GLMs, we extracted \( t \)-statistics per subject and session (from the fixed effects analysis) as a measure of BOLD-sensitivity for each trial type (cf. Puckett et al., 2018). To statistically compare protocols, we used a linear mixed effects model (Gelman and Hill, 2007) with these \( t \)-values dependent variables, and ROI, trial type, and protocol as independent variables including all possible interaction terms. ROI was included as an independent variable to be able to study whether the protocols differ in their BOLD-sensitivity across ROIs. Trial type was included in as an independent variable because it is unlikely that all ROIs show the same BOLD-response in all trial types. Alternatively, it would have been possible to analyze only one of the three trial types, but this would require an arbitrary choice as to which trial type is analyzed. A second possible approach was to fit a single GLM to data from both sessions (after rescaling each run’s timeseries to percent signal change). Unlike the whole-brain GLMs, these data were not spatially smoothed to ensure that signal within each ROI is not contaminated with signal from outside the ROI. Then, we fit the GLM using an AR(1)-model as implemented in Python package statsmodels (Seabold and Perktold, 2010) to account for temporal autocorrelation in the timesseries. The whitened residuals were inspected to ensure there were no between-protocol differences in the remaining autocorrelation, which could bias \( t \)-values. GLMs were fit to data from both sessions separately.

Random intercepts were included for subjects. To ensure larger \( t \)-values always indicate a larger effect, the sign of \( t \)-values of terms for which the group effect was negative was flipped. Homogeneity of the residuals as well as the normality of residuals were assessed. Follow-up paired \( t \)-tests (one per combination of ROI, trial type and data type) were used to provide detail on the possible interaction terms. The reported \( q \)-values are \( p \)-values adjusted for the false discovery rate (critical value \( q < 0.05 \)).

In frequentist tests, the lack of a significant finding cannot be interpreted as evidence against the presence an effect (e.g., Wetzel and Wagenmakers, 2012). To allow for quantifying evidence against effects, we repeated the statistical analyses using Bayesian linear models (Liang et al., 2008; Rouder and Morey, 2012) as implemented in the R package BayesFactor (version 0.9.12–4.2; Morey et al., 2018).

Exploratory analyses were included to assess how BOLD-sensitivity changes with TE in the various ROIs. To do so, we first fit, for every participant and ROI, a linear regression model predicting the \( t \)-value using the echo time as predictor. The slope of the resulting model, \( \Delta t/TE \), indicates the amount of change in \( t \)-values per millisecond increase in TE. Values of \( \Delta t/TE \) were calculated for each trial type (vs baseline), and then averaged across trial types per participant and ROI. Then, for every participant, we used the \( T_2^* \) per ROI (obtained by fitting monoexponential decay functions through ROI timeseries per echo) to predict \( \Delta t/TE \). A one-sample \( t \)-test was used to test whether the slope of this model is different from 0 on a group level.

3. Results

3.1. Behavioral analyses

Descriptive statistics summarizing the behavioral data are presented in Table 1. In both sessions, the data are typical for the stop-signal response task (e.g., De Hollander et al., 2017; White et al., 2014). The percentages of successful inhibition are close to 50%, indicating that the staircasing procedure successfully adjusted SSDs to participants’ SSRTs. Between sessions, no significant differences were observed in any of the dependent variables after multiple comparison correction (using the FDR with critical value \( q < 0.05 \)). Further, median go RTs were not correlated with SSRTs, in line with the assumption of independence between the go and stop processes in the horse-race model (Logan et al., 1984).

3.2. tSNR and CNR

As a first assessment of functional data quality, we analyzed the temporal signal-to-noise ratios. Fig. 1 illustrates the whole-brain tSNR for each of the protocols. tSNR values within each ROI are also summarized in Table 2. The single echo tSNRs are in line with earlier findings using a...
similar protocol (De Hollander et al., 2017). Across the all the ROIs, the tSNR of the single echo protocol is higher than the tSNR of the optimally combined multi echo protocol (smallest difference observed in rIFG, \(t(16) = 11.08, p < 0.001\)). In the multi echo protocol, the decrease in tSNR across the three echoes deviated only marginally from the decrease in tSNR that is expected based on \(T^*\) decay (results not shown).

Contrast-to-noise depends on both the tSNR and the ratio between the TE and the \(T^*\) of the tissue (Eq. (1)). The short TE in the single echo sequence may have resulted in a high tSNR, but the TE is much shorter than the \(T^*\) in many brain areas (Fig. 2 provides a \(T^*\) map to illustrate the between-region variability in \(T^*\)). For such brain areas, the single echo protocol samples from the \(T^*\)-decay curve too early, which can lead to decreased BOLD contrast-to-noise. The BOLD CNR of the optimally combined multi echo data is expected to profit from the additional samples from the \(T^*\)-decay curve.

To provide a comparison between the protocols that considers both tSNR and TEs, we estimated the theoretically expected relative CNR-values. Fig. 3 shows the voxel-wise CNR-values across the entire brain. Table 3 summarizes these CNR-values for the ROIs. The results are shown in Table 3 and Fig. 3. The expected CNR in rPreSMA and rIFG is higher in the optimally combined multi echo data (\(t(16) = 6.54, p < 0.001\) and \(t(16) = 4.76, p < 0.001\), respectively), but the single echo data has a higher expected CNR in subcortical areas (smallest difference for rSTR, \(t(16) = 8.15, p < 0.001\)).
echoes) are likely largely noise and not BOLD-signal. Since there appears to be limited
TE-dependency in GPe and GPi, the data observed (at least in the later
stimulus onset, followed by a post-stimulus undershoot until approximately
15 s after stimulus onset. The deconvolved signal of the individual
responses to the three trial types. Deconvolved responses are shown in
Fig. 4. In every ROI, the deconvolved responses, while not identical, look
qualitatively similar between protocols. With the exception of GPe and
GPi, the amount of signal increases until approximately 7
s after stimulus onset, followed by a post-stimulus undershoot until approximately
15 s after stimulus onset. The deconvolved signal of the individual
responses indicates that the signal changes are due to changes in T$_2^*$, and thus likely BOLD-responses (Kundu et al.,
2012; Menon et al., 1993). Since there appears to be limited
TE-dependency in GPe and GPi, the data observed (at least in the later
echoes) are likely largely noise and not BOLD-signal.

### 3.3. Signal deconvolution

Deconvolution analyses were performed to estimate the signal
responses to the three trial types. Deconvolved responses are shown in
Fig. 4. In every ROI, the deconvolved responses, while not identical, look
qualitatively similar between protocols. With the exception of GPe and
GPi, the amount of signal increases until approximately 7–10 s after stimulus onset, followed by a post-stimulus undershoot until approximately
15 s after stimulus onset. The deconvolved signal of the individual
echoes show clear TE-dependence (again with the exception of GPe and
GPi) with higher percentages signal change observed at later TEs. The TE-
dependence of the observed responses indicates that the signal changes are due to changes in $T_2^*$, and thus likely BOLD-responses (Kundu et al.,
2012; Menon et al., 1993). Since there appears to be limited
TE-dependency in GPe and GPi, the data observed (at least in the later
echoes) are likely largely noise and not BOLD-signal.

### Table 2

<table>
<thead>
<tr>
<th>Data/ROI</th>
<th>Cortex</th>
<th>Subcortex</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rM1</td>
<td>rPreSMA</td>
</tr>
<tr>
<td>Single echo</td>
<td>91.06 (16.25)</td>
<td>86.74 (13.48)</td>
</tr>
<tr>
<td>Multi echo (OC)</td>
<td>60.49 (8.63)</td>
<td>60.43 (7.51)</td>
</tr>
<tr>
<td>Echo 1</td>
<td>72.66 (8.99)</td>
<td>67.08 (6.58)</td>
</tr>
<tr>
<td>Echo 2</td>
<td>40.61 (4.29)</td>
<td>42.18 (4.43)</td>
</tr>
<tr>
<td>Echo 3</td>
<td>24.95 (2.86)</td>
<td>27.60 (3.58)</td>
</tr>
<tr>
<td></td>
<td>rSTR</td>
<td>rGPe</td>
</tr>
<tr>
<td>Single echo</td>
<td>72.13 (10.86)</td>
<td>55.38 (9.76)</td>
</tr>
<tr>
<td>Multi echo (OC)</td>
<td>43.52 (4.15)</td>
<td>31.54 (3.66)</td>
</tr>
<tr>
<td>Echo 1</td>
<td>44.82 (3.88)</td>
<td>22.03 (3.82)</td>
</tr>
<tr>
<td>Echo 2</td>
<td>23.11 (2.72)</td>
<td>10.89 (1.55)</td>
</tr>
<tr>
<td>Echo 3</td>
<td>12.28 (1.81)</td>
<td>5.55 (0.60)</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Data/ROI</th>
<th>Cortex</th>
<th>Subcortex</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rM1</td>
<td>rPreSMA</td>
</tr>
<tr>
<td>Single echo</td>
<td>91.06 (16.25)</td>
<td>86.74 (13.48)</td>
</tr>
<tr>
<td>Multi echo (OC)</td>
<td>60.49 (8.63)</td>
<td>60.43 (7.51)</td>
</tr>
<tr>
<td>Echo 1</td>
<td>72.66 (8.99)</td>
<td>67.08 (6.58)</td>
</tr>
<tr>
<td>Echo 2</td>
<td>40.61 (4.29)</td>
<td>42.18 (4.43)</td>
</tr>
<tr>
<td>Echo 3</td>
<td>24.95 (2.86)</td>
<td>27.60 (3.58)</td>
</tr>
<tr>
<td></td>
<td>rSTR</td>
<td>rGPe</td>
</tr>
<tr>
<td>Single echo</td>
<td>72.13 (10.86)</td>
<td>55.38 (9.76)</td>
</tr>
<tr>
<td>Multi echo (OC)</td>
<td>43.52 (4.15)</td>
<td>31.54 (3.66)</td>
</tr>
<tr>
<td>Echo 1</td>
<td>44.82 (3.88)</td>
<td>22.03 (3.82)</td>
</tr>
<tr>
<td>Echo 2</td>
<td>23.11 (2.72)</td>
<td>10.89 (1.55)</td>
</tr>
<tr>
<td>Echo 3</td>
<td>12.28 (1.81)</td>
<td>5.55 (0.60)</td>
</tr>
</tbody>
</table>

### Table 4

Descriptives of the dependent variables of interest for the behavioral analyses.

<table>
<thead>
<tr>
<th>Median go RT (SD), ms</th>
<th>Median failed stop RT (SD), ms</th>
<th>% errors (SD)</th>
<th>SSRT (SD), ms</th>
<th>SSD (SD), ms</th>
<th>% successful inhibition (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single echo</td>
<td>438.82 (63.54)</td>
<td>424.70 (56.64)</td>
<td>0.64</td>
<td>233.50 (83.54)</td>
<td>0.64</td>
</tr>
<tr>
<td>Multi echo</td>
<td>463.42 (78.49)</td>
<td>446.30 (65.32)</td>
<td>0.46</td>
<td>229.56 (94.49)</td>
<td>0.46</td>
</tr>
</tbody>
</table>

### Table 5

<table>
<thead>
<tr>
<th>Go RT (SD), ms</th>
<th>Stop RT (SD), ms</th>
<th>Failed trials (SD), ms</th>
<th>Median RT (SD), ms</th>
<th>rM1 (SD)</th>
<th>rPreSMA (SD)</th>
<th>rFG (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single echo</td>
<td>438.82 (63.54)</td>
<td>424.70 (56.64)</td>
<td>0.64</td>
<td>233.50 (83.54)</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>Multi echo</td>
<td>463.42 (78.49)</td>
<td>446.30 (65.32)</td>
<td>0.46</td>
<td>229.56 (94.49)</td>
<td>0.46</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5. Group-level SPMs of the three main contrasts of interest for both the single echo and optimally combined multi echo data. Colors indicate z-values. Thresholds
were set using the FDR method ($q < 0.05$). Since this method determines the critical z-value based on the distribution of observed z-values, the critical values obtained
for each contrast differs between protocols. To prevent within-contrast varying thresholds between protocols, the highest of the two critical z-values (per contrast) was
chosen as the threshold. Contours indicate regions of interest, thresholded at a 30% probability level (rIFG in white, striatum in blue, GPe and GPi in dark and light
green, respectively, and STN in light blue). Slice locations are in MNI2009c space (1 mm).

Interestingly, the data suggest a difference in the post-stimulus undershoot amplitude between the cortical and subcortical responses. Whereas the amplitude of the undershoot in cortical areas is approximately 1/5th of the peak amplitude (in line with the canonical HRF; Glover, 1999), the undershoot amplitude in subcortical areas (rSTR, rSTN) is roughly equal to the peak. This pattern is present in the single echo data, in the optimally combined multi echo data, and in each of the individual echoes where it scales with TE. Further, we also observed it in the left hemisphere ROIs (results not shown), and it appears present in earlier findings as well (De Hollander et al., 2017). We briefly return to this observation in the Discussion.

A second notable aspect of the responses is that there appears to be a
delay in the hemodynamic response in rM1 during Successful Stop trials (onset roughly 3 s later compared to the other two trial types). Since the delay appears present in rM1 only, but in both protocols, we assume it is both task and region-specific. Furthermore, the delay also appears present in the deconvolved hemodynamic response of the same contrast observed by Aron and Poldrack (2006, their Fig. 2).

Due to the (on average) higher TEs of the multi echo sequence, the optimally combined multi echo BOLD-responses reach a higher peak-level percent signal change than the responses in the single echo data. However, this mean increase in signal is accompanied by an increase in model misfit, as quantified by $\sigma^2$. As a consequence, the higher percentage signal change does not necessarily translate into an increase in detection power. In the next section, we formally compare protocols.

### 3.4. GLMs: Whole-brain results

We first fit GLMs using the canonical HRF to each voxel. Group-level whole-brain statistical parametric maps (SPMs) of the contrasts of interest (failed stop – successful stop, successful stop – go, and failed stop – go) are shown in Fig. 5.

Both protocols lead to highly similar SPMs. Clusters of differential BOLD-responses between failed stop – successful stop are present in IFG, insula, preSMA, dorsal anterior cingulate cortex (dACC), and parietal cortex. The single echo data further appears to show a negative cluster around M1, which is not visible in the combined multi echo data. Subcortically, a differential BOLD-response is present in thalamus, extending to substantia nigra (SN). The subcortical BOLD-responses are widespread and not confined to individual subcortical nuclei, as has been observed before (De Hollander et al., 2017).

The successful stop – go contrast shows relatively little differential BOLD-responses. There is a small positive cluster in IFG and insula, which appears slightly stronger in the single echo data than in the combined multi echo data. Both data sets show a negative cluster around M1 for this contrast. Finally, the pattern of clusters in the failed stop – successful stop cluster is similar to (although somewhat weaker than) the failed stop – go contrast, with clusters of activation present in IFG, preSMA, ACC, thalamus, STN, and SN. The clusters in this contrast appear more widespread and stronger in the combined multi echo data than in the single echo data. Note, however, that the failed stop – successful stop contrast is based on least trials due to the design of the stop-signal response paradigm (approximately 12.5% of all trials are failed stops, and another 12.5% of all trials are successful stops).

#### 3.5. GLMs: ROI analyses

To statistically quantify the differences between the protocols, we first fit a set of GLMs using the canonical HRF to the timeseries within each ROI. Per session, for each participant and ROI, we calculated the $t$-values of each trial type regressor against baseline (Fig. 6, diagonal). Using a linear mixed effects model, we subsequently tested whether ROI, protocol, and trial type (and interaction effects) had an effect on the $t$-value. The results are shown in Table 4 (upper half). Reported $q$-values are $p$-values adjusted for the false discovery rate (Yekutieli and Benjamini, 1999), although the exact same conclusions would be reached without correcting for multiple comparisons. Most importantly, there was no evidence for a main effect of protocol on $t$-values, in line with the marginal difference between mean $t$-values for both protocols (95% confidence intervals: single echo [2.48, 5.20], combined multi echo [2.47, 5.19]). There was also no evidence for an interaction effect of protocol with other terms in the model.

Furthermore, follow-up paired $t$-tests per trial type and ROI showed no significant differences between protocols after multiple comparison corrections. Only without multiple comparisons corrections (and assuming an alpha-level of 0.05), one would have concluded that there was an increase in $t$-values in the multi echo protocol compared to the single echo protocol in rPresMA for failed stop ($t(16) = -2.242, p = 0.04$) and successful stop ($t(16) = -2.50, p = 0.024$) trials. Vice versa, without multiple comparison correction, one would have concluded there was a decrease in $t$-values in the multi echo protocol in rGPe for go ($t(16) = 3.184, p = 0.006$) and successful stop trials ($t(16) = 2.493, p = 0.024$).

To quantify the strength of the evidence against an effect of protocol on first-level $t$-values, we computed Bayes factors by comparing linear

![Fig. 6. Comparison of $t$-values between protocols. Bar graphs illustrate $t$-values per ROI for the main effects (diagonal) and contrasts (lower left). Error bars indicate 67% confidence intervals, obtained by bootstrapping with 1000 bootstrap replications. SPMs ($y = -13$) are included from Fig. 5 for comparison.](image-url)
models (with subjects as random effects). The results are presented in Table 5 (upper half). The model with the overall highest Bayes factor (against a model including the intercept only) included main effects of trial type and ROI, and an interaction between trial type and ROI. Model comparison showed that there was substantial to very strong evidence against an additional effect of protocol and/or interactions between protocol and ROI on t-values.

Finally, we repeated both the frequentist as the Bayesian analyses using the three contrasts of interest as independent variables, and the corresponding t-values as dependent variable (Fig. 6, lower left triangle). The main results are the same as in the previous analyses. The frequentist tests (Table 4, lower half) showed no evidence for an effect of protocol (or interactions between other terms) on these t-values (95% confidence intervals: single echo [0.82, 1.34], combined multi echo [1.02, 1.54]). Similarly, the Bayesian model with the highest overall Bayes Factor included a main effect of ROI, contrast, and an interaction between ROI and contrast. Bayesian model comparison indicated that there was anecdotal to decisive evidence against an additional effect of protocol and/or interactions between protocol and ROI. In summary, the differences between the two protocols were marginal.

Using the individual echoes of the multi echo protocol (Fig. 7, left panel), we tested the effect of echo time on BOLD-sensitivity for each ROI. For each subject, we used a second linear model to predict Δt/TE = f(Δt/TE) for each ROI. This may explain why the single echo sequence still provides good BOLD-sensitivity in cortical areas despite the low TE.

### Table 4
Frequentist F-tests on the effect of terms in the linear mixed effects models, using Satterthwaite’s method (Satterthwaite, 1941) to approximate the denominator degrees of freedom. Both models included random intercepts for subjects. q-values are p-values adjusted for the false discovery rate, and are considered significant if q < 0.05. The same conclusions would be reached without any multiple comparison correction.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Term</th>
<th>Frequentist test</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-value of trial type</td>
<td>Protocol</td>
<td>F(1, 656) = 0.0001, q = 0.99</td>
</tr>
<tr>
<td>Trial type</td>
<td>F(2, 656) = 56.47, q &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>ROI</td>
<td>F(6, 656) = 80.23, q &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>Protocol x Trial type</td>
<td>F(2, 656) = 0.19, q = 0.69</td>
<td></td>
</tr>
<tr>
<td>Protocol x ROI</td>
<td>F(6, 656) = 1.73, q = 0.20</td>
<td></td>
</tr>
<tr>
<td>ROI x Trial type</td>
<td>F(12, 656) = 3.13, q &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>ROI x Trial type x Protocol</td>
<td>F(12, 656) = 0.99, q = 0.99</td>
<td></td>
</tr>
<tr>
<td>Protocol</td>
<td>F(1, 656) = 3.16, q = 0.13</td>
<td></td>
</tr>
<tr>
<td>Contrast</td>
<td>F(2, 656) = 7.92, q &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>ROI</td>
<td>F(6, 656) = 26.25, q &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>Protocol x Contrast</td>
<td>F(2, 656) = 0.53, q = 0.68</td>
<td></td>
</tr>
<tr>
<td>Protocol x ROI</td>
<td>F(6, 656) = 0.85, q = 0.66</td>
<td></td>
</tr>
<tr>
<td>ROI x Contrast</td>
<td>F(12, 656) = 9.99, q &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>ROI x Contrast x Protocol</td>
<td>F(12, 656) = 0.50, q = 0.92</td>
<td></td>
</tr>
</tbody>
</table>

### Table 5
Results of the Bayesian linear mixed effects models. All models included random intercepts for subjects. H0 was the overall winning model (against a model including only an intercept). Bayes Factors (BF01) quantify the relative likelihood of the data under two models; e.g., a BF01 = 10 entails that the data is 10 times more likely under the null hypothesis H0 than under the alternative hypothesis H1. Verbal qualifications of the strength evidence derived from the Bayes Factors (“anecdotal” to “decisive” evidence) are based on taken from Wetzels and Wagenmakers (2012), who adjusted Jeffreys’ (1961) original formulation.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Model</th>
<th>Bayesian test</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-value of trial type</td>
<td>H0: t - ROI + Trial type + ROI x Protocol</td>
<td>BF01 &lt; 0.58%</td>
</tr>
<tr>
<td>H1: t - ROI + Trial type + ROI x Protocol + Contrast</td>
<td>BF01 &gt; 98.67%</td>
<td></td>
</tr>
<tr>
<td>H1: t - ROI + Trial type + ROI x Protocol + Contrast + Protocol x ROI</td>
<td>BF01 &gt; 0.88%</td>
<td></td>
</tr>
<tr>
<td>Trial type</td>
<td>BF01 &gt; 11.78%</td>
<td></td>
</tr>
<tr>
<td>Contrast</td>
<td>BF01 &gt; 7.98%</td>
<td></td>
</tr>
<tr>
<td>ROI</td>
<td>BF01 &gt; 98.67%</td>
<td></td>
</tr>
<tr>
<td>Protocol</td>
<td>BF01 &gt; 11.78%</td>
<td></td>
</tr>
<tr>
<td>ROI x Contrast</td>
<td>BF01 &gt; 0.88%</td>
<td></td>
</tr>
<tr>
<td>Protocol x ROI</td>
<td>BF01 &gt; 0.88%</td>
<td></td>
</tr>
<tr>
<td>H1: t - ROI + Contrast + ROI x Protocol + Contrast</td>
<td>BF01 &gt; 62.43%</td>
<td></td>
</tr>
<tr>
<td>H1: t - ROI + Contrast + ROI x Protocol + Contrast + Protocol x ROI</td>
<td>BF01 &gt; 0.88%</td>
<td></td>
</tr>
</tbody>
</table>

### 4. Discussion
When studying both cortical and subcortical structures using conventional single echo 2D EPI fMRI protocols, researchers are faced with a tradeoff between BOLD-sensitivity in the cortex and in the subcortex. Here, we studied whether a multi echo EPI protocol can be used to avoid this tradeoff and obtain optimal BOLD-sensitivity across the brain. We directly compared a multi echo protocol with a single echo protocol, both optimized for subcortex by using a relatively high spatial resolution combined with short echo times. Our results showed only marginal differences between protocols, which both led to similar conclusions about BOLD-activity in the stop-signal task. The relatively small difference between protocols appears to be caused by two factors.

Firstly, the advantage of combining multiple volumes in the multi echo sequence is counteracted by the relatively low tSNR per echo. Contrary to earlier empirical studies that employed different experimental designs (Poser and Norris, 2009; Posse et al., 1999; Puckett et al., 2018), this resulted in a CNR that was similar to (in cortical areas) or even lower than (in subcortical areas) the CNR of the single echo sequence. An important difference with previous experiments is the amount of acceleration used. Earlier studies typically kept the amount of acceleration identical between sequences and minimized TR, whereas we kept the TRs identical and minimized the amount of acceleration. Acquiring multiple echoes after one excitation pulse with an isotropic spatial resolution of 1.6 mm required the otherwise disadvantageous usages of highly accelerated imaging and an increase in the receiver bandwidth. An increased acceleration factor not only decreases the number of acquired k-space lines but also increases the so-called g-factor penalty (Larkman and Nunes, 2007) especially at the center of the brain. Increasing the GRAPPA acceleration from IPAT = 3 to 4 while using our 32-channel phased array headcoil led to a g-factor increase of approximately 30% at the position of the STN (see Appendix C). This is in line with similar technical setups using GRAPPA (Salomon et al., 2014) or SENSE (Sengupta et al., 2016) for unfolding the aliased data (although differences in slice location, phase encoding direction, and IPAT factor hinder an exact comparison). As a result, both the reduced amount of
acquired k-space and the increased g-factor penalty lead to reduced SNR in highly accelerated acquisitions (see equation 8 in Larkman and Nunes, 2007) and finally to reduced t-values in the activation maps obtained with the multi echo approach. The multiband factor 2 in the multi echo protocol likely also had a disadvantageous effect on the SNR, although previous studies at different field strengths showed that the effect of low multiband factors is relatively small, especially when CAIPI shifts are used (Demetriou et al., 2018; Preibisch et al., 2015; Setsompop et al., 2012; Todd et al., 2017, 2016; Uğurbil et al., 2013; Xu et al., 2013), compared to the effect of increasing iPAT from 3 to 4. Also in combination with iPAT 2, the effect of low multiband factors of 2–3 has been shown to be relatively limited (Demetriou et al., 2018; Setsompop et al., 2012), although more impact on g-factors is expected in combination with iPAT factors of 3–4. In addition to in-plane and slice acceleration, the increased receiver bandwidth in the multi echo protocol (compared to the single echo protocol) resulted in an increased amount of ramp sampling. This effect is discussed in section 2.3. The complex interactions between acceleration and physiology leads to highly non-uniform effects on BOLD-sensitivity across the brain, often with a large decline in BOLD-sensitivity in subcortical areas with higher acceleration (Todd et al., 2017).

Secondly, the CNR of the single echo protocol in cortical areas was higher than what may have been expected based on its short TE. Per-echo analyses of the multi echo data also showed similar first-level t-values across TEs for most cortical regions, in line with previous reports (Gorno-Tempini et al., 2002; Hyde et al., 2001; Weiskopf et al., 2005). In particular, it has been suggested that physiological noise underlies the relatively small effect of echo time on BOLD-sensitivity (Van De Moortele et al., 2008).

We specifically compared the BOLD-sensitivity of two protocols with high spatial resolution and whole-brain coverage. Our task-based analyses indicate that, under these requirements, the protocols provided similar BOLD-sensitivity. However, the single echo protocol used in our study could be optimized further and has the potential for additional gains in statistical power. With the amount of acceleration used in the present study, for example, the temporal resolution can already be increased, and gains in statistical power may potentially be obtained by the use of multiband acceleration (Larkman et al., 2001; Moeller et al., 2010). In contrast, although the contrast-to-noise of the multi echo protocol clearly benefitted from the combination of multiple echoes, the individual echoes themselves had relatively low tSNR. Further increasing the amount of acceleration by increasing the multiband factor to 3 (or higher), while retaining the GRAPPA factor of 4 (which could not be decreased without substantially increasing the echo times), would likely worsen the conditioning of the parallel imaging solution and increase temporal instabilities and therefore decrease the tSNR. This would thus be disadvantageous for the contrast-to-noise, although gains in statistical power might be obtained from the increased total amount of volumes. One may have specific reasons to prefer multi echo acquisition. One example is the reduced susceptibility-related signal loss compared to standard single echo 2D EPI sequences (Kirilina et al., 2016). Another reason may be the promise of denoising techniques. Recent papers (Kundu et al., 2017, 2012) suggested that denoising methods based on multiple echoes can significantly benefit statistical analyses by filtering out components in signal that are not TE-dependent and thus unlikely to be caused by fluctuations in tissue $T_2^*$. Unfortunately, the application of the current publicly available code for this technique on our data led to no detected BOLD-signals. We speculate that this may be caused by the combination of high spatial resolution and low temporal resolution (1.6 mm and 3 s vs. 2.5 mm and 1.8 s in Kundu et al., 2017), which leads to lower tSNR per voxel, as well as fewer volumes (data points) to assess TE-dependence. Thus, we think that the inability to detect BOLD-responses is likely due to the peculiar type of data we collected, more than a shortcoming of the denoising approach, and hope that the techniques are further developed to handle this type of data.

Deconvolved cortical and subcortical responses showed an interesting difference in both experiments: The undershoots of subcortical BOLD-responses appear large relative to the undershoots of the cortical BOLD-responses. This was observed in the data of both sequences and in each echo separately. It is also in line with the deconvolved BOLD-responses observed in De Hollander et al. (2017, their Fig. 6). Substantial interindividual differences (Aguirre et al., 1998; D’Esposito et al., 1999; Handwerker et al., 2004) and interregional differences (Boillat and Van der Zwaag, 2019; Handwerker et al., 2004; Miezin et al., 2000) in the HRF have been reported before. It is important to investigate in future research whether the large undershoots are a consequence of the specific neural response evoked by the task used in the current study, or perhaps a signature of a different hemodynamic response than often assumed, for example caused by differences in vasculature. Such studies would require specifically designed experiments, preferably using multiple behavioral paradigms. If a more substantial investigation of subcortical hemodynamics would confirm these observations, the canonical HRF may be adjusted for these areas. Since mismatches between the assumed hemodynamic response function in a GLM and the actual response inflates the noise term of the model, this may improve statistical power. However, because the larger undershoots were present in both protocols of the present study, it is unlikely that they biased the protocol comparison in the present study.

The GLM results with respect to the stop-signal task replicate earlier results using the same paradigm (Aron and Poldrack, 2006; De Hollander et al., 2017; Li et al., 2008). For example, we replicated the typical finding that RFG (as well as insula) are more active during failed stops compared to go trials and successful stops (Aron et al., 2014). However, we observed an increase in STN activity in failed stop trials compared to successful stop and go trials. This is in line with some (De Hollander et al., 2017; Li et al., 2008), but not all (Aron and Poldrack, 2006) earlier
studies, and it is an open question whether this observation can be reconciled with current theories of the STN’s functioning in the stop-signal task (Aron and Poldrack, 2006).

A potential limitation of the current study is that we did not correct for potential cardiac and respiratory artefacts. High-frequency noise such as cardiac and respiratory artefacts are prominent in inferior parts of the brain (e.g., Barry et al., 2013; Todd et al., 2017). However, these effects would only bias the obtained t-values if the cardiac and respiratory responses are spuriously correlated with the task design. Furthermore, in an earlier study with a very similar single echo protocol and the same behavioral paradigm, it was shown that removing respiratory and cardiac effects from the signal using RETROICOR (Glaver et al., 2000) had only marginal effects on tSNR and task-based t-statistics (De Hollander et al., 2017). This is in line with an earlier report showing that only a small part of the signal variance in midbrain regions could be explained by cardiac and respiratory effects (Barry et al., 2013).

In conclusion, we observed at best marginal differences between the BOLD-sensitivity of a protocol optimized for subcortex and a multi echo protocol that both retain whole-brain coverage and have a high spatial resolution. A single echo protocol attuned to subcortical areas (with a reduced TE and high spatial resolution) retained adequate sensitivity in the cortex. This single echo protocol can be optimized further (e.g., increased temporal resolution, with potential benefits from multi-band acceleration), and the resulting data are better suited for established procedures of data processing and analyses. Therefore, when studying small, subcortical nuclei in parallel with cortical areas with current typical 7 T hardware, we recommend to use an optimized single echo sequence unless one has additional reasons (e.g., reduced susceptibility-related signal loss or when one plans to use denoising techniques that rely on multiple contrasts) to collect multiple echoes.

Code & data availability

All analysis code can be found at https://github.com/IMCN-UvA/fMRI_protocol_comparison_se.me. All data will be uploaded to https://osf.io/6qwjz/ upon publication.

Declarations of competing interest

The Max Planck Institute for Human Cognitive and Brain Sciences has an institutional research agreement with Siemens Healthcare. NW was a speaker at an event organized by Siemens Healthcare and was reimbursed for the travel expenses.

CRediT authorship contribution statement

Steven Miletic: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization. Pierre-Louis Bazin: Methodology, Investigation, Software, Supervision, Writing - review & editing. Nikolaus Weiskopf: Resources, Supervision, Project administration, Funding acquisition, Writing - review & editing, Wietse van der Zwaag: Resources, Investigation, Writing - review & editing. Birte U. Forstmann: Conceptualization, Methodology, Resources, Supervision, Project administration, Funding acquisition, Writing - review & editing. Robert Trampel: Conceptualization, Methodology, Investigation, Resources, Software, Project administration, Writing - review & editing, Supervision.

Acknowledgements

We would like to thank Anke Kummer, Mandy Jochemko, Simone Wipper, and Domenica Willting for their help collecting the data. We would like to thank Felix Breuer and Renzo Huber for insightful discussions. This work was supported by an NWO-Vici grant (BUF). NW received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ ERC grant agreement n° 616905; from the European Union’s Horizon 2020 Research and Innovation Programme under the grant agreement No 681094; from the BMBF (01EW1711A & B) in the framework of ERA-NET NEURON.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.116992.

References


De Hollander, G., Keucen, M.C., van der Zwaag, W., Forstmann, B.U., Trampel, R., 2017. Comparing functional MRI protocols for small, iron-rich basal ganglia nuclei such as the subthalamic nucleus at 7T and 3T. Hum Brain Mapp. 38, 3226–3248. https://doi.org/10.1002/hbm.23586.


