Supplementary Materials for Group-level impacts of within- and between-subject hemodynamic variability in fMRI

Solveig Badillo[a,b], Thomas Vincent[a,c], and Philippe Ciuciu[a,b,*]

[a] *CEA/DSV/I^{2}BM/NeuroSpin, CEA Saclay, Bât. 145, Point Courrier 156, 91191 Gif-sur-Yvette cedex, France*
[b] *Parietal team, INRIA Saclay Ile-de-France, NeuroSpin, Bât 145, 91191 Gif-sur-Yvette cedex, France*
[c] *Mistis team, Inria Grenoble Rhône-Alpes, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier Cedex, France*

Appendix A. Complementary results in activation detection performance: A - V contrast

For the **A - V** contrast and $R = 2$, all group-level analyses found large bilateral and approximately symmetric activation clusters in the temporal lobes, as illustrated in Fig. A.1(a)-(c). All statistical analyses were significant irrespective of the intra-subject modeling, with larger t-values for GLM and SAGLM-based analyses. For instance, the largest activation clusters were obtained using the SAGLM modeling. JDE-based inference provides us with less sensitive yet significant results; see Tab. A.1[Top] for details.

As regards $R = 4$, the activation clusters recovered from the different analyses have also approximately the same localization along the superior temporal sulci (auditory cortices). However, their size differences are much more important than those we found for $R = 2$: JDE-based inference yield very small but significant clusters compared to those that obtained with the GLM or SAGLM counterparts: see Tab. A.1[Bottom] for details. The latter actually gave more sensitive results, both at the cluster and voxel levels with $p_{\text{val}} < 0.05$. In contrast to our observations for **Lc - Rc** (see subsection 5.1.1 of the paper), here the sensitivity at the voxel and cluster levels appeared higher for the GLM-based analysis. These results showed that a strong loss of sensitivity is induced by estimating the HRF shape in auditory cortices. As advocated in [1], this seems to also indicate that the canonical HRF shape is already a good candidate for the auditory system. The same conclusions can be drawn from the **V - A** contrast (results not shown).

*Corresponding author (philippe.ciuciu@cea.fr).

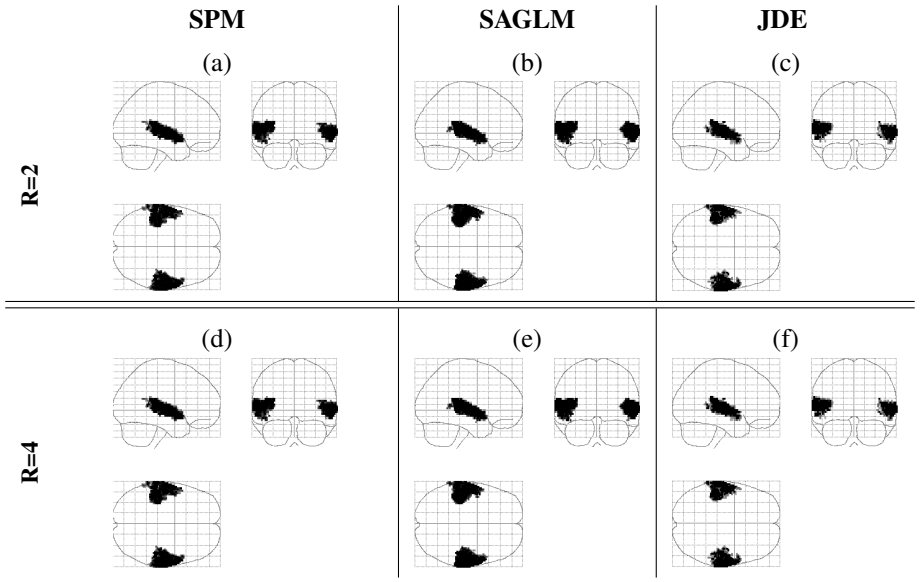

Figure A.1: RFX statistical analysis for the **A - V** contrast and acceleration factors $R = 2$**(a-c)** and $R = 4$ **(d-f)**, derived using the GLM **(a,d)**, SAGLM **(b,e)** and JDE **(c,f)** subject-level analyses. Statistical maps were thresholded at $p_{\text{val}}^{\text{vox}} < 0.001$ uncorrected, and corrected for multiple comparisons at $p_{\text{val}}^{\text{clust}} < 0.05$, using calibration by permutations.

Table A.1: Group-level comparison for the **A - V** contrast using the JDE, SAGLM and GLM subject-level inferences ($R = 2$ and $R = 4$). Cluster-level FWER correction at $p_{\text{val}}^{\text{clust}} < 0.05$. Significant values are reported in **bold font** ($p_{\text{val}} < 0.05$).

		Group-level statistical results				
		Cluster-level		Voxel-level (peak)		
	Subj.-level inference	corr. $p_{\text{val}}^{\text{clust}}$	Size	corr. $p_{\text{val}}^{\text{vox}}$	t-score	Position (mm)
$R = 2$	GLM	$\mathbf{< 0.001}$	**1023**	$\mathbf{< 0.001}$	**12.57**	[64 -10 0]
		$\mathbf{< 0.001}$	**1003**	$\mathbf{< 0.001}$	**9.42**	[-60 -8 -6]
	SAGLM	$\mathbf{< 0.001}$	**1108**	$\mathbf{< 0.001}$	**10.6**	[64 0 -3]
		$\mathbf{< 0.001}$	**1135**	**0.005**	**10.12**	[-44 -24 9]
	JDE	**0.001**	**909**	**0.001**	**9.96**	[-68 -14 3]
		$\mathbf{< 0.001}$	**903**	**0.008**	**8.79**	[-42 -30 12]
$R = 4$	GLM	$\mathbf{< 0.001}$	**597**	$\mathbf{< 0.001}$	**12.33**	[56 4 -3]
		$\mathbf{< 0.001}$	**543**	**0.007**	**8.65**	[-44 -28 9]
	SAGLM	$\mathbf{< 0.001}$	**505**	**0.009**	**8.52**	[44 -26 12]
		$\mathbf{< 0.001}$	**447**	**0.017**	**8.18**	[-42 -18 6]
	JDE	**0.04**	**28**	0.068	6.98	[64 -8 0]
		0.005	**65**	0.09	6.76	[-60 -14 9]

Appendix B. Comparative study for hemodynamic inference

To complete our investigations concerning the performance of different BOLD signal modeling in terms of HRF estimation, we have performed a comparative study at the subject-level between the parcelwise JDE approach and the voxelwise regularized Finite Impulse Response (RFIR) technique described in [2, 3]. The major goal that sustains this comparison is to measure the actual impact of introducing regional spatial aggregation in JDE-based HRF inference in terms of SNR and to compare these results with those deriving from spatial averaging of RFIR HRF estimates, the latter only relying on temporal regularization. In other words, this consists of testing how tenable is the assumption that all voxels of a given parcel share the same HRF shape.

The unsupervised RFIR approach [2]. For the sake of completeness, let us briefly summarize how the voxelwise RFIR approach proceeds. The generative BOLD signal modeling in the FIR context is linear, time-invariant (convolution model) and additive over experimental conditions and thus reads as follows:

$$\forall j \in [\![1;J]\!], \qquad \boldsymbol{y}_j = \sum_{m=1}^{M} \boldsymbol{X}^m \boldsymbol{h}_j^m + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{b}_j = \mathbb{X}\mathbb{h}_j + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{b}_j \qquad \text{(B.1)}$$

In Eq. (B.1), vector $\boldsymbol{h}_j^m = \left(h_{j,d\Delta t}^m\right)_{d=0,..,D}^{\mathrm{t}}$ represents the unknown HRF time course in voxel j which is associated with the m^{th} experimental condition and sampled every Δt. J corresponds to the total number of voxels considered in the functional mask of the brain. Taken together through $\mathbb{h}_j = \left[(\boldsymbol{h}_j^1)^{\mathrm{t}}, \ldots, (\boldsymbol{h}_j^M)^{\mathrm{t}}\right]^{\mathrm{t}}$, there are $M \times (D+1)$ unknown HRF coefficients to be estimated. As in the JDE formulation, $\boldsymbol{X}^m = \left[\boldsymbol{x}_{t_1}^m, \ldots, \boldsymbol{x}_{t_N}^m\right]^{\mathrm{t}}$ is the $N \times (D+1)$ binary occurrence matrix consisting of the lagged stimulus covariates $\boldsymbol{x}_{t_n}^m = \left(x_{t_n - d\Delta t}^m\right)_{0 \leqslant d \leqslant D}^{\mathrm{t}}$. Consequently, the global occurrence matrix reads $\mathbb{X} = \left[\boldsymbol{X}^1 \mid \cdots \mid \boldsymbol{X}^M\right]$. Matrix $\boldsymbol{P} = \left[\boldsymbol{p}_1, \ldots, \boldsymbol{p}_G\right]$ of size $N \times G$ are the values at times t_n of an orthonormal basis (*i.e.,* $\boldsymbol{P}^{\mathrm{t}}\boldsymbol{P} = \boldsymbol{I}_G$) consisting of G functions $\boldsymbol{p}_g = (\boldsymbol{p}_{g,t_n})^{\mathrm{t}}$ that take a potential drift and any other nuisance effect into account. Vector $\boldsymbol{\ell}_j = (l_{g,j})_{1 \leqslant g \leqslant G}^{\mathrm{t}}$ contains the corresponding unknown coefficients in j. Vector $\boldsymbol{b}_j = (b_{j,t_n})^{\mathrm{t}}$ defines the noise term in voxel j and is supposed to be white, normally-distributed with variance $v_{\boldsymbol{b}_j}$ and independent of the HRFs.

Most of reliable RFIR estimators have taken place in the Bayesian framework and constraint the non-parametric HRF to be temporally smooth [4, 5, 2, 3, 6, 7] using a Gaussian prior density for the HRF and by automatically tuning the amount of regularization using either full Bayesian inference [5, 3] schemes or Expectation-Maximization (EM) algorithms [2, 6, 7]. The most often used prior was $\mathbb{h}_j \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \mathrm{diag}\left[\Sigma_1, \ldots, \Sigma_M\right]$, $\Sigma_m = v_{\boldsymbol{h}_j^m}\boldsymbol{R}$ and $\boldsymbol{R} = (\boldsymbol{D}_2^{\mathrm{t}}\boldsymbol{D}_2)^{-1}$ where $\boldsymbol{D}_2$ is the second-order finite difference matrix enforcing local smoothness by penalizing abrupt changes quadratically. This prior has ensured stable recovery of HRF profiles according to the maximum a posteriori (MAP) voxelwise estimator:

$$\forall j = 1:J, \quad \left\{ \begin{array}{rcl} \widehat{\mathbb{h}}_j^{\mathrm{MAP}} & = & \left(\mathbb{X}^{\mathrm{t}}\mathbb{X} + v_{\boldsymbol{b}_j}\boldsymbol{\Sigma}^{-1}\right)^{-1}\mathbb{X}^{\mathrm{t}}(\boldsymbol{y}_j - \boldsymbol{P}\widehat{\boldsymbol{\ell}}_j^{\mathrm{ML}}) \\ \widehat{\boldsymbol{\ell}}_j^{\mathrm{ML}} & = & \boldsymbol{P}^{\mathrm{t}}\left(\boldsymbol{y}_j - \mathbb{X}^{\mathrm{t}}\widehat{\mathbb{h}}_j^{\mathrm{MAP}}\right). \end{array} \right. \qquad \text{(B.2)}$$

where the nuisance variables $\boldsymbol{\ell}_j$ are jointly estimated with the HRFs $\mathbb{h}_j$ in every voxel j but in the maximum likelihood (ML) sense. For computational and inference details about the variance parameters, see [2].

Intra-subject comparison between RFIR and JDE. In this comparative study, we considered one subject's data set acquired with $R = 2$ (higher SNR). To make sense, our analysis took place in brain regions eliciting evoked activity during the localizer paradigm. For the sake of consistency, we therefore investigated the BOLD response in the temporal, right motor and parietal regions as already done in Subsection 4.4 of the paper. The selected parcels were chosen according to their activation extent inferred from JDE-based analysis, i.e. as the regions-of-interest (ROIs) with the largest number of activated voxels. Fig. B.2(a)-(c) shows the localization of the selected parcels. Moreover, in each parcel, we focused on the two experimental conditions that induced the most significant BOLD effects in terms of NRL estimates (JDE inference). This allowed us to infer the corresponding RFIR HRF estimates. More precisely, we paid attention to the *i.)* "Auditory Sentences" (**AS**) and "Auditory Right Click" (**ARc**) conditions in the **left temporal** cortex, the *ii.)* "Auditory Left Click" (**ALc**) and "Visual Left Click" (**VLc**) conditions in the **right motor** cortex and to the *iii.)* "Auditory Computation" (**AC**) and "Visual Computation" (**VC**) conditions in the **left parietal** cortex.

Fig. B.2(d)-(f) depicts the corresponding RFIR and JDE-based HRF estimates. As regards RFIR inference, the spatially average HRF time courses are plotted for the two conditions of interest together with their standard deviations ($\pm\sigma$) in order to report how variable in space the voxelwise RFIR-estimates are.

As already observed with JDE-based HRF inference in Section 5.2 of the paper, the closest average RFIR estimate to the canonical shape is retrieved in the temporal parcel (Fig. B.2(d), **AS** condition, blue curve), whereas the average RFIR HRF profiles showing the strongest discrepancy with the canonical version are found in the parietal parcel irrespective of the experimental condition. In the motor parcel, the RFIR estimates show an intermediate behavior.

As illustrated in Fig. B.2(c), there is a converging evidence to the same *non-canonical* HRF time course between the RFIR and JDE inference schemes in the parietal parcel. In addition, the RFIR HRF profiles reported for the two conditions of interest (auditory vs. visually-induced calculations) are very close to each other. This confirms that the parietal cortex performs *supramodal calculations* which are independent of the modality used for stimulus presentation. These results highlight that performing spatial aggregation as proposed by JDE inference or assuming the same HRF profile over the whole parcel and for all experimental conditions is a tenable assumption in the parietal cortex. This strategy thus enables to decrease the number of unknown HRF parameters to be estimated and leaves the available supplementary degrees of freedom to perform robust detection of evoked activity in the JDE formalism. In contrast, stronger differences between JDE and RFIR HRF estimates were observed in the temporal and motor parcels (see Fig. B.2(d)-(f)). Moreover, for the RFIR HRF estimates computed over the temporal and motor parcels, we noticed some discrepancy between conditions in terms of time-to-peak (TTP) and full width at half-maximum (FWHM). Also, the average RFIR HRF time courses associated with the **AS** and **ALc** conditions in the temporal and motor parcels suffer from spurious artifacts after the undershoot. The

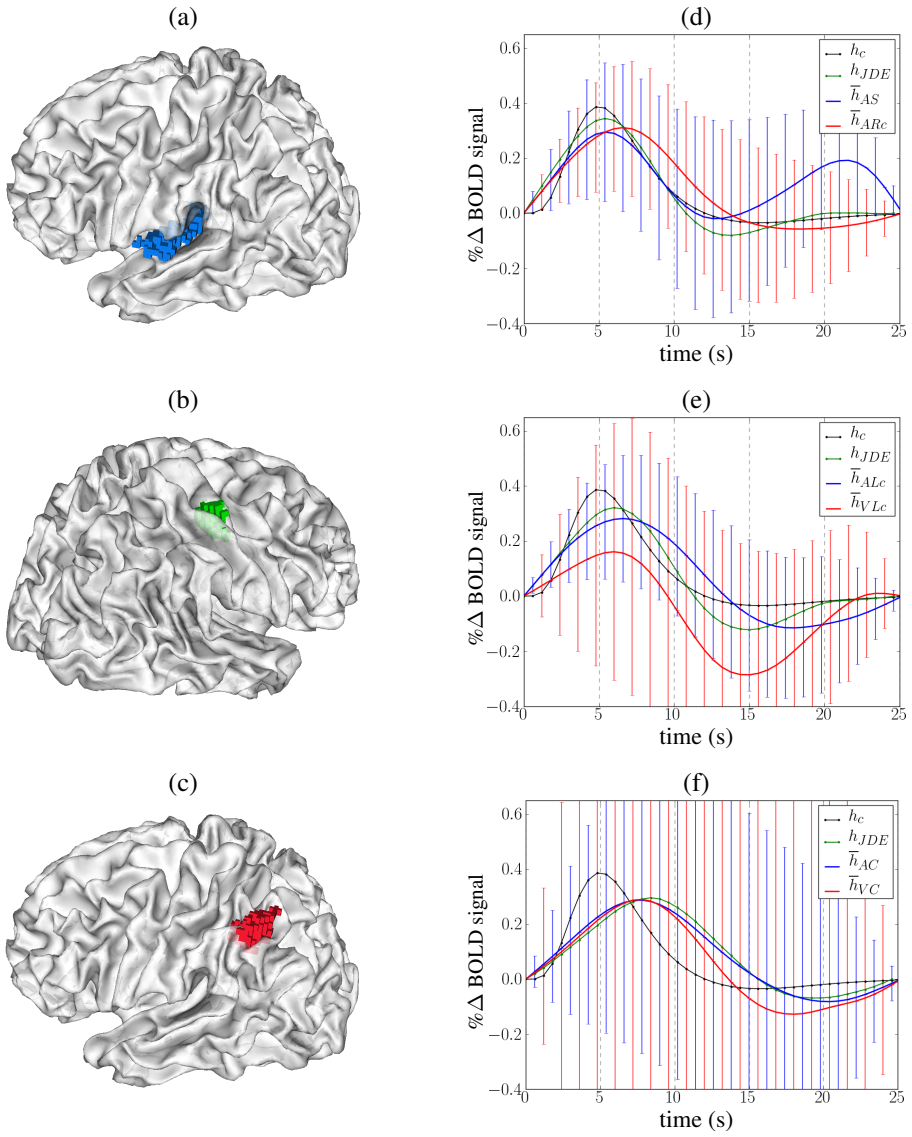

Figure B.2: Parcel-average RFIR HRF estimates of one subject with superimposed error bars ($\pm\sigma$) for the left temporal **(top row)**(189 voxels), right motor **(middle row)**(151 voxels) and left parietal **(bottom row)**(200 voxels) regions. **(Left column)**: localization of regions, respectively, superimposed on the segmented cortical fold. In each parcel, the RFIR HRF estimates corresponding to two conditions eliciting the largest evoked activity are plotted in blue and red. The JDE-based HRF estimates are plotted in green and the canonical HRF $\boldsymbol{h}_c$ in black. To make the comparison easier, all panels are scaled within the same variation range. In **(c)**, exceeding error bars range from ... to ....

Table B.2: Information on standard deviations of RFIR HRF profiles shown in Fig. B.2. The mean ($\bar{\sigma}$) and maximum ($\sigma_{\max}$) values of standard deviations are reported in terms of $\%\Delta$ BOLD signal.

Region r	Conditions	Mean std $\bar{\sigma}$	Max std $\sigma_{\max}$ (corresponding time point in s.)
Temporal	**AS**	0.1	0.36 (12.6)
	ARc	0.1	0.3 (13.2)
Motor	**ALc**	0.1	0.28 (13.8)
	VLc	0.18	0.51 (8.4)
Parietal	**AC**	0.2	0.6 (12.6)
	VC	0.52	1.53 (14.4)

latter seems non-physiological and shows that estimating several HRFs per voxel may turn out unreliable owing to noise overfitting issue.

Besides, through large error bars Fig. B.2(d)-(f) indicate a pretty strong spatial variability of voxelwise RFIR HRF estimates, in all brain parcels that we consider and despite their size difference (between 150 and 200 voxels). For illustration purposes, the full length of the error bars does not appear in all figures, so as to better visualize the HRF profiles. Complementary information on standard deviations is given in Tab. B.2. In agreement with our JDE-based results in Subsection 5.2 of the paper, the largest spatial variability for RFIR estimates was found in the parietal parcel for the **AC** condition: the maximum and mean standard deviations were at least three times larger than in other ROIs:$(\sigma_{\max}, \bar{\sigma}) = (1.53, 0.52)\%\Delta$ BOLD signal. This clearly reflects a pretty strong SNR variability in space in the parietal area that makes RFIR identification techniques less robust than the JDE approach. Also, as reported in Tab. B.2, spatial variability of the RFIR HRF profiles significantly fluctuates across the two experimental conditions in the motor and parietal parcels by a factor of at least 1.8: the larger fluctuations were retrieved for the **VLc** and **VC** conditions in the motor and parietal regions and were thus associated with the visual modality. Since the two RFIR HRF time courses in the motor parcel were also quite different, this may indicate that spatial aggregation performed for HRF inference in the JDE formalism is questionable. In the temporal regions, since both conditions refer to the auditory modality, the RFIR HRF time courses show stronger homogeneity.

Appendix C. Complementary statistical comparison of JDE and GLM-based inference schemes

Tab. C.3 provides the complete set of values for group-average HRF estimates at each time point, for all regions and both acquisition factors.

Tab. C.4 provides region-specific group-average values of the TTP and FWHM parameters and brings complementary evidence of what is already shown in Figs. 13-14 of the paper.

In order to assess any significant change of inference-specific hemodynamic parameters between SNR values and regions, we entered the subject-dependent parameter estimates $(\text{TTP}_{r,s}^{R,\psi})$ and $(\text{FWHM}_{r,s}^{R,\psi})$ with $\psi \in \{\text{JDE}, \text{GLM}\}$ in several 2-way repeated measures ANOVAs involving two factors, namely the acceleration factor $R = (2, 4)$

Table C.3: Values of group-average HRF estimates at each time point, for left parietal (P), left motor (M), temporal (T) and occipital (O) regions and for both acceleration factors $R = 2$ and $R = 4$.

	HRF values for $R = 2$				**HRF values for $R = 4$**			
time points (s)	**P**	**M**	**T**	**0**	**P**	**M**	**T**	**0**
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.6	0.019	0.03	0.034	0.021	0.027	0.041	0.034	0.04
1.2	0.041	0.061	0.073	0.051	0.054	0.082	0.073	0.082
1.8	0.069	0.093	0.118	0.099	0.082	0.122	0.118	0.129
2.4	0.1	0.129	0.166	0.157	0.111	0.159	0.166	0.177
3	0.134	0.168	0.216	0.216	0.141	0.194	0.216	0.222
3.6	0.171	0.208	0.264	0.271	0.17	0.226	0.264	0.261
4.2	0.21	0.246	0.306	0.317	0.199	0.253	0.306	0.292
4.8	0.245	0.279	0.334	0.349	0.225	0.275	0.334	0.312
5.4	0.271	0.303	0.344	0.361	0.245	0.289	0.344	0.319
6	0.285	0.315	0.333	0.35	0.259	0.294	0.333	0.312
6.6	0.286	0.317	0.305	0.314	0.266	0.29	0.305	0.291
7.2	0.278	0.306	0.263	0.264	0.266	0.278	0.263	0.26
7.8	0.26	0.283	0.213	0.206	0.26	0.257	0.213	0.223
8.4	0.237	0.251	0.159	0.147	0.25	0.23	0.159	0.182
9	0.21	0.212	0.109	0.091	0.236	0.198	0.109	0.141
9.6	0.182	0.171	0.066	0.042	0.219	0.165	0.066	0.103
10.2	0.156	0.133	0.032	0.006	0.201	0.132	0.032	0.07
10.8	0.131	0.098	0.008	-0.019	0.182	0.102	0.008	0.044
11.4	0.11	0.067	-0.008	-0.032	0.164	0.074	-0.008	0.024
12	0.093	0.041	-0.017	-0.04	0.145	0.051	-0.017	0.009
12.6	0.08	0.021	-0.021	-0.046	0.129	0.033	-0.021	-0.002
13.2	0.072	0.007	-0.022	-0.047	0.115	0.019	-0.022	-0.009
13.8	0.068	-0.002	-0.023	-0.046	0.105	0.009	-0.023	-0.013
14.4	0.065	-0.009	-0.025	-0.045	0.095	0.001	-0.025	-0.015
15	0.062	-0.013	-0.026	-0.047	0.086	-0.004	-0.026	-0.018
15.6	0.058	-0.016	-0.027	-0.049	0.078	-0.008	-0.027	-0.02
16.2	0.055	-0.017	-0.028	-0.052	0.071	-0.011	-0.028	-0.022
16.8	0.053	-0.019	-0.03	-0.053	0.065	-0.014	-0.03	-0.023
17.4	0.05	-0.021	-0.031	-0.053	0.059	-0.017	-0.031	-0.024
18	0.046	-0.024	-0.033	-0.053	0.053	-0.02	-0.033	-0.024
18.6	0.042	-0.026	-0.033	-0.054	0.048	-0.024	-0.033	-0.024
19.2	0.04	-0.028	-0.032	-0.053	0.043	-0.026	-0.032	-0.023
19.8	0.038	-0.029	-0.03	-0.051	0.04	-0.028	-0.03	-0.022
20.4	0.037	-0.031	-0.026	-0.049	0.035	-0.029	-0.026	-0.021
21	0.035	-0.032	-0.021	-0.049	0.03	-0.029	-0.021	-0.021
21.6	0.033	-0.032	-0.016	-0.05	0.025	-0.028	-0.016	-0.021
22.2	0.03	-0.029	-0.01	-0.046	0.02	-0.024	-0.01	-0.019
22.8	0.026	-0.023	-0.007	-0.039	0.015	-0.02	-0.007	-0.015
23.4	0.019	-0.016	-0.005	-0.028	0.011	-0.014	-0.005	-0.011
24	0.01	-0.008	-0.003	-0.015	0.006	-0.008	-0.003	-0.007
24.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table C.4: Group-average $\overline{\text{TTP}}_r^{R,\psi}$ and $\overline{\text{FWHM}}_r^{R,\psi}$ parameter estimates with $\psi = \{\text{JDE}, \text{GLM}\}$, $R = (2, 4)$ computed in all regions $r \in \Upsilon = \{\text{P}, \text{M}, \text{T}, \text{O}\}$. Grand mean and standard deviation over regions are also reported as summary statistics.

		TTP (s)		**FWHM (s)**	
	Region r	$\overline{\text{TTP}}_r^{R,\text{JDE}}$	$\overline{\text{TTP}}_r^{R,\text{GLM}}$	$\overline{\text{FWHM}}_r^{R,\text{JDE}}$	$\overline{\text{FWHM}}_r^{R,\text{GLM}}$
$R = 2$	Parietal	6.68	5.17	9.92	5.1
	Motor	6.4	5.6	7.26	5.08
	Temporal	5.48	5.36	6	5.25
	Occipital	5.4	4.8	5.55	5.21
	Mean	5.99	5.23	7.18	5.16
	Std	0.64	0.34	1.96	0.08
$R = 4$	Parietal	7.48	5	9.63	5.11
	Motor	6.16	5.57	8.36	5.22
	Temporal	6.08	5.47	6.72	5.24
	Occipital	5.68	4.9	7.33	5.18
	Mean	6.35	5.23	8.0	5.19
	Std	0.78	0.33	1.27	0.06

and the region $r \in \Upsilon$. In a first round, these ANOVAs were conducted separately for the TTP and FWHM parameters and also according to each intra-subject inference scheme. Our results are summarized in Tab. C.5 where the significance threshold was set to 0.05. Regarding the JDE-based results, the region factor r drove significant differences for both parameters, whereas the acceleration factor R induced statistically significant differences only for the FWHM. Also, the level of significance was stronger for the region factor. Not surprisingly, GLM-based inference captured the same statistically significant TTP difference between regions. The region factor induced significant variations of FWHM values too but at a lower significance level. No interaction between R parameter and region exceeded the significance threshold irrespective of the inference scheme and the hemodynamic parameter.

In a second round, we performed 2-way repeated measures ANOVAs on the difference of parameter estimates provided by GLM and JDE inference schemes keeping the same two factors as before (acceleration factor R and region r): $\delta\text{TTP}_{r,s}^R = \text{TTP}_{r,s}^{R,\text{JDE}} - \text{TTP}_{r,s}^{R,\text{GLM}}$ and $\delta\text{FWHM}_{r,s}^R = \text{FWHM}_{r,s}^{R,\text{JDE}} - \text{FWHM}_{r,s}^{R,\text{GLM}}$ for $r \in \Upsilon$, $s = 1 : S$ and $R = (2, 4)$. The underlying idea was to assess the overall significance of parameter estimate difference for each hemodynamic feature in turn. Our results are summarized in Tab. C.6. The region factor was again the sole parameter generating significant changes between the TTP differences. As regards FWHM differences, the two factors taken separately achieved significance owing to the large region and SNR-based FWHM variability we already reported for JDE inference in Fig. 14 of the main document and Tab. C.4. Also, no interaction between region and acceleration factor was significant because R had only a limited impact on the hemodynamic parameter estimates in any given region.

Table C.5: 2-way repeated measures ANOVA results based on inference-specific estimates of $\text{TTP}_{r,s}^{R,\psi}$ (top) and $\text{FWHM}_{r,s}^{R,\psi}$ (bottom) for $\psi = \{\text{JDE}, \text{GLM}\}$, $r \in \Upsilon$, $s = 1 : S$ and $R = (2, 4)$. Significant p-values ($p_{\text{val}} < 0.05$) appear in **bold font**.

Inference	Param.	Source	F score	p-val.
JDE-based	$\text{TTP}_{r,s}^{R}$	SNR	2.44	0.14
		Region	9.34	$\mathbf{7.48\ 10^{-5}}$
		SNR × Region	0.87	0.46
JDE-based	$\text{FWHM}_{r,s}^{R}$	SNR	5.04	**0.04**
		Region	12.9	$\mathbf{4.15\ 10^{-6}}$
		SNR × Region	0.95	0.43
GLM-based	$\text{TTP}_{r,s}^{R}$	SNR	0	1
		Region	8.69	$\mathbf{1.33\ 10^{-4}}$
		SNR × Region	0.43	0.73
GLM-based	$\text{FWHM}_{r,s}^{R}$	SNR	0.89	0.36
		Region	2.89	**0.047**
		SNR × Region	1.3	0.29

Table C.6: 2-way repeated measures ANOVA results based on the differences between JDE and GLM-based inference of hemodynamic parameters: $\delta\text{TTP}_{r,s}^{R} = \text{TTP}_{r,s}^{R,\text{JDE}} - \text{TTP}_{r,s}^{R,\text{GLM}}$ (top) and $\delta\text{FWHM}_{r,s}^{R} = \text{FWHM}_{r,s}^{R,\text{JDE}} - \text{FWHM}_{r,s}^{R,\text{GLM}}$ for $r \in \Upsilon$, $s = 1 : S$ and $R = (2, 4)$. Significant p-values ($p_{\text{val}} < 0.05$) appear in **bold font**.

Param.	Source	F score	p-val.
$\delta\text{TTP}_{r,s}^{R}$	SNR	2.6	0.13
	Region	7.45	$\mathbf{4.16\ 10^{-4}}$
	SNR × Region	0.99	0.4
$\delta\text{FWHM}_{r,s}^{R}$	SNR	4.79	**0.045**
	Region	13.52	$\mathbf{2.59\ 10^{-6}}$
	SNR × Region	0.94	0.43

Appendix D. Nonparametric tests over hemodynamic parameters

In this section, we summarize how we performed non-parametric statistical tests at the group-level on the scalar hemodynamic parameters, namely the activation delay and activation duration.

Let θ_s be the parameter of interest for subject s and $\widehat{\theta}_s$ its pointwise estimate provided by any inference scheme. In what follows, we tested the null hypothesis $H_0 : \overline{\theta} = \theta_{\mathrm{c}}$ where $\overline{\theta}$ corresponds to the group-level mean of parameters $\{\theta_s\}_{s=1:S}$ and θ_{c} the reference or canonical value against which the test was performed. Here, we only implemented random-effect analyses and thus we neglected the within-subject variance component. Further attention should be paid if one wants to relax this assumption, for instance by considering mixed effect analysis (see [8] for details).

Under the null hypothesis that the S estimates are independent, identically and symmetrically distributed around the canonical value θ_{c}, the following region-level sign permutation procedure was implemented for each parameter in turn, the TTP and FWHM estimates [9, 10].

The multiple comparison problem was managed by applying Bonferroni correction to control the Familywise Error Rate (FWER) owing to its simplicity and its ability to handle one-tailed and two-tailed nonparametric tests in the same context as well as to change the number of comparisons to be accounted for (regions only or regions and hemodynamic parameters). The corresponding procedure is detailed in Algorithm 1.

Algorithm 1 Nonparametric statistical tests of hemodynamic parameters using FWER control based Bonferroni correction.

Set $\alpha = 0.05$ and the threshold of type-I error rate and K the number of simultaneous hypotheses to be tested: $K = 4$ for controlling the regions of interest only and $K = 8$ if additional control over both TTP and FWHM is performed so that $\alpha_K = \alpha/K$.
Define the $S \times 4$-dimensional zero-mean sample $\boldsymbol{X} = [\boldsymbol{x}_1 \mid \cdots \mid \boldsymbol{x}_4]$ where S is the number of subjects, $\boldsymbol{x}_r = \widehat{\boldsymbol{\theta}}_r - \theta_{\mathrm{c}}$ and $\boldsymbol{\theta}_r = [\theta_{r,1}, \ldots, \theta_{r,S}]^{\mathrm{t}}$ with $\theta =$ TTP or $\theta =$ FWHM and θ_{c} its canonical value.
Compute $\forall r \in \Upsilon$, $T_r^{\mathrm{obs}} = \frac{\mu_r}{\sigma_r/\sqrt{S}}$ where μ_r and σ_r are the empirical mean and standard deviation of $\boldsymbol{x}_r$, respectively.
Set $I = 2^S$ the number of requested permutations depending on the expected accuracy a on the nonparametric p-value: $a = \sqrt{(\alpha - \alpha^2)/I}$ such as for $\alpha = 0.05$, we get $a = 0.2\%$.
for $i = 1$ to I **do**
  for $r = 1$ to 4 **do**
    Randomly permute the sign of vector $\boldsymbol{x}_r^i = \left[x_{r,1}^i, \ldots, x_{r,S}^i\right]^{\mathrm{t}}$ with probability 0.5
  end for
  Get a new $4 \times S$ sample $\boldsymbol{X}^i = \left[\boldsymbol{x}_1^i \mid \cdots \mid \boldsymbol{x}_4^i\right]$
  Compute the t-statistic T_r^i associated with $\boldsymbol{x}_r^i$ in each region $r \in \Upsilon$: $T_r^i = \frac{\mu_r^i}{\sigma_r^i/\sqrt{S}}$ where μ_r^i and σ_r^i are the empirical mean and standard deviation of $\boldsymbol{x}_r^i$, respectively.
end for
Compute uncorrected p-values $\forall r \in \Upsilon$ $p_{r,\mathrm{unc.}}$ as $\mathrm{Card}\left[T_r^i > T_r^{\mathrm{obs}}\right]_i / I$ i.e. the relative frequency of the event $\left\{T_r^i > T_r^{\mathrm{obs}}\right\}_i$ in case of one-tailed *right-sided* test and as $2 \times \mathrm{Card}\left[T_r^i > T_r^{\mathrm{obs}}\right]_i / I$ in case of two-tailed test given the symmetry assumption.
Compute $\forall r \in \Upsilon$, $h_r = I\left(p_{r,\mathrm{unc.}} \leqslant \alpha_K\right)$ with $I(x) = 1$ if x is true and 0 otherwise, and reject the null hypothesis H_0 in region r whenever $h_r = 1$.
return Compute the corrected p-value $p_{r,\mathrm{corr}} = \min(1, K\, p_{r,\mathrm{unc.}})$.

References

[1] G. H. Glover, "Deconvolution of impulse response in event-related BOLD fMRI," *Neuroimage*, vol. 9, pp. 416–429, 1999.

[2] P. Ciuciu, J.-B. Poline, G. Marrelec, J. Idier, Ch. Pallier, and H. Benali, "Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment," *IEEE Trans. Med. Imag.*, vol. 22, no. 10, pp. 1235–1251, Oct. 2003.

[3] G. Marrelec, P. Ciuciu, M. Pélégrini-Issac, and H. Benali, "Estimation of the hemodynamic response function in event-related functional MRI: Bayesian networks as a framework for efficient Bayesian modeling and inference," *IEEE Trans. Med. Imag.*, vol. 23, no. 8, pp. 959–967, Aug. 2004.

[4] C. Goutte, F. A. Nielsen, and L. K. Hansen, "Modeling the haemodynamic response in fMRI using smooth filters," *IEEE Trans. Med. Imag.*, vol. 19, no. 12, pp. 1188–1201, Dec. 2000.

[5] G. Marrelec, H. Benali, P. Ciuciu, M. Pélégrini-Issac, and J.-B. Poline, "Robust Bayesian estimation of the hemodynamic response function in event-related BOLD MRI using basic physiological information," *Hum. Brain Mapp.*, vol. 19, no. 1, pp. 1–17, May 2003.

[6] S. Donnet, M. Lavielle, P. Ciuciu, and J.-B. Poline, "Selection of temporal models for event-related fMRI," in *Proc. 2th Proc. IEEE ISBI*, Arlington, VA, Apr. 2004, pp. 992–995.

[7] P. Ciuciu, J. Idier, A. Roche, and Ch. Pallier, "Outlier detection for robust region-based estimation of the hemodynamic response function in event-related fMRI," in *2th Proc. IEEE ISBI*, Arlington, VA, Apr. 2004, pp. 392–395.

[8] A. Roche, S. Mériaux, M. Keller, and B. Thirion, "Mixed-effects statistics for group analysis in fMRI: A nonparametric maximum likelihood approach," *Neuroimage*, vol. 38, pp. 501–510, 2007.

[9] T.E. Nichols and A.P. Holmes, "Nonparametric permutation tests for functional neuroimaging: A primer with examples," *Hum. Brain Mapp.*, vol. 15, pp. 1–25, 2002.

[10] M. Lindquist, J. Meng Loh, L.Y. Atlas, and T.D. Wager, "Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling," *Neuroimage*, vol. 45, no. 1 Suppl, pp. S187–98, Mar. 2009.