



## Temporal multivariate pattern analysis (tMVPA): A single trial approach exploring the temporal dynamics of the BOLD signal

Luca Vizioli<sup>a,\*</sup>, Alexander Bratch<sup>a,b,1</sup>, Junpeng Lao<sup>c,1</sup>, Kamil Ugurbil<sup>a</sup>, Lars Muckli<sup>d</sup>, Essa Yacoub<sup>a</sup>

<sup>a</sup> Center for Magnetic Resonance Research (CMRR), University of Minnesota, Minneapolis, MN, United States

<sup>b</sup> Department of Psychology, University of Minnesota, Minneapolis, MN, United States

<sup>c</sup> Department of Psychology, University of Fribourg, Fribourg, Switzerland

<sup>d</sup> Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, Scotland, UK

### ARTICLE INFO

#### Keywords:

Multivariate  
Mean BOLD amplitude  
Temporal analysis  
BOLD  
MVPA

### ABSTRACT

**Background:** fMRI provides spatial resolution that is unmatched by non-invasive neuroimaging techniques. Its temporal dynamics however are typically neglected due to the sluggishness of the hemodynamic signal.

**New Methods:** We present *temporal multivariate pattern analysis* (tMVPA), a method for investigating the temporal evolution of neural representations in fMRI data, computed on single-trial BOLD time-courses, leveraging both spatial and temporal components of the fMRI signal. We implemented an expanding sliding window approach that allows identifying the time-window of an effect.

**Results:** We demonstrate that tMVPA can successfully detect condition-specific multivariate modulations over time, in the absence of mean BOLD amplitude differences. Using Monte-Carlo simulations and synthetic data, we quantified family-wise error rate (FWER) and statistical power. Both at the group and single-subject levels, FWER was either at or significantly below 5%. We reached the desired power with 18 subjects and 12 trials for the group level, and with 14 trials in the single-subject scenario.

**Comparison with existing methods:** We compare the tMVPA statistical evaluation to that of a linear support vector machine (SVM). SVM outperformed tMVPA with large N and trial numbers. Conversely, tMVPA, leveraging on single trials analyses, outperformed SVM in low N and trials and in a single-subject scenario.

**Conclusion:** Recent evidence suggesting that the BOLD signal carries finer-grained temporal information than previously thought, advocates the need for analytical tools, such as tMVPA, tailored to investigate BOLD temporal dynamics. The comparable performance between tMVPA and SVM, a powerful and reliable tool for fMRI, supports the validity of our technique.

### 1. Introduction

Over the past quarter century, functional Magnetic Resonance Imaging (fMRI) has become one of the most powerful non-invasive tools for investigating human neural processing. By exploiting the coupling between oxygenated blood flow and neuronal firing (Goense and Logothetis, 2008; Logothetis et al., 2001; Ogawa et al., 1993), fMRI infers cortical activity by measuring changes in the Blood Oxygen Level Dependent (BOLD) signal (Goense and Logothetis, 2008; Logothetis et al., 2001; Ogawa et al., 1993). The sluggish nature of the hemodynamic based BOLD signal (requiring several seconds to peak following stimulus presentation – Boynton et al., 1996; Ogawa et al., 1993), paired with the high spatial precision of fMRI recordings, has resulted

in a focus on BOLD spatial information in most applications, neglecting any temporal dynamics. More recently, developments in fMRI pulse sequences that allow significant increases in temporal resolution (Feinberg et al., 2010; Moeller et al., 2010) have been thus far primarily exploited to improve statistical power in fMRI analysis. These developments however, further offer the possibility of resolving BOLD temporal dynamics that were previously elusive.

While focus has been primarily on the spatial domain of the BOLD signal, this is not to say that the fMRI temporal domain has been entirely ignored. For example, several attempts have been made to target local stimulus-distinct characteristics of the BOLD time series. Specifically, these investigations have sought to understand stimulus-specific temporal effects in the context of decision making (Mcguire and

\* Corresponding author at: Centre for Magnetic Resonance Research, Department of Radiology, University of Minnesota, 2021 6th Street SE, 55455 Minneapolis, MN, United States.

E-mail address: [lvizioli@umn.edu](mailto:lvizioli@umn.edu) (L. Vizioli).

<sup>1</sup> Both authors equally contributed to the formulation of this manuscript.

<https://doi.org/10.1016/j.jneumeth.2018.06.029>

Received 10 April 2018; Received in revised form 29 June 2018; Accepted 29 June 2018

Available online 30 June 2018

0165-0270/ © 2018 Elsevier B.V. All rights reserved.

Kable, 2015), auditory (Baumann et al., 2010), and semantic and visual processing (Avossa et al., 2003; Bailey et al., 2013; Formisano et al., 2002; Gentile et al., 2017; Siero et al., 2011; Vu et al., 2016). In conjunction, animal studies have sought to understand the precise relationship between the BOLD temporal dynamics and the neural activity elicited from such domains (Silva and Koretsky, 2002; Yen et al., 2018). Additionally, it is worth noting that a variety of both high complexity and real world stimuli operate at the temporal resolution available to fMRI. For example, in the visual domain, a number of visual illusions are characterized by their slowly transforming, bi-stable nature (Ernst and Bu, 2004; Schrater et al., 2004). Furthermore, biological motion (Johansson, 1973; Maier et al., 2008; Troje, 2002) and other motion-based complex stimuli (Ball and Sekuler, 1982; Shadlen and Newsome, 1998) are typically presented over large temporal windows. BOLD latency measurements have likewise been shown to be relevant in the auditory and multisensory domain, where, for example, phonemic boundaries shift across temporal gradients when presented in isolation (Lee et al., 2012) or within specific visual contexts (Gribble, 1996). Moreover, analyses of neural responses to any long duration stimuli, such as film or real-world dynamic scenes, necessitate a technique that directly measures the temporal evolution of the BOLD signal.

Importantly, a number of studies have more recently suggested that fMRI may carry neuronal information at a much faster temporal scale than previously (Lewis et al., 2016; Siero et al., 2011; Vu et al., 2016). Siero et al. (2011), for example, indicated that neurovascular coupling takes place on a shorter timescale than had been previously reported in the human brain. Moreover, Lewis et al. (2016) have suggested that, due to recent advances in MR hardware and software as well as analytical strategies, fMRI can measure neural oscillations up to 1 Hz. Additionally, Vu et al. (2016) successfully demonstrated that with the use of multivoxel pattern analysis (MVPA), it is possible to extract word timing information with fast TRs (i.e. 500 ms). Along the same lines, in a visual illusion experiment, Edwards et al. (2017) showed that as little as 32 ms difference in stimulus presentation is reliably detected in the BOLD time-course.

These observations highlight the growing interest in the temporal dynamics of the BOLD signal. However, to fully exploit the potential neuro-temporal information carried by the BOLD time-course, MR hardware and software (e.g. pulse sequences) developments have to be paired with suitable analytical tools that maximize the sensitivity to BOLD temporal information. Thus far, the majority of temporal analyses have only examined mean BOLD amplitude temporal differences between stimuli or stimulus conditions (i.e., latency differences on average amplitude). While such data is useful for understanding the propagation of neural activation throughout the brain as a function of time, it fails to capture the representational content as conveyed by multivariate patterns as well as how these representations transform over time. Multivariate approaches to analyzing fMRI data offer a different, albeit complementary outlook on the neural information carried by the BOLD signal (Kriegeskorte and Bandettini, 2007). It has been suggested that multivoxel pattern analysis, or MVPA (Haxby et al., 2001; Kamitani and Tong, 2005), has the ability to optimally probe neuronal information existing in voxel populations with conventional fMRI methods (Carlson et al., 1999; Cox and Savoy, 2003; Haxby et al., 2005; Kriegeskorte and Bandettini, 2007; Strother et al., 2002). Even at 3 T, where voxels traditionally measure 2–3 mm isotropic resolutions, MVPA can successfully extract neural information – such as orientation preference (Kamitani and Tong, 2005) – which exists at a much finer spatial scale than the resolution of single voxels. These approaches are believed to increase the sensitivity to such fine-grained information present in lower resolution images by exploiting the micro-feature-selective biases of single voxels that stem from the variability of the distribution of cortical columns or their vascular architecture (Op de Beeck, 2010; Freeman et al., 2011; Kamitani and Tong, 2005; Mannion et al., 2009, 2015; Sasaki et al., 2006).

Inspired by the demonstrated fine sensitivity of MVPA to finer scale

spatial information, here we apply multivariate analysis to BOLD time-courses in order to maximize sensitivity to neuro-temporal information. Capitalizing on the growing interest surrounding the temporal domain of fMRI, we propose a method that captures the temporal characteristics of the BOLD signal at the multi-voxel pattern level. The method, first introduced in Ramon et al. (2015), consists of probing single trial events to investigate how the associated representational pattern of activity (Kriegeskorte et al., 2008; Kriegeskorte and Kievit, 2013) for a given stimulus evolves over time. This enables the creation of Single Trial Representational Dissimilarity Matrices (stRDMs), which allows assessing the temporal evolution of the (dis)similarity of these activity patterns.

As previously shown on real data (Ramon et al., 2015), here we demonstrate on synthetically generated data that our approach can detect multivariate differences over time in the *absence of mean BOLD* amplitude modulations across conditions. As such, our temporal multivoxel pattern analysis (tMVPA) offers a different albeit potentially complementary approach to examining BOLD temporal dynamics. We further present a sliding window statistical analysis of these stRDMs that allows quantifying the precise temporal window displaying the effect of interest. We estimate the power and family wise error rate (FWER) of the technique using Monte Carlo simulations. Moreover, we benchmarked tMVPA performance against one of the most widely used multivariate tools in fMRI: linear support vector machine (SVM).

## 2. Methods

### 2.1. Procedure and MRI acquisition

Note that the acquired data were used as a starting point to generate synthetic data with realistic signal properties. Thus, within the context of this paper, the original purpose and the hypothesis of the experiment are irrelevant.

#### 2.1.1. Participants

20 healthy right-handed subjects (age range: 18–31) participated in the study. Of these, 10 were Western Caucasian (WC; 5 females; mean age, 24) and 10 were East Asian (EA; 4 females; mean age, 22). Three participants (1 WC 2 EA) were excluded from the analysis due to excessive motion during scanning (details below). All subjects had normal, or corrected vision and provided written informed consent. The ethical committee of College of Medical, Veterinary and Life Sciences at the University of Glasgow approved the experiments.

#### 2.1.2. Stimuli and procedure

The experimental procedure consisted of a standard block design face localizer and a simple slow event-related face paradigm. All visual stimuli used for the face localizer consisted of front-view gray scale photographs depicting 20 different faces (5 identities  $\times$  2 genders  $\times$  2 races, taken from the JACFEE database (Matsumoto and Ekman, 1988)), houses (Husk et al., 2007) and textures of noise, respectively. Noise texture stimuli were created by combining the mean amplitude spectrum across faces and houses with random phase spectra sampled from a Gaussian distribution, thereby lending them to contain the same amplitude spectrum as the face and house stimuli. For the main slow event-related experiment, a different set of images used in previous studies (Michel et al., 2006) was utilized which also consisted of 20 front-view gray scale photographs of WC and EA (again 5 identities  $\times$  2 genders  $\times$  2 races). All images subtended approximately  $3.75 \times 4.25^\circ$  of visual angle. Face stimuli were cropped to remove external features; none had particularly distinctive features and male faces were clean-shaven. The stimuli were centered in a  $52 \times 52$  cm background of average luminance (25.4 cd/m<sup>2</sup>, 23.5–30.1). All images were equated in terms of luminance, contrast and spatial frequency content by taking the average of the amplitude spectra of all stimuli and combining that average spectrum with the original phase spectra to reconstruct each

individual stimulus. The root mean square contrast (i.e. the standard deviation of the pixel intensities) was also kept constant across stimuli. Stimuli were projected from the back of the scanner on a round screen situated in the scanner tunnel and occupying the whole width of the tunnel (i.e. 60 cm of diameter). Participants viewed the images through a mirror placed on the head coil.

All participants completed two runs of the block design face localizer fMRI experiment to define the areas responding preferentially to faces (~12 min/run), and three runs of the main event-related design experiment aimed at measuring the neural activity elicited by individual SR and OR identities (~16 min/run).

### 2.1.3. Face localizer

Face localizer runs involved presentation of blocks of WC or EA faces, houses and noise textures. Each run began with presentation of black fixation cross displayed on grey background for 20 s and consisted of 24 randomly presented blocks of images. Each block (6 blocks/category; separated by a 12 s fixation) involved presentation of 10 different stimuli randomly presented for 800 ms, separated by a 400 ms ISI. To minimize attentional confounds on the BOLD signal related to the race of the stimuli, we implemented an orthogonal task. Participants were instructed to respond to red or green stimuli which (10% of the images, i.e. one red or green stimulus per block), by pressing a button on a response pad held in their right hand.

### 2.1.4. Event-related experiment

Each run of the event-related face experiment began and ended with 20 s fixation and consisted of 80 events (10 identities per race x 2 races x 4 repetitions per identity). Face stimuli were displayed for 850 ms followed by a 11.15 s fixation cross; participants were instructed to maintain fixation on a central fixation cross throughout each 12 s event. As for the face localizer scans, an orthogonal task was employed with participants responding to a change in the color of the fixation cross (red or green, for 200–1200 ms at a random time within an event, before reverting to its original color) by pressing a button.

### 2.1.5. MRI acquisition protocol

All MRI data were collected with a 3-T Siemens Tim Trio System with a 32-channel head coil and integrated parallel imaging techniques (IPAT factor: 2). Functional MRI volumes were collected using an echoplanar acquisition sequence [localizer runs: repetition time (TR), 2000 ms; echo time (TE), 30 ms; field of view (FOV), 210 x 210 mm; flip angle (FA), 77°; 36 axial slices; spatial resolution, 3 mm isotropic voxels; event-related runs: TR, 1000 ms; TE, 30 ms; FOV, 210 x 210 mm; FA, 62°; 16–18 axial slices; spatial resolution, 3 x 3 x 4 mm voxels]. Slices were positioned to maximize coverage of occipito-temporal regions. T1-weighted anatomical images were obtained using an MPRAGE sequence (192 slices; TR, 1900 ms; FOV, 256 x 256 mm; flip angle, 9°; TE, 2.52 ms; spatial resolution, 1 mm isotropic voxels). For participants who were re-scanned due to movement artifacts, separate anatomical scans were recorded for each scanning session to facilitate realignment of the functional data.

### 2.1.6. MRI data preprocessing

fMRI data were preprocessed in native space using BrainVoyager QX version 2.1 (Brain Innovation). Functional images were slice-scan time corrected, three-dimensional motion corrected with reference to the functional volume taken just before the anatomical scan, high-pass filtered using a Fourier basis set of three cycles per run (including linear trend). Images were co-registered with the anatomical set and spatially normalized into Talairach space (Talairach and Tournoux, 1988); images from localizer runs were spatially smoothed with a full-width at half-maximum of 4 mm.

### 2.1.7. Functional ROI definition

Five functional ROIs were identified from the localizer runs.

Individual participants bilateral FFA, bilateral OFA, and right AIT were identified by performing F-tests on all the voxels in the brain and determining the peak voxel of the activation clusters identified by the contrast (WC + AC) faces > (Houses + Noise) located in the bilateral fusiform and inferior occipital gyrus, respectively. To control for type I errors, False positive Discovery Rate (FDR) was implemented as a multiple comparison correction. The significance threshold was set to  $q < .05$  for all ROIs and participants. The corresponding masks for these ROIs were exported into MATLAB (MathWorks) for subsequent analyses. Across all participants from both groups (WC and EA), we identified 86 ROIs in total. While bilateral FFA and right OFA were identified in all participants, a few subjects did not have a clear definition of left OFA and right AIT. The average number of voxel across all ROIs was 47.9 (std: 16.7).

### 2.1.8. BOLD percent signal change and epochs definition

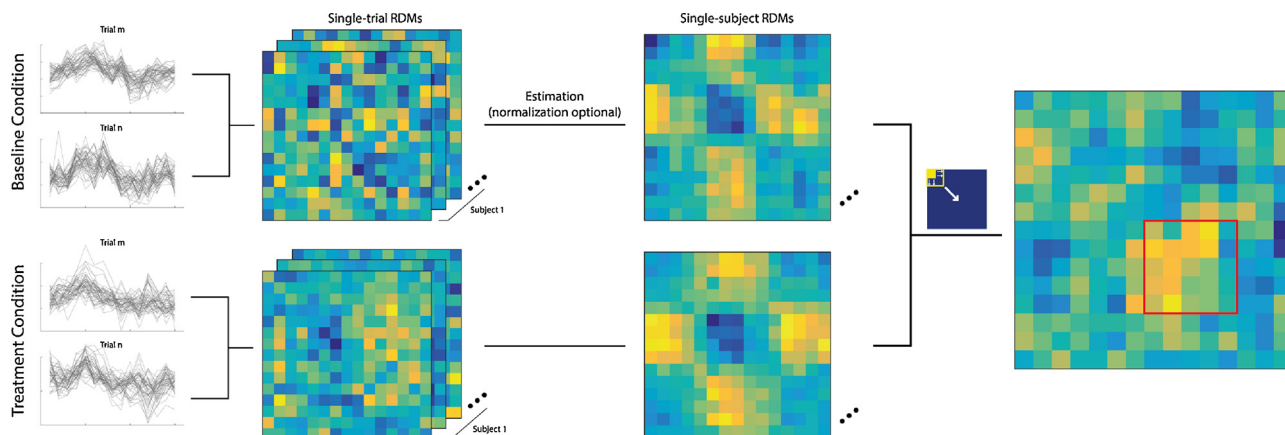
For each voxel, we computed BOLD percent signal change by dividing the raw BOLD time course by its mean. We then defined the epochs of interest as those portions of the whole BOLD time series ranging from 1 TR prior to 14 TRs after stimulus onset. For each single trial we extracted these 15-TR long time-courses from all the voxels within each ROI of every subject. These BOLD percent signal change epochs were saved as a matrix that we used to generate synthetic data using Monte Carlo simulations (details below).

## 2.2. Temporal multivariate pattern analysis (tMVPA)

In this paper, we developed a novel multivariate temporal analysis for the BOLD time-course, inspired by representational similarity analysis (Kriegeskorte and Kievit, 2013). This approach assesses the temporal evolution of the degree of dissimilarity of neural representations - defined as the pattern of BOLD response across all voxels - elicited by different time points (Ramon et al., 2015). It involves computing Single Trial Representational Dissimilarity matrices (stRDMs) within a selected ROI between two conditions (e.g., *baseline* and *treatment* condition). We compute stRDMs on the BOLD percent signal change independently per subject and condition as follows: for each condition, we iteratively correlated (Pearson  $r$ ) the values of all the voxels at one time point with all the remaining ones amongst the epochs of two different trials (e.g. the time course elicited by trial 1 and that elicited by trial 2) and calculated the correlation distance (i.e.  $1-r$ ; see Fig. 1). This procedure was repeated across all possible trial pair combinations. The resulting matrices were fisher-z transformed to render the skewed Pearson- $r$  distribution approximately normal. We then averaged (10% trimmed mean) the single trial correlational distance matrices to obtain the single subject stRDM.

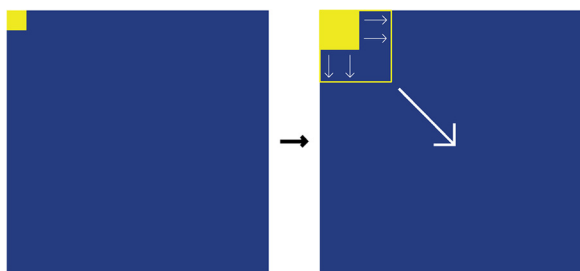
While all subsequent statistical analyses were performed on the fisher-z transformed values, for visualization purposes (Fig. 1) and to render the values within the stRDMs interpretable, we performed the inverse of the fisher-z normalization on the final averaged stRDM.

To test for statistically significant differences between the stRDMs from different conditions (i.e., *baseline* and *treatment* condition), we implemented an expanding sliding window approach. We started by computing a simple subtraction between the stRDMs of the 2 conditions of interest. We then centered a 2 x 2 pixel window (Fig. 2) on the first point of the diagonal of the matrix. We then computed the 10% trimmed mean across the values within the window. We divided this mean by the standard error of the values within the window. Given that the standard error is a function of the variance weighted by the number of data points, this procedure was implemented to partially account for the relative difference in terms of data points and variance across windows of different sizes. We then performed (1- $\alpha$ ) bootstrap confidence interval (CIs) analyses by sampling subjects with replacement 500 times. Importantly, we adjusted the threshold ( $\alpha$  above) for determining high and low CIs as a function of the total number of windows to account for multiple comparison problems (i.e. Bonferroni



**Fig. 1.** The temporal multivariate pattern analysis (tMVPA) procedure. Cool colors indicate higher similarity between neural representations elicited by any 2 given time points. Warm colors indicate higher dissimilarity or distinctiveness amongst neural representations. Each row and column represents a single TR.

correction). The analysis was repeated on increasingly larger windows that expanded by 1 pixel in each direction (when applicable), centered on each point of the diagonal (Fig. 2). Differences between conditions were inferred when the btCIs did not include zero. This expanding sliding window approach allows investigating whether potential differences across stRDMs encompass a few time points or whether these are sustained over a larger time window.



**Fig. 2.** Expanding Sliding Window approach. The panel on the left depicts the starting window size and location, while the panel on the right represents this same window “sliding” (as indicated by the large white arrow) and “expanding” (as indicated by the thin pairs of white arrows).

### 2.3. Synthetic data generation and validation

The following sections describe the procedure we implemented for the synthetic data generation process and the approach we adopted to estimate the power and Family-wise error rate (FWER) of our proposed multivariate temporal analysis. In brief, we employed Monte Carlo (MC) simulation on synthetic data to estimate the FWER and the statistical power of our proposed method, explicitly manipulating a number of parameters (see the *Manipulated parameters paragraph*). In order to reproduce realistic fMRI noise and signal properties, we generated synthetic data starting from the BOLD signal recorded during the event-related experiment. We created a complete dataset comprised of 2 conditions (i.e. *Baseline* and *Treatment*). Importantly, we generated Baseline and Treatment conditions under 2 distinct scenarios: 1) under H0 (i.e. no multivariate differences between conditions), thus being in the ideal context to measure our approach’s FWER, as any statistical difference detected by our approach would be a false positive; and 2) under H1 (i.e. artificially introducing multivariate pattern differences between conditions – see *synthetic multivariate effect*) to test our approach’s power (see below for more details).

#### 2.3.1. Synthetic data generation

Starting from the single trial BOLD time course matrix (see the *BOLD percent signal change and epochs definition paragraph*), we

extracted single trial epochs from one of the 20 conditions for one participant across one run and using just a single ROI. We saved the extracted BOLD values in a 3D *Raw\_singletrials\_BOLD* matrix with dimensions [number of trials \* number of voxels \* number of time points]. From the *Raw\_singletrials\_BOLD* matrix we calculated the mean and the variance across voxels, and then saved these 2 metrics in 1D vectors of size [number of time points]. We refer to these vectors, representing respectively the average HRF for a given ROI and the voxel-wise variance within that same ROI, as *mu\_BOLD(time point)* and *var\_BOLD(time point)*. We then calculated the residual between the single trials epochs and their mean (across trials) for each voxel and time point, and then saved these values in a [number of trials \* number of voxels \* number of time points], a 3D matrix that we refer to as *sigma\_BOLD(trial, voxel, time point)*.

We repeated the procedure described above for all conditions, runs, ROIs, and subjects. The resulting *mu\_BOLD*, *var\_BOLD*, and *sigma\_BOLD* were flattened and saved in 2 dimensional matrices: **E**, **V**, and **S**. Note that the matrices **E**, **V**, and **S** have an equal numbers of columns, corresponding to the number of time points per epoch of interest (i.e. 15), but a different number of rows. For the matrices **E** and **V**, containing, respectively, the mean time courses across voxels and the variance across voxels, the number of rows was equal to [number of subjects \* number of runs \* number of conditions \* number of ROIs]; while the number of rows for matrix **S**, containing the single-trial residual for each voxel, was equal to [number of subjects \* number of runs \* number of conditions \* number of ROIs \* number of trials \* number of voxels per ROI].

The raw BOLD signal was thus fully represented in matrices **E**, **V**, and **S**. To generate synthetic data for one subject we randomly sampled one row vector from **E** and **V** and generated a 2D [number of voxel \* number of time points] matrix, representing the mean (across trials) time course for all voxels within a given ROI. We then injected the trials’ variation from their mean by randomly sampling from **S** (see below for details).

In order to generate the Baseline and Treatment conditions, we implemented very similar, albeit slightly different procedures. The first step of the data generation process (step 0) was the same regardless of the generation goal. For each MC simulation, we began by randomly selecting a row vector **e** from matrix **E**, representing the group average time course for a hypothetical ROI.

For the Baseline condition, independently per subject we generated a number of voxels (*nv*) \* number of time points (*ntp*) \* number of trials (*ntrial*) matrix **MB**, following the 9 step algorithm below:

- Step 1, we randomly selected a row vector **v** from matrix **V**, and *nv*\**ntrial* rows vectors from matrix **S** to get **sv**.



To have full control of the simulation study, we kept the variance across time points within a single voxel and a single trial constant by setting  $v_1 = v_2 = \dots = v_{ntp} = \text{mean}(\mathbf{v})$  and  $sv_{1,i} = sv_{2,i} = \dots = sv_{ntp,i} = \text{mean}(sv)$  for  $i \sim [1, nv]$ .

- Step 2, we repeated  $nv$  copies of array  $e$  and transformed them into a  $nv \times ntp$  matrix  $\mathbf{ev}$ .
- Step 3, we repeated  $nv$  copies of array  $\mathbf{v}$  and transformed them into a  $nv \times ntp$  matrix  $\mathbf{vv}$ .
- Step 4, we generated a  $nv \times ntp$  matrix  $\mathbf{dv1}$  to represent the variance across voxel. Each element in  $\mathbf{dv1}$  was generated following one of 3 distributions: either *Normal* ( $\mu = 0, sd = 1$ ), *Uniform* ( $lower = -, upper =$ ), or *Exponential* ( $\lambda = 1$ ) - 1. These three distributions all have mean equal to 0 and variance equal to 1.
- Step 5, the mean BOLD time course for each voxel  $\mathbf{Mp}$  was generated following the equation:

$$\mathbf{Mp} = \mathbf{ev} + \mathbf{dv1} \cdot \sqrt{\mathbf{vv}}$$

Where "." indicates the element-wise multiplication. By doing this,  $\mathbf{Mp}$  satisfies  $\text{mean}(\mathbf{Mp}) = \mathbf{e}$  and  $\text{var}(\mathbf{Mp}) = \mathbf{v}$ .  $\mathbf{Mp}$  is an  $nv \times ntp$  matrix representing the single voxel BOLD time course.

- Step 6, we repeated  $ntrial$  copies of matrix  $\mathbf{Mp}$  and transformed them into a  $nv \times ntp \times ntrial$  matrix  $\mathbf{MP}$ .
- Step 7, we reshaped the residual matrix  $\mathbf{sv}$  into an  $nv \times ntp \times ntrial$  matrix and computed the variance across trials. The resulting  $nv \times ntp$  matrix was then repeated and reshaped into an  $nv \times ntp \times ntrial$  matrix  $\mathbf{svt}$  representing the single trials residuals for each voxel and time-point.
- Step 8, we generated an  $nv \times ntp \times ntrial$  matrix  $\mathbf{dv2}$ . Similar to  $\mathbf{dv1}$ , each element in  $\mathbf{dv2}$  followed one of 3 distributions: either *Normal* ( $\mu = 0, sd = 1$ ), *Uniform* ( $lower = -, upper = \sqrt{3}$ ), or *Exponential* ( $\lambda = 1$ ) - 1.  $\mathbf{dv2}$  represents the noise at the single trial level for each voxel.
- Step 9, finally, we computed the single trials BOLD time course matrix  $\mathbf{MB}$  following the equation:

$$\mathbf{MB} = \mathbf{MP} + \mathbf{dv2} \cdot \sqrt{\mathbf{svt}}$$

Notice that the mean and variance across trials for  $\mathbf{MB}$  satisfies  $\text{mean}(\mathbf{MB}) = \mathbf{Mp}$  and  $\text{var}(\mathbf{MB}) = \mathbf{svt}$ .

These 9 steps were repeated for all subjects.

Similar to the baseline conditions, we generated an  $nv \times ntp \times ntrial$   $\mathbf{MT}$  Treatment condition matrix for each subject following the same 9 steps.

When no effect was introduced in the Treatment condition (i.e. FWER estimation, see below), the  $\mathbf{MT}$  matrix creation began directly at step 7 (through to 9), starting from the same  $\mathbf{Mp}$  and  $\mathbf{sv}$  generated for the Baseline condition using steps 1 to 6. Thus, the  $\mathbf{MT}$  mean and variance across trials satisfies  $\text{mean}(\mathbf{MT}) = \mathbf{Mp}$  and  $\text{var}(\mathbf{MT}) = \mathbf{svt}$ .

### 2.3.2. Synthetic multivariate effect

Our procedure to introduce multivariate differences between the baseline and treatment conditions consisted of rendering the voxel response for some selected time points in the treatment condition highly correlated across trials. To achieve this, we first repeated steps 1–9 to generate matrix  $\mathbf{Mp}'$ , containing the treatment condition mean BOLD time course across all trials for all voxels within a given ROI;  $\mathbf{MP}'$ , containing the single trials BOLD time-course for all voxels within any given ROI;  $\mathbf{svt}'$ , containing the residuals between the single trials and average across trials for each voxel, time-point, and trial; and  $\mathbf{MT}'$ , containing the single trials' BOLD time courses for all voxels within a given ROI. We therefore modulated  $k$  consecutive time points in matrix  $\mathbf{MT}'$  to introduce correlation in the synthetic signal by rotating the data matrix  $\mathbf{MT}$  to reduce the multivariate distance across trials.<sup>2</sup>

Independently for each of the  $k$  time points, we first repeated step 8 to generate a new independent and identically distributed (i.i.d.) noise matrix  $\mathbf{dv2}'$ . We then computed the BOLD time course for the treatment condition  $\mathbf{MT}$  following the equation:

$$\mathbf{MT}'[:, k, :] = \mathbf{MP}'[:, k, :] + \text{diag}(\sqrt{\mathbf{svt}'[:, k, 1]}) * \mathbf{L} * \mathbf{dv2}'[:, k, :],$$

where  $\mathbf{L}$  is the Cholesky factor of a correlation matrix randomly sampled from a LKJ correlation distribution (Lewandowski et al., 2009). Therefore, the variance across voxels for  $k$ -th time points of some selected voxels was identical:

$$\text{var}(\mathbf{Mp}'_{t:t+k}) = \text{var}(\mathbf{Mp}_{t:t+k}) = \mathbf{v}.$$

Notice that the mean BOLD amplitude in  $\mathbf{Mp}'$  was kept constant:  $\text{mean}(\mathbf{Mp}') = \mathbf{e}$ . Moreover, the mean and variance across trials for  $\mathbf{MT}'$  also satisfies  $\text{mean}(\mathbf{MT}') = \mathbf{Mp}'$  and  $\text{var}(\mathbf{MT}') = \mathbf{svt}'$ .

The resulting data matrix  $\mathbf{MB}$  and  $\mathbf{MT}'$  represented the full synthetic dataset for one subject. We repeated the above 9 steps to generate  $k$  (number of subject)  $\mathbf{MB}$  and  $\mathbf{MT}'$  matrices. We therefore implemented our TMPVA analysis to test for multivariate differences between the treatment and baseline conditions. We repeated this MC simulation 1000 times for each combination of parameters (details below).

### 2.3.3. Manipulated parameters

In an attempt to maximally parameterize our validation procedure while keeping within the boundaries of reasonable computational demands, we manipulated the following 4 parameters: 1) number of trials per condition, 2) number of subjects per group, 3) number of time points at which the effect was introduced, and 4) the percentage of subjects (or trials for the single subject validation procedure) in which the effect was introduced (i.e. the target power).

- 1 the number of trials varied across 4 different levels: 4, 8, 12, and 16.
- 2 for the number of subjects, we tested 4 sample sizes: 6, 10, 14, 18 participants.
- 3 while the multivariate effect always began at TR 5, the number of time points at which the effect was introduced varied across 4 different levels: 2, 3, 4, 5.
- 4 the percentage of sample showing effect (i.e. power), varied across three different levels: 50%, 65%, and 80%.

Additionally, the number of voxels (range [30, 60]) per simulated subject was randomized across all MC simulations. We thus ran independent MC simulations for all possible combinations of the different parameter levels. This parameterization of the MC simulation was implemented to evaluate the reliability and sensitivity of our method in different experimental contexts and assess the relative impact of different parameters on tMVPA statistical performance. Note that we introduced a multivariate effect for our power analysis at time point 7 (up to time point 11, depending on the number of manipulated time points). For the estimation of FWER, only number of trials and number of subjects were relevant parameters. For each unique parameter combination, we computed 95% bootstrap CI based on 500 bootstraps, and repeated this procedure 1000 times.

Importantly, we validated our tMVPA approach within two different settings: *group analysis* and *single subject analysis*. In the group analysis setting, to manipulate the target power we varied the percentage of subjects in which we introduced correlation across voxels (i.e. the synthetic multivariate effect). In the single subject validation setting, the target power was instead manipulated by varying the percentage of

(footnote continued)

between 2 points in a multidimensional space. In the same vein, we can think of increase in correlation (and therefore decrease in correlational distance) between these 2 points as a rotation of axis of the the multidimensional space for point 1.

<sup>2</sup>Note that correlational distance  $1-r$  can be conceptualized as distance

trials in which the multivariate pattern was introduced (i.e. 50%, 65% or 80% of the trials).

Furthermore, we performed the same validation analysis using linear support vector machines (SVM). To this end, we divided our synthetic data into 4 runs. We performed a classic general linear model (GLM) analysis (least squared minimization) on the synthetic data timecourses. We estimated one beta weight per condition, voxel and run, summarizing BOLD activation. We computed the design matrix for the GLM by convolving a double gamma function with a “boxcar” function (representing onset and offset of the image stimuli). We thus trained SVM on 3 runs and tested its accuracy on the 4<sup>th</sup> independent run. We repeated this procedure for all possible combination of runs. Like tMVPA, we evaluated SVM decoding accuracy across all parameters (see the “manipulated parameters” paragraph). Decoding analysis was performed using linear support vector machine (SVM) algorithms as implemented by the LIBSVM toolbox (Lin, 2011), with default parameters (notably  $C = 1$ ). Note that the activity of each voxel for the training was normalized within a range of -1 to 1 before input to the SVM.

#### 2.3.4. FWER estimation

To estimate the FWER, we performed tMVPA and SVM analyses (as described above) to test for multivariate differences between baseline and treatment conditions, prior to introducing correlation across voxels at selected time points. We thus counted the number of significant events detected by both approaches. We repeated this procedure 1000 times. Since baseline and treatment conditions were created under  $H_0$  (i.e. no differences between them), significant differences detected by our approach were considered to be false positives (i.e. type I error). The FWER was thus computed as the total number of significant time windows divided by 1000 (i.e. the total number of MC simulation).

#### 2.3.5. Statistical power estimation

For statistical power estimation we, instead, generated 1000 *treatment* conditions following a procedure similar to the generation of the *baseline* condition (i.e. steps 1–9 as described earlier). We additionally introduced multivariate differences between conditions (see *Synthetic multivariate effect*) in a number of subjects by manipulating the pattern of voxels within a given ROI over some selected time points (see *Manipulated parameters for more details*). Importantly, no mean BOLD amplitude differences (i.e. no differences between the time courses averaged across voxels - see Figs. 3 and 4) between the two conditions existed over these time points. The target power of the tMVPA and SVM approaches was represented by the percentage of subjects for whom we introduced multivariate differences between conditions. For example, if we introduced correlation across voxels in 80% of the subjects, we expected the tMVPA and SVM to report significant differences 80% of

the time across all simulations where the effect was introduced. Statistical power was thus computed as the total number of significant time windows detected divided by the total number of MC simulations.

### 3. Results

95% bootstrap confidence intervals (btCIs) computed across our MC simulations showed that manipulating the number of time points at which we introduced the synthetic multivariate effect did not significantly ( $p > .05$ ) impact FWER and power estimations (see supplementary section). Additionally, we observed that the distribution from which we sampled the synthetic noise did not significantly ( $p > .05$ ) modulate FWER and power estimations (see supplementary section). We therefore only report the results for synthetic data with a multivariate effect over 3 time-points, generated by sampling noise from a normal distribution. Figures and results for the remaining levels of these 2 parameters as well as detailed tables reporting mean and bootstrap CIs can be found in the supplementary section.

In the following paragraph we report the mean across all MC simulations and standard deviation (std) of the peak amplitude of the BOLD % signal change time course. We further report the mean std across voxels, trials, and time course. In the MC simulations for the group study, the mean peak amplitude (across subjects and MCs) of the generated synthetic BOLD % signal change was 1.222 (std = .531), while a mean std across time 0.353 (std = .137). Moreover, the average std across voxels was 2.815 (std = 2.643) and the average std across trials 1.343 (std = .348). As for the MC simulation for the single subject study, the generated synthetic data set had a mean (across MCs) peak amplitude of 1.247 (std = .533), with a mean std across time 0.357 (std = .106). The mean std across voxels was 2.996 (std = 2.409), and the mean std across trials was 1.364 (std = .362).

Fig. 4 (panels a through e) shows the BOLD time course of our synthetic data for the 18 subjects and 16 trials scenario. Error bars represent the 95% bootstrap confidence intervals (btCIs). We infer robust statistical significance ( $p < .05$ ) when the error bars do not overlap. Our analyses revealed no significant mean BOLD amplitude differences across the whole time course between the *baseline* (red line) and the *treatment* (blue line) conditions for all the parameter manipulations (see manipulated parameters). Importantly, this absence of mean BOLD amplitude differences persisted even after we synthetically introduced multivariate effects at selected time-points. Our tMVPA approach, thus, crucially revealed robust *genuine* multivariate differences across conditions that are not evident in mean BOLD amplitude differences. Note that the introduced multivariate effect is visible by computing the sRDM, as shown in Fig. 4f for the *baseline* condition and Fig. 4h–j for the *effect* condition.

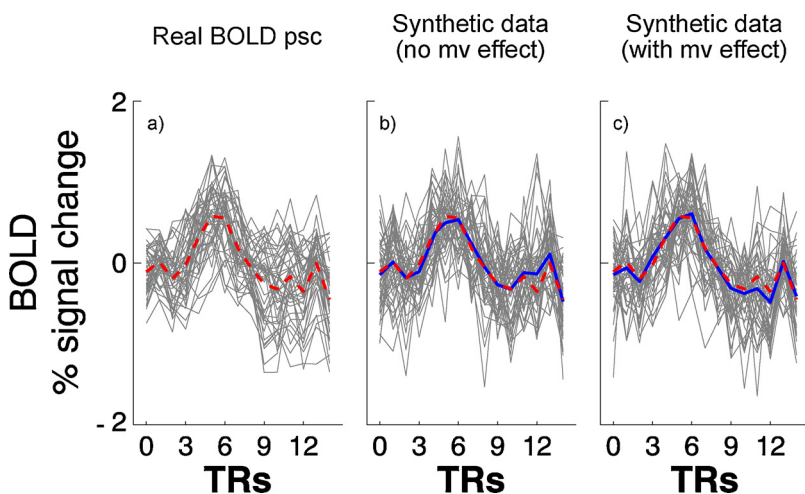
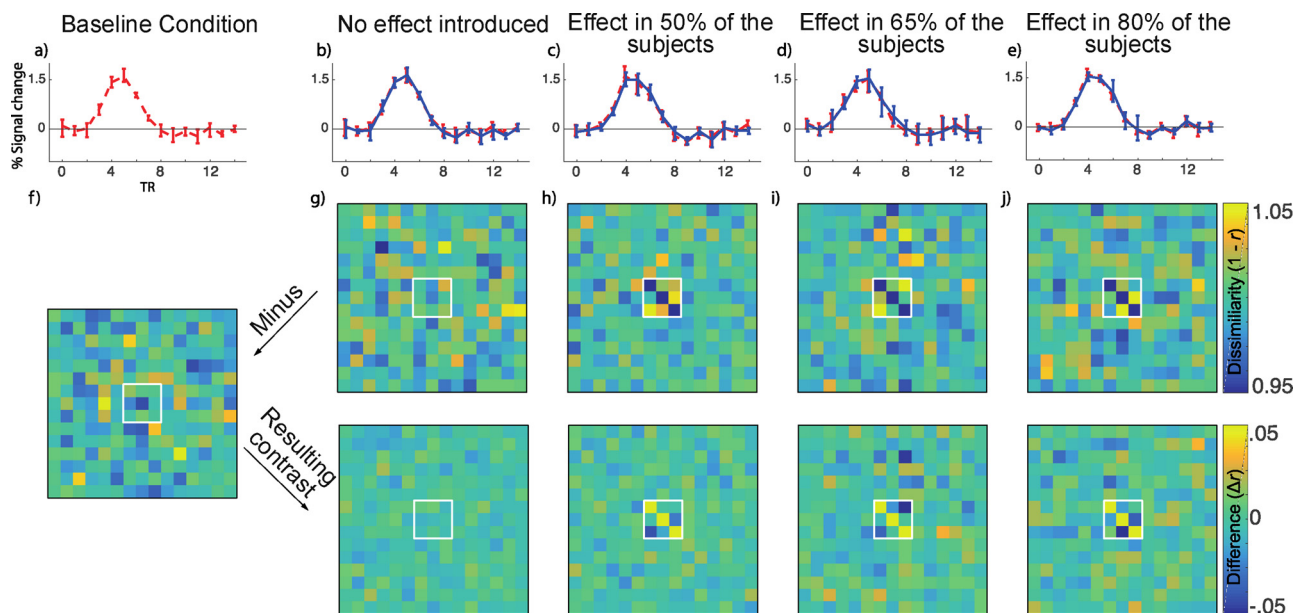


Fig. 3. Panel a) portrays an example of real BOLD percentage signal change (psc) time course for all voxels in a given ROI for a single subject. The thin grey line plots show the BOLD time course for each voxel, while red dashed line shows the average BOLD time course. Panels b) and c) depict the generated synthetic BOLD time course created using the same mean and variance of the real BOLD time course. Panel b) shows an example of the synthetic baseline condition - i.e. no multivariate (mv) effect; and Panel c) shows an example of a synthetic treatment condition where we introduced a mv effect (see the Synthetic multivariate effect paragraph) over time points 5–7. Thin grey line plots show single voxels, the red dashed line shows the average time course of the real signal, the solid blue line shows the average time course of the synthetic data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** Synthetic data for the 18 subjects and 14 trials group a-e) Average time course across voxel participant within a ROI. The dashed red line shows baseline condition (a) and the solid blue line shows Treatment condition. Error-bars shows 95% bootstrapped confidence interval across subjects for each time point. f) stRDM of the baseline condition. g) stRDM of the treatment condition when no effect is introduced (to estimate FWER). h-j) stRDM of the treatment condition when different strengths of the multivariate effect is introduced over time-points.5–7. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 3.1. Family-wise error rate (FWER) under $H_0$

For both the group and single subject scenarios, to estimate the FWER we computed the frequency of significant outputs detected by our approach across MC settings, before introducing the multivariate effect. As explained earlier, prior to introducing correlation across voxels over a number of selected time points, we generated the synthetic baseline and treatment data under  $H_0$  (i.e. no differences between conditions). We were, therefore, in the ideal context to estimate FWER, as statistically significant differences between conditions were mere type I errors.

#### 3.1.1. Group-level analysis

**3.1.1.1. tMVPA.** 95% bootstrap confidence intervals (btCIs) show that FWERs were significantly below .05 in all MC simulation with sample size  $> 6$  (Fig. 5). For  $N = 6$  the mean of the estimated FWER was, instead, consistently above .05 (mean FWER: .058), regardless of the number of trials. The 95% btCIs (mean btCIs [.044 .073]), however, indicated that even for  $N = 6$ , FWER were not significantly larger than .05 (see Fig. 5), which according to Westfall and Young (1993) is an indication that the group analysis is valid. This is because the FWER for  $N = 6$  were significantly larger than those estimated for all other sample size (6 subjects simulation lowest mean FWER and btCIs: .056; [.042 .07]; highest FWER and btCIs across the remaining MC simulations: .03; [.02 .042]). For a complete table of all FWER and btCIs see supplementary section). **Overall**, within these experimental settings **our approach achieved the desired FWER at 5% under the group analysis setting.**

**3.1.1.2. SVM.** 95% bootstrap confidence intervals (btCIs) show that, overall, FWERs were either significantly ( $p < .05$ ) below or at .05 in all MC simulations regardless of the number of subjects and trials. We observed that, to an extent, SVM FWER were also modulated by the number of subjects. We report significantly ( $p < .05$ ) larger FWER for the sample size of 6, compared to a sample size of 18. This was true for all trial scenarios, except for the 12 trials simulations, where, while the mean FWER was larger for the sample size 6 compared to 18, this

difference was not significant ( $N = 6$ ; trials = 12 – .053 [.04 .067] and  $N = 18$ ; trials = 12 – .036 [.027 .049]). We estimated FWER of .059 [.046 .075] at the lowest number of subjects and trials (6 subjects with 4 trials each), and .03 [.021 .043] at the highest tested number of subjects and trials (18 subjects with 16 trials each). Moreover, a direct comparison between tMVPA and SVM further indicates that the number of subjects differentially modulates FWER. Unlike SVM, for tMVPA we report significant ( $p < .05$ ) differences between  $N = 6$  and  $N = 10$ , for all trial scenarios. Finally, we observed that for  $N = 6$  and  $N = 14$ , the FWER estimated for the 16 trial scenarios were significantly larger ( $p < .05$ ) for SVM than tMVPA. All mean FWER and btCI for the SVM group analysis are reported in the supplementary section.

#### 3.1.2. Single-subject analysis

**3.1.2.1. tMVPA.** Similarly, FWERs were not significantly above .05 in all MC simulations, regardless of the number of trials, as shown in Fig. 5 above. The highest FWER is 0.056 [0.043, 0.071] in the simulation with 4 trials, and the lowest FWER is 0.006 [0.003, 0.013] in the simulation with 16 trials (For a complete table of all FWER and btCIs see supplementary section). The 4 trials scenario produced significantly higher FWER than all other trials groups. Nonetheless, overall, within these experimental contexts **the simulation result clearly showed that in a single-subject analysis setting, our approach achieved the desired FWER at 5% even with as little as 4 trials.**

**3.1.2.2. SVM.** Other than for the 16 trials scenario, where FWER were slightly above .05 (.067 [.053 .084]), here too FWERs were not significantly ( $p < .05$ ) above .05 in all MC simulations, regardless of the number of trials. However, compared to tMVPA single subjects FWER, SVM single subject FWER was not modulated by the number of trials, as no significant differences were observed across the number of trials (Fig. 5). All Mean FWER and btCI for the SVM single subject analysis are reported in the supplementary section.

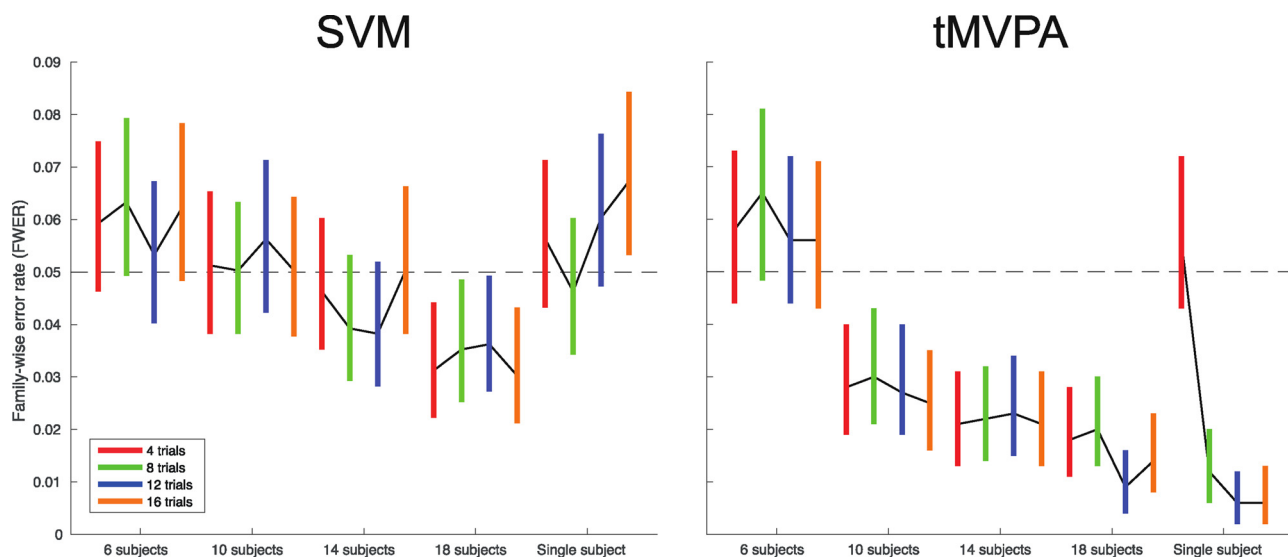


Fig. 5. FWER for all trials numbers, subjects groups and for the single subject scenario for SVM (left) and tMVPA (right). Here we show the family-wise error rate for the Monte Carlo simulated synthetic data with noise sampled from a Normal distribution. Error-bars represent the 95% bootstrap confidence interval of the Monte-Carlo simulation.

### 3.2. Power analysis

#### 3.2.1. Group-level analysis

**3.2.1.1. tMVPA.** btCIs analysis generally revealed that for the group scenario, regardless of the number of trials, tMVPA was relatively underpowered when differences across conditions were present in 50% and (to a lesser extent) 65% of the subjects. As shown in Fig. 6, the power of our approach increases as the number of subjects and the number of trials increases. With the effect introduced in 50% of the subjects, we estimated a power of 0.15 [0.132, 0.176] at the lowest number of subjects and trials (6 subjects with 4 trials each), to 0.27 [0.247, 0.297] at the highest tested number of subjects and trials (18 subjects with 16 trials each). Importantly, when we introduced the effect in 80% of the subjects, the 16 trials simulations led to significantly ( $p < .05$ ) higher power than the 8 and 4 trials scenarios for all sample sizes. Moreover, while generally displaying higher mean power, the 16 trials simulation never significantly ( $p > .05$ ) differed

from the 12 trials one. It is also worth noting that when  $N = 18$ , both the 16 and 12 trials simulations led to significantly higher power ( $p < .05$ ) compared to the 4 and 8 trials simulations, regardless of the number of subjects in which we introduced an effect. Furthermore, for the 14 subjects simulations only, the power estimated for the 4 trials scenario was significantly lower than all other group sizes, regardless of the number of subjects displaying the effect.

Not surprisingly, the highest statistical power was reached in the 18 subjects simulations with a minimum of 12 trials. Within this context, the tMVPA approaches 0.8 when we introduced the multivariate effect in 80% of the subjects (0.76 [0.731, 0.784], see also Fig. 6). A detailed report of mean power and btCIs for all MC simulations can be found in the supplementary section.

**3.2.1.2. SVM.** btCIs analysis generally revealed that, for the group scenarios, the sample size and number of trials have a greater modulatory impact on SVM's compared to tMVPA's power (Fig. 6). As

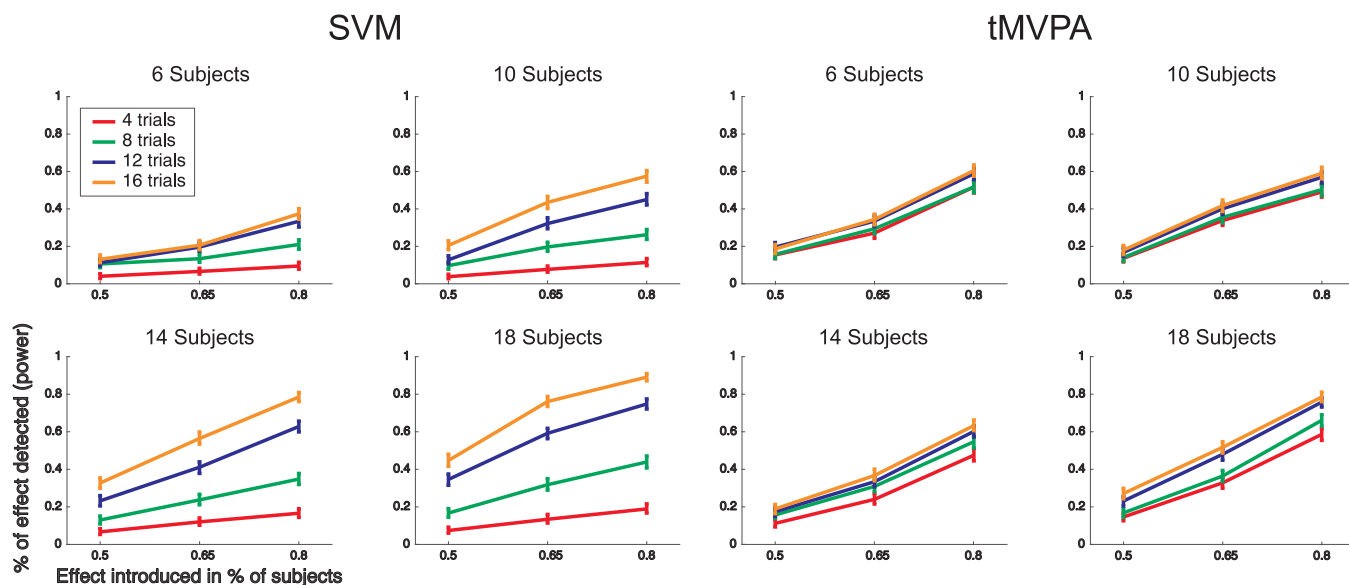


Fig. 6. Statistical power of the group-level analysis for SVM (left) and tMVPA(right). Error-bars represent the 95% bootstrapped confidence intervals across Monte-Carlo simulations.



shown in Fig. 6, with the effect introduced in 50% of the subjects, we estimated a power of .045 [.034 .06] at the lowest number of subjects and trials (6 subjects with 4 trials each), to .43 [.44 .46] at the highest tested number of subjects and trials (18 subjects with 16 trials each). Moreover, when we introduced the effect in 80% of the subjects, the 16 trials simulations led to significantly ( $p < .05$ ) higher power than the 8 and 4 trials scenarios for all sample sizes. Unlike tMVPA, we observed significantly ( $p < .05$ ) larger power for the 16 trials simulation compared to the 12 trials one, for all simulations, other than for the sample size  $N = 6$ . For this sample size, we additionally noted that tMVPA had significantly larger power ( $p < .05$ ) than SVM for all trials and % of data showing the effect, other than when the effect was introduced in 50% of the subjects. In this case, while both techniques only successfully detected the effect of interest less than 20% of the time and were therefore greatly underpowered, tMVPA (0.15 [0.132, 0.176]) showed significantly ( $p < .05$ ) larger power estimates than SVM in the 4 trials scenario.

For sample sizes of 14 and 18 instead, but not for samples of 6 and 10, the 16 trials simulation led to significantly higher power for SVM compared to tMVPA. We further observed that for the low trials numbers (i.e. 4 and 8) SVM is generally underpowered regardless of the sample size and percentage of subjects showing the effect. For the higher trials number instead, SVM power was significantly ( $p < .05$ ) modulated by the sample size and the percentage of subjects showing the effect. Here, too, the highest statistical power was reached in the 18 subjects and 16 trials. All mean power and btCI for SVM groups analysis are reported in the supplementary section.

### 3.2.2. Single-subject analysis

**3.2.2.1. tMPVA.** As shown in Fig. 7, the power of our approach increases as the number of trials increase. With the effect introduced in 50% of the trials, we estimated the power of our proposed approach at 0.10 [0.087, 0.126] with 4 trials, and at 0.28 [0.260, 0.317] with 16 trials. With the total number of 16 trials, the statistical power of the proposed approach reached 0.8 when the effect was introduced in 80% of the trials (0.80 [0.776, 0.827]). Importantly, regardless of the percentage of trials in which we introduced the effect, the 16 trials simulations led to significantly ( $p < .05$ ) higher power compared to all other simulations. Moreover, while significantly ( $p < .05$ ) lower than its 16 trials counterpart, the 12 trials simulations also led to significantly ( $p < .05$ ) higher power than the 2 remaining trials scenarios (see Fig. 7), peaking when 80% of the trials showed a multivariate effect (mean power: 0.719; btCIs: [0.691, 0.747]). A detailed report of mean power and btCIs for all MC simulations can be found in the supplementary section.

**3.2.2.2. SVM.** At the single subject level power analysis btCIs indicated that tMVPA generally performed better than SVM. SVM was generally underpowered, regardless of the sample size, number of trials and % of data showing the effect. When the effect was introduced in 80% of the trials, while the number of trials significantly ( $p < .05$ ) modulated SVM's power (ranging from .085 [.07 .1] for 4 trials to .45 [.42 .48] for 16 trials), across all trials number, SVM failed to reach the desired power of 0.8. Moreover, for the 16 trials scenario, tMVPA **always** reached a significantly ( $p < .05$ ) higher power than SVM. For all other trial numbers, no significant differences were observed across approaches when the effect was introduced in 50% of the trials. Within the simulation, we estimated the power of SVM at .089 [.072 .11] with 4 trials, and at .15 [.13 .18] with 16 trials. When the effect was introduced in more than 50% of the trials instead, tMVPA reached a significantly ( $p < .05$ ) higher power compared to SVM. A detailed report of mean power and btCIs for all MC simulations can be found in the supplementary section.

## 4. Discussion

In this paper, we present temporal Multivariate Pattern analysis (tMVPA), a method that we developed to quantify the temporal evolution of single trial dissimilarity across multivoxel patterns evoked by a given stimulus within a defined ROI. tMVPA builds upon the generation of single trial Representation Dissimilarity Matrices (stRDM) independently per ROI and condition: for all trials pairs, we iteratively cross-correlate the multivoxel pattern of BOLD % change across all possible time point combinations and we calculate its correlation distance (1-r). We then implemented a robust expanding sliding window approach to identify the temporal loci where statistically significant differences between conditions can be inferred (see methods). We validated this method for group and single subject analyses on data that were synthetically generated using noise (e.g. std across voxels, trials and time points) and signal (e.g. the BOLD time course) parameters derived from real fMRI data. We further benchmarked tMVPA performance against one of the most widely used multivariate tools in fMRI, SVM. This procedure was implemented to provide a relative reference against which to interpret FWER and power results. As such, SVM reports are not discussed in as much detail as tMVPA's.

Our validation analysis revealed three main findings: 1) our tMVPA approach reached the desired FWER ( $< = .05$ ) for both the group and single subject approach; and 2) Our power analysis showed that: a) for the *group* scenario, within this specific experimental context, tMVPA reached the desired power with a sample size of 18 subjects, each with 12 trials or more, when 80% of the participants displayed the desired multivariate effect. In all other contexts (i.e.  $< 18$  subjects,  $< 12$  trials

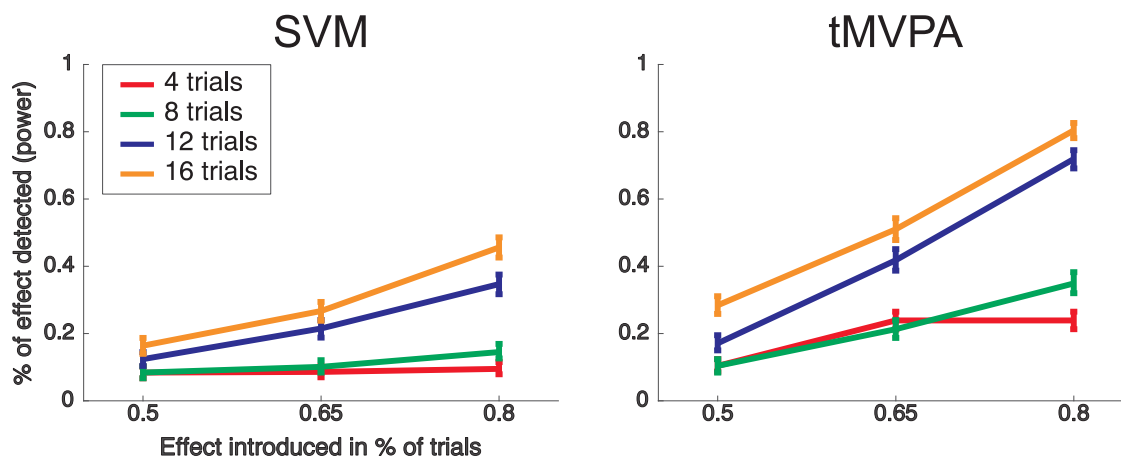


Fig. 7. Statistical power of the single-subject analysis for SVM (left) and tMVPA (right). Error bars represent the 95% bootstrapped confidence intervals across Monte-Carlo simulations.

and < 80% of subjects showing the effect), our method tended to be relatively underpowered and b) similarly, for the *single subject* scenario, our approach reached the desired power with at least 12 trials, when the multivariate effect of interest was present in 80% of them. All other simulation scenarios failed to reach the target power. 3) tMVPA performed better than SVM in the single subject scenario and in the group scenarios with low subject number (i.e.  $N = 6$ ;  $N = 10$ ); conversely with larger  $N$  (i.e.  $N = 14$ ;  $N = 18$ ), SVM outperformed tMVPA. Our findings are discussed in detail below.

#### 4.1. tMVPA FWER and power analyses

##### 4.1.1. Group analysis

Simulation results indicate that when the sample size is less than 8 subjects, regardless of the number of trials per condition or percentage of effect introduced, our technique is significantly ( $p < .05$ ) below the lower margin of the desired FWER (.05) (Fig. 7). Moreover, it is worth noting that for  $N = 6$ , FWER is not significantly larger than .05, a finding which advocates the validity of the group analysis (Westfall and Young, 1993; Westfall et al., 1993) (at least in terms of false positive rate). Nonetheless, we observe that when  $N = 6$ , the estimated FWERs are significantly larger than all other samples and MC simulations (see Fig. 5), which may significantly inflate the occurrence of Type I errors for this specific sample size.

Furthermore, the results of our power analysis suggest that, within these experimental settings, a minimum of 18 subjects with at least 12 trials per condition is required to achieve adequate statistical power. While a sample size of 18 subjects could be regarded as sufficient for the majority of current fMRI studies, low  $N$  is considered one of the main culprits for the so called “replication crisis” (Button et al., 2013; Maxwell et al., 2015; Schooler, 2014). Consequently, the field of science as a whole, and specifically disciplines such as psychology and cognitive neuroscience, is undergoing a targeted endeavor aimed at augmenting the experimental sample size, in an effort to increase statistical power and produce replicable results (Button et al., 2013; Maxwell et al., 2015; Schooler, 2014). Within this context, a sample size of 18 participants does not therefore seem prohibitive.

##### 4.1.2. Single subject analysis

One of the main advantages of the tMVPA analysis is the exploitation of single trials in computing temporal RDMs. Generating RDMs by correlating all possible single trial pairs leads to a distribution of single trial RDMs (stRDMs), which allows one to carry out second order inferential statistics at the single subject level. This procedure permits full exploitation of the trial-by-trial variability, which is lost in the group-level approach due to averaging. It is worth noting that, while still not significantly larger than the desired FWER of .05, the single subject validation procedure indicates that the 4 trial scenario produces significantly more FWER than all other trials groups. Not surprisingly, the peak statistical power is achieved for the 16 trials simulations (Fig. 7). The 12 trials simulations, however, led to significantly higher statistical power than its 4 and 8 trials counterparts. Crucially, when the multivariate effect of interest is present in at least 80% of the trials, our approach achieves the desired power with 12 trials or more. With a minimum of 12 trials across runs, our approach reaches the desired power and FWER. This finding makes our tMVPA appealing and powerful, not only to carry out single subject statistics, but to investigate issues that have thus far been elusive to the world of cognitive neuroscience, such as individual differences in the BOLD response. Moreover, the ability to conduct single subject statistics is additionally advantageous for both piloting experimental designs and for analyzing experiments which are limited by low subject numbers due to, amongst other things, the time required in preprocessing and by-hand analysis (e.g., 7 T laminar/columnar studies). Importantly, we show that we can carry out single subject analysis with a relatively parsimonious experimental design, which does not require a large number of trials.

#### 4.2. tMVPA vs. SVM

While both draw on multivariate patterns, tMVPA and SVM are two fundamentally different approaches, the latter representing a supervised machine learning decoding algorithm; and the former estimating the single trial temporal evolution of the dissimilarity of multivoxel patterns of activation. In spite of these differences, both methods led to comparable patterns of power and FWER. A direct comparison between these two approaches showed that: a) for the single subject analysis tMVPA performed generally better than SVM; and b) for the group scenario, SVM slightly outperformed tMVPA with larger subject and trial numbers; conversely, tMVPA did better in the low  $N$  and low trials scenarios. These results are not surprising. The better performance observed for tMVPA in low trials and sample sizes simulations is related to the use of single trial computations. We designed tMVPA to exploit information and statistical power conveyed by single trials. As shown here, this can boost statistical power in power-limited scenarios (such as, for example, low sample sizes), while maintaining the desired FWER, thus improving the reliability of the method. The superior performance of SVM with higher  $N$  and trials can be related to a number of factors, including additional preprocessing steps (i.e. using GLM coefficients instead of raw BOLD signal) and the lower dimensionality (i.e. the lack of temporal dimension). These points are discussed in details below.

##### 4.2.1. Group analysis: FWER

Like tMVPA, overall SVM FWER were either significantly ( $p < .05$ ) below or at .05. However, we observed a different modulatory impact of sample size on FWER estimated for tMVPA compared to those estimated for SVM. tMVPA FWER was significantly lower for  $N = 10$  compared to  $N = 6$ . This was not true for SVM FWER. Moreover, for the 16 trials scenario, SVM FWER were significantly larger than tMVPA's for  $N = 10$  and  $N = 14$ . These results indicate that tMVPA is either comparable or likely to commit fewer Type I errors compared to SVM. These observations can be potentially accounted for by some of the steps in the analysis procedure. Firstly, hypothesis testing was performed by comparing the bounds of non-parametric bootstrap confidence intervals. This technique builds upon sampling the raw data points with replacement to form bootstrapped distribution from which confidence intervals can be estimated. As such, one of the main strengths of this non-parametric approach is the relatively low sensitivity to the less frequently occurring extreme values, when  $N$  is large (Wilcox, 2005). Conversely, however, the opposite can be true when  $N$  is low (Wilcox, 2005). Consider, for example, the case where we have one extreme value in a sample size of 6. This value represents  $\sim 16.6\%$  of the data. One extreme value in a sample size of 18 instead, only represents  $\sim 5.5\%$ . In this example, when  $N = 6$ , the one extreme value is over 300% more likely to be sampled than when  $N = 18$ , leading to relatively larger variance for the bootstrap sample generated from  $N = 6$ . Here, FWER estimation was performed by testing for significant differences between 2 conditions with identical mean (i.e.  $H_0$ ; see methods) but different variance. As such, false positives are a direct outcome of sample variance, with larger variance leading to higher FWER. These considerations would explain why  $N$  modulated FWER for both SVM and tMVPA and specifically why FWER is larger for  $N = 6$  compared to  $N = 18$ . Additionally, for tMVPA we adjust the alpha threshold for statistical significance to account for the multiple tests carried out across time (i.e. multiple comparison correction – MCC). In light of these considerations, we argue that the observed differences in FWER rates between the two approaches can be explained as follows. At the lowest  $N$ , subject variability dominates, thus modulating FWER estimation regardless of the adjusted alpha threshold. On the other hand, at the highest  $N$ , the impact of extreme values is combatted by the adequate sample size, stabilizing FWER and leading to a convergence of this metric across the 2 approaches. At middle  $N$ s instead, where sample size, albeit larger, is not quite adequate (see tMVPA group analysis

paragraph), adjusting the alpha threshold has the largest impact on FWER estimation, leading to the reported differences between the approaches.

#### 4.2.2. Group analysis: power

Sample size and number of trials had a greater modulatory impact on SVM's compared to tMVPA's power (Fig. 6). For the 2 lower sample sizes tMVPA generally achieved higher power than SVM (other than when  $N = 10$  and trials = 16). At higher  $N$ s instead, SVM power was consistently larger than tMVPA's for the highest trial simulations and lower for the lower trial simulations. These findings are not surprising and can be construed as mainly rooted in tMVPA's and SVM's differential exploitation of the single trials. One of the main features of our technique is in fact the ability to draw on single trials to compute single trial representational dissimilarity matrices (i.e. stRDMS). In doing so, we exploit all possible trial pairings, thus effectively increasing the usable amount of data. The 4 trials scenario therefore generates 6 stRDMS, by computing the correlational distance between all unique trial pairings. This procedure generates 50% more data points than those available with SVM, therefore leading to higher statistical power. Our results suggest that, in lower  $N$  scenarios, the power gain conferred by single trial exploitation counters the power loss related to a small number of subjects. Contemporaneously, it would be plausible to assume that training SVM with only 3 trials would lead to over-fitting, with consequently poor generalizability of the model and low accuracy scores. This observation would account for the observed low SVM power in trial contexts 4 and 8, regardless of  $N$  and the number of data points showing an effect.

Finally, the larger SVM compared to tMVPA power estimates reported for the highest trial context can be explained by at least 2 factors. Firstly, unlike tMVPA, computed over the raw BOLD percent signal change, SVM classification accuracy was computed on GLM beta weights. As such, the latter involved one further preprocessing step (fitting a GLM), which further removes putative noise from the data (i.e. the GLM error term), thus indirectly increasing SVM's ability to detect multivariate differences across conditions. Secondly, as mentioned above, MCC, implemented to account for multiple tests over time, renders tMVPA more conservative than SVM.

#### 4.2.3. Single subject: FWER and power

Both FWER and power estimation clearly indicate that, at least within these experimental contexts, tMVPA performs significantly better than SVM in the single subject context. Again, this is not surprising as this method was originally developed for single subject analysis and later adapted to perform group level analyses. As such, tMVPA exploits the advantage conferred by single trial analysis. Power and FWER considerations raised above for the group analysis (see the *Group tMVPA vs. SVM* paragraphs) also apply for the single subject analysis. In short, the overall larger power displayed by tMVPA can be related to the computation of stRDMS across all unique trial pairings, which effectively increases the number of data points (and consequently the statistical power) used to infer statistical significance. At the same time, adjusting the alpha threshold to account for multiple comparison over time, lead to significantly lower tMVPA FWER compared to SVM for all but the 4 trial case.

#### 4.3. General considerations on FWER and power analysis

Though tMVPA was underpowered in simulations where 65% or fewer data points contained the effect of interest for both the group and single subject analyses, we argue that this is a potential strength rather than a weakness of our approach. While more likely to incur Type II errors (i.e. failing to reject  $H_0$ ), we would question the sensitivity, validity, and especially the generalizability of a method reporting statistical significance when only 65% or fewer data points display the effect being claimed. This argument becomes even more relevant in light of

the recent emphasis of the scientific community on producing highly replicable studies, following the so called “replication crisis” (Schooler, 2014). We advocate the use of relatively more conservative statistical approaches, as we believe that overpowered statistical approaches can be regarded as one of the causes of the aforementioned replication crisis (Anderson and Maxwell, 2017). Furthermore, it is worth noting that the values estimated here (and the considerations that follow) are specific to our experimental settings and image acquisition parameters. Statistical power, for example, is a direct function of sample size, effect size and alpha threshold. It therefore becomes difficult to interpret its absolute values and generalize them beyond the experimental setting within which they are computed. The direct comparison between tMVPA and SVM, however, provides a relative context that can allow extrapolation of our findings beyond the current settings. In light of its widespread use and demonstrated statistical reliability (Song et al., 2011), SVM represent an ideal candidate to provide a relative frame of reference for interpreting tMVPA statistical evaluation. For example, the relatively higher SVM power in higher  $N$  and trials scenarios and the opposite pattern observed for lower  $N$  and trials could represent more general properties of these techniques, related, as explain earlier, to precise analytical steps and computational features. Moreover, tMVPA's statistical evaluation provides additional meaningful information if we interpret power and FWER estimates within the realm of our highly parameterize space, by indicating, for example, which parameter had a *relatively* higher modulatory impact on the reliability of our technique.

It is also worth noting that we chose a stimulation paradigm (i.e. 850 ms visual stimulation; 4 trials per run) that is likely to lead to low evoked BOLD amplitude and, consequently, low experimental SNR (i.e. BOLD amplitude over trials measurement error). Under different stimulation regimes, such as longer stimulus presentation or block design experiments, we would expect higher statistical power or lower  $N$  to achieve the desired power. Moreover, at higher fields (i.e. 7 T or above) the increase in both temporal and image SNR (Ugurbil, 2014) will be paired with a boost in statistical power. As such, the statistical power computed here in a relatively low SNR regime, represents a conservative estimate for the proposed approach.

#### 4.4. Temporal multivariate approach to fMRI

Traditionally, due to the sluggish nature of the hemodynamic based BOLD signal (Boynton et al., 1996; Ogawa et al., 1993), fMRI's temporal resolution has traditionally been overlooked, deemed to be too inaccurate to measure the temporal dynamics of neural processing. More recently, however, a number of animal studies have begun exploring the temporal dimension of the BOLD signal. Functional images have been recorded in marmosets with a temporal resolution of 200 ms (Yen et al., 2018) and in rats with 40 ms (Silva and Koretsky, 2002). Furthermore, human recordings have suggested that increasing fMRI temporal resolution may reveal insights into the temporal dynamics of neural processing. For example, recent evidence put forward by Lewis et al. (2016) suggests that fMRI can measure neural oscillatory activity at a much higher rate than previously suggested, specifically up to 1 Hz. Accordingly, Siero et al. (2011) showed that, away from large draining, vessels the hemodynamic response function peaks  $\sim 2$  s earlier and is approximately 1 s narrower than previously reported, thus indicating that the neurovascular coupling occurs at a much shorter time-scale. Additionally, Vu et al.'s (Vu et al., 2016) work also advocates the importance of the BOLD temporal dimension. These authors showed that that, using MVPA, it is possible to extract timing information at fast TRs (i.e. 500 ms) that would otherwise be inaccessible (Vu et al., 2016).

These observations highlight the growing interest in the temporal dynamics of the BOLD signal, motivating the need for novel analytical tools specifically tailored to extract BOLD temporal information. Within this context, the method we developed is highly advantageous in that it incorporates the multivariate dimension in the temporal analysis of the

BOLD signal, rendering potentially unexplored temporal features accessible. This multivariate dimension comes from considering the spatial pattern of BOLD activity across the voxels population within a given ROI at every time-point. As such, tMVPA extends the power of fMRI, which has historically been in the spatial domain, to the much less studied temporal dimension.

tMVPA thus allows investigating the temporal evolution of neural representation, which is incredibly valuable for exploring a wide range of phenomena, from visual illusions (Ernst and Bu, 2004; Schrater et al., 2004), real world scenes, and a variety of auditory paradigms (Baumann et al., 2010; Lee et al., 2012). As such, our method can be broadly applied to a large domain of stimulus paradigms.

Another interesting feature of tMVPA is the fact that paradigms utilizing active behavioral judgments of stimuli (as in Ramon et al., 2015) may choose to align the analysis with either the stimulus onset or the behavioral response. This allows investigating response- as well as stimulus-locked modulations of neural representations over time.

It is also worth considering the nature of the effect being observed with tMVPA. Our technique measures multivariate activity at the population level accessible with fMRI [ $\sim 640,000$  neurons (Lent et al., 2012)], and is as such constrained by the temporal lag of the BOLD signal (Ogawa et al., 1993). While these constraints limit its temporal precision, especially relatively to the resolution available using invasive electrophysiological techniques (Meyers et al., 2015), tMVPA does provide valuable insights into the *relative* temporal dynamics of the neural processes captured with fMRI. In essence, while tMVPA won't provide direct insights into the actual temporal window of neural processing, the careful investigation of temporal aspects of the BOLD signal could provide important information regarding the neural substrates of cognition (Ogawa et al., 2000; Smith et al., 2012). For example, the relative BOLD latency differences between experimental conditions can be related to diverse cognitive processes (Gentile et al., 2017; Henson et al., 2002).

tMVPA analysis already proved useful by revealing crucial differences in the temporal processing of familiar and unfamiliar faces in the left fusiform face area and in the bilateral amygdala (Ramon et al., 2015). Importantly, in Ramon et al. (2015) these differences would have remained undetected using traditional temporal mean BOLD amplitude analysis techniques, as we did not observe significant differences between the average (across voxels and trials) BOLD time courses of familiar and unfamiliar faces. Accordingly, our simulations were carried out on synthetic data that were carefully generated with the absence of mean BOLD amplitude differences across conditions (Fig. 3). We thus replicated what we originally showed in Ramon et al. (2015), namely, the ability of the tMVPA approach to detect genuine temporal multivariate effects or ones not driven by mere mean BOLD amplitude differences.

It must be noted that the differences between this work and Ramon et al. (2015) are substantial both in terms of stimulation paradigm and MR acquisition parameters. Their functional scans were acquired using a repeated, single-shot echo planar imaging sequence with 3.5-mm isotropic voxel, a  $64 \times 64$  matrix, a TE of 50 ms, TR of 1250 ms, FA of  $90^\circ$  and FOV of 224 mm. Moreover, Ramon et al. (2015) used a novel visual paradigm where a face stimulus was kept on screen for a duration of approximately 19–21 TRs, followed by a fixation period lasting 6–8 TRs. Yet, in spite of these differences, in both datasets our technique uncovered effects that were not detected when using traditional mean BOLD amplitude methods focusing on amplitude differences between average time courses.

#### 4.4.1. Validation on synthetic versus real data

It is important to consider that the multivariate data used to assess this technique were generated synthetically (see methods). Our technique was initially conceived for use with experimentally derived data (Ramon et al., 2015). As the goal of the present study is to assess the experimental parameters and conditions under which our technique is

most useful, the ability to manipulate these variables is crucial and thus synthetic data is ultimately necessary. As previously mentioned, in an effort to generate a synthetic data set with realistic signal and noise properties, we used noise and signal estimates from real fMRI data. We approximated the fMRI signal by averaging BOLD time courses across voxels, trials, and conditions, and the amount of noise by measuring the variability (i.e. standard deviation) across voxels, trials and time-points. Hybrid approaches to synthetic data generations, such as the one implemented here, are highly beneficial (Welvaert and Rosseel, 2013). They provide full control over the data set, while preserving realistic signal to noise estimates and, according to (Welvaert and Rosseel, 2013), may represent the ideal data generation procedure for statistical validation. Our data generation approach, however, builds upon random sampling of variance and signal properties across voxels, ROIs, conditions, and subjects (see methods). This procedure effectively impairs the original temporal and spatial autocorrelation present in fMRI data. In the present study, we did not attempt to re-inject temporal and spatial autocorrelation in the synthetic data. The reason behind this choice is twofold. Firstly, fMRI has multiple sources of noise (e.g. thermal, physiological, motion, task), each of which is characterized by different distributions and parameters, making it difficult to accurately and comprehensively model all noise sources. As such, an exhaustive model that allows generation of realistic fMRI noise has yet to be formulated. In order to introduce synthetic but *realistic* spatio-temporal auto-correlated noise in simulated fMRI data, there is first a need to formulate a comprehensive and realistic noise model. However, the quest for an exhaustive model for fMRI data (including noise) generation is challenging enough to require a study in and of itself tailored to tackle this specific endeavor (Davis et al., 2014) and, as such, is well beyond the scope of this article. Additionally, given the lack of a “ground-truth” noise model, noise estimates may be inaccurate or misrepresent the contribution of difference noise sources and, as such, noise injection may have a negative impact on the validation procedure as a whole. Secondly, we argue that the impact of spatio-temporal auto-correlated noise is minimal within these specific settings. The structure of the stRDMs when considering real, as opposed to synthetic, data can be seen in Fig. 1. Patches of similarity (cool colors) and dissimilarity (warm colors) exist in clusters of approximately 3–4 TRs. Such structure is due to the inherent spatiotemporal autocorrelation present in the BOLD signal, which is not dependent on experimental manipulations. Rather, it is a direct outcome of the HRF response properties. Specifically, BOLD activation for all voxels will synchronously rise for approximately the first 6 s after stimulus onset (varying depending on stimulus presentation time), and then decrease for the following 6 s, thus generating the structure visible in the matrices in Fig. 1. This structure will therefore be shared across conditions and subtracted out when performing the linear contrast between the stRDMs across conditions (see methods). As such, the inherent presence of autocorrelation in fMRI data, which is shared across conditions, becomes irrelevant in evaluating the validity of our validation procedure.

## 5. Conclusion

In summary, we have developed a method for examining the representational content of fMRI data as a function of time, thereby enabling the investigation of the temporal evolution of neural representation. The method, that builds upon fMRI most recognized strength – namely its spatial resolution – to analyze BOLD temporal dynamics, consists of creating Single Trial Representational Dissimilarity Matrices (stRDMs) to measure the dissimilarity between the neural representations elicited by each acquired time point of a BOLD time course. Along with the addition of a multivariate dimension to BOLD temporal analyses, tMVPA permits performing single subject's inferential statistics by considering single trial distributions. Importantly, we implemented an expanding, sliding window method for inferring statistical significance that allows precise identification of



the temporal window of an effect. We validated our temporal multivariate pattern analysis (tMVPA) in both group and single subject settings using synthetically generated data and compared its performance to that of SVM. Our results show that tMVPA outperforms SVM in low N and trials contexts and in single subject analysis; and that, conversely, SVM outperformed tMVPA in high N and trials contexts. We relate these findings to tMVPA's exploitation of single trial distributions. We also show that, both in simulated as well as real settings (see Ramon et al., 2015), our tMVPA is capable of detecting multivariate effects between experimental conditions in the absence of mean BOLD amplitude differences. The technique presented here expands on traditional multivariate fMRI analyses, facilitating investigations of the evolution of neural representations over time.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.jneumeth.2018.06.029>.

## References

- Anderson, S.F., Maxwell, S.E., 2017. Addressing the “replication crisis”: using original studies to design replication studies with appropriate statistical power addressing the “replication crisis”: using original studies to design replication studies with appropriate statistical power. *Multivar. Behav. Res.* 52 (3), 305–324. <https://doi.org/10.1080/00273171.2017.1289361>.
- Avossa, G., Shulman, G.L., Corbetta, M., Shulman, G.L., Iden, M.C., 2003. Identification of Cerebral Networks by Classification of the Shape of BOLD Responses. pp. 360–371.
- Bailey, C.J., Sangahalli, B.G., Herman, P., Blumenfeld, H., Gjedde, A., 2013. Analysis of Time and Space Invariance of BOLD Responses in the Rat Visual System. January. <https://doi.org/10.1093/cercor/bhs008>.
- Ball, K., Sekuler, R., 1982. A specific and enduring improvement in visual motion discrimination. *Science* 218 (4573), 697–698.
- Baumann, S., Grif, T.D., Rees, A., Hunter, D., Sun, L., Thiele, A., 2010. Characterisation of the BOLD response time course at different levels of the auditory pathway in non-human primates. *NeuroImage* 50, 1099–1108. <https://doi.org/10.1016/j.neuroimage.2009.12.103>.
- Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J., 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* 16 (13), 4207–4221.
- Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14 (May), 365–376. <https://doi.org/10.1038/nrn3475>.
- Carlson, T.A., Schrater, P., He, S., 1999. Patterns of Activity in the Categorical Representations of Objects. pp. 704–717.
- Cox, D.D., Savoy, R.L., 2003. Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage* 19, 261–270. [https://doi.org/10.1016/S1053-8119\(03\)00049-1](https://doi.org/10.1016/S1053-8119(03)00049-1).
- Davis, T., Larocque, K.F., Mumford, J.A., Norman, K.A., Wagner, A.D., Poldrack, R.A., 2014. What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *NeuroImage* 97, 271–283. <https://doi.org/10.1016/j.neuroimage.2014.04.037>.
- Edwards, G., Vetter, P., Mcgruer, F., Petro, L.S., Muckli, L., 2017. Predictive feedback to V1 dynamically updates with sensory input. *Sci. Rep.* 1–12. <https://doi.org/10.1038/s41598-017-16093-y>.
- Ernst, M.O., Bu, H.H., 2004. Merging the senses into a robust percept. *Trends Cogn. Sci.* 8 (4), 162–169. <https://doi.org/10.1016/j.tics.2004.02.002>.
- Feinberg, D.A., Moeller, S., Smith, S.M., Auerbach, E., Ramanna, S., Glasser, M.F., et al., 2010. Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. *PLoS One* 5 (12).
- Formisano, E., Linden, D.E.J., Salle, F., Di Trojano, L., Esposito, F., Sack, A.T., et al., 2002. Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. *Neuron* 35, 185–194.
- Freeman, J., Brouwer, G.J., Heeger, D.J., Merriam, E.P., 2011. Orientation decoding depends on maps. *Not. Commun.* 31 (13), 4792–4804. <https://doi.org/10.1523/JNEUROSCI.5160-10.2011>.
- Gentile, F., Ales, J., Rossion, B., 2017. Being BOLD: the neural dynamics of face perception. *Hum. Brain Mapp.* 139, 120–139. <https://doi.org/10.1002/hbm.23348>. March 2016.
- Goense, J.B., Logothetis, N.K., 2008. Neurophysiology of the BOLD signal in awake monkeys. *Curr. Biol.* 18 (9), 631–640.
- Gribble, P., 1996. Temporal constraints on the McGurk effect. *Percept. Psychophys.* 58 (3).
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293 (September), 2425–2431.
- Haxby, J.V., Toole, A.J.O., Jiang, F., 2005. Partially distributed representations of objects and faces in ventral temporal cortex. *J. Cogn. Neurosci.* 580–590.
- Henson, R.N.A., Price, C.J., Rugg, M.D., Turner, R., Friston, K.J., 2002. Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. *NeuroImage* 97, 83–97. <https://doi.org/10.1006/nimg.2001.0940>.
- Husk, J.S., Bennett, P.J., Sekuler, A.B., 2007. Inverting houses and textures: investigating the characteristics of learned inversion effects. *Vision Res.* 47, 3350–3359. <https://doi.org/10.1016/j.visres.2007.09.017>.
- Johansson, G., 1973. Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14 (2), 201–211.
- Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8 (5), 679–685. <https://doi.org/10.1038/nn1444>.
- Kriegeskorte, N., Bandettini, P., 2007. Analyzing for information, not activation, to exploit high-resolution fMRI. *NeuroImage* 38, 649–662. <https://doi.org/10.1016/j.neuroimage.2007.02.022>.
- Kriegeskorte, N., Kievit, R.A., 2013. Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn. Sci.* 17 (8), 401–412. <https://doi.org/10.1016/j.tics.2013.06.007>.
- Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2 (November), 4. <https://doi.org/10.3389/fnro.2008.06.004.2008>.
- Lee, Y., Turkeltaub, P., Granger, R., Raizada, R.D.S., 2012. Categorical speech processing in Broca's Area: an fMRI study using multivariate pattern-based analysis. *J. Neurosci.* 32 (11), 3942–3948. <https://doi.org/10.1523/JNEUROSCI.3814-11.2012>.
- Lent, R., Azevedo, F.A.C., Andrade-moraes, C.H., Pinto, A.V.O., 2012. How many neurons do you have? Some dogmas of quantitative neuroscience under revision. *Eur. J. Neurosci.* 35, 1–9. <https://doi.org/10.1111/j.1460-9568.2011.07923.x>.
- Lewandowski, D., Kurowiczka, D., Joe, H., 2009. Generating random correlation matrices based on vines and extended onion method. *J. Multivar. Anal.* 100 (9), 1989–2001. <https://doi.org/10.1016/j.jmva.2009.04.008>.
- Lewis, L.D., Setsompop, K., Rosen, B.R., Polimeni, J.R., 2016. Fast fMRI can detect oscillatory neural activity in humans. *PNAS* 113 (43), E6679–E6685. <https://doi.org/10.1073/pnas.1608117113>.
- Lin, C.-C.C.-J., 2011. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 1. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412 (6843), 150.
- Maier, A., Wilke, M., Aura, C., Zhu, C., Ye, F.Q., Leopold, D.A., 2008. Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nat. Neurosci.* 11 (10), 1193–1200. <https://doi.org/10.1038/nn.2173>.
- Mannion, D.J., McDonald, J.S., Clifford, C.W.G., 2009. Discrimination of the local orientation structure of spiral glass patterns early in human visual cortex. *NeuroImage* 46 (2), 511–515. <https://doi.org/10.1016/j.neuroimage.2009.01.052>.
- Mannion, D.J., Kersten, D.J., Olman, C.A., 2015. Scene coherence can affect the local response to natural images in human V1. *Eur. J. Neurosci.* 42, 2895–2903. <https://doi.org/10.1111/ejn.13082>.
- Matsumoto, D., Ekman, P., 1988. Japanese and Caucasian Facial Expressions of Emotion and Neutral Faces. *els Autors*.
- Maxwell, S.E., Lau, M.Y., Howard, G.S., Scott, E., 2015. Is psychology suffering from a replication crisis? *Am. Psychol.* 70 (6), 487–498.
- Mcquire, J.T., Kable, J.W., 2015. Medial prefrontal cortical activity reflects dynamic re-evaluation during voluntary persistence. *Nat. Neurosci.* 18 (5). <https://doi.org/10.1038/nn.3994>.
- Meyers, E.M., Borzello, M., Freiwald, W.A., Tsao, D., 2015. Intelligent information loss: the coding of facial identity, head pose, and non-face information in the Macaque face patch system. *J. Neurosci.* 35 (18), 7069–7081. <https://doi.org/10.1523/JNEUROSCI.3086-14.2015>.
- Michel, C., Rossion, B., Han, J., Chung, C.-S., Caldara, R., 2006. Holistic processing is finely tuned for faces of one's own race. *Psychol. Sci.* 17 (7), 608–615.
- Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Ugurbil, 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. *Magn. Reson. Med.* 63 (5), 1144–1153. <https://doi.org/10.1002/mrm.22361>.
- Ogawa, S., Menon, R.S., Tank, D.W., Kim, S.G., Merkle, H., Ellermann, J.M., et al., 1993. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. *Biophys. J.* 64 (3), 803–812. [https://doi.org/10.1016/S0006-3495\(93\)81441-3](https://doi.org/10.1016/S0006-3495(93)81441-3).
- Ogawa, S., Lee, T., Steposki, R., Chen, W., Zhu, X., Ugurbil, K., 2000. An approach to probe some neural systems interaction by functional MRI at neural time scale down to milliseconds. *PNAS* 97 (20), 11026–11031.
- Op de Beeck, H., 2010. Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? *NeuroImage* 49 (3), 1943–1948. <https://doi.org/10.1016/j.neuroimage.2009.02.047>.
- Ramon, M., Vizioli, L., Liu-shuang, J., Rossion, B., 2015. Neural microgenesis of personally familiar face recognition. *PNAS*. <https://doi.org/10.1073/pnas.1414929112>.
- Sasaki, Y., Rajimehr, R., Kim, B.W., Ekstrom, L.B., Vanduffel, W., Tootell, R.B.H., 2006. The Radial Bias: A Different Slant on Visual Orientation Sensitivity in Human and Nonhuman Primates. pp. 661–670. <https://doi.org/10.1016/j.neuron.2006.07.021>.
- Schooler, J.W., 2014. Metascience could rescue the “replication crisis”. *Nature* 515, 9.
- Schrater, P., Kersten, D., Murray, S.O., 2004. Perceptual grouping and the interactions between visual cortical areas. *Neural Netw.* 17 (5–6), 695–705. <https://doi.org/10.1016/j.neunet.2004.03.010>.
- Shadlen, M.N., Newsome, W.T., 1998. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* 18

- (10), 3870–3896.
- Siero, J.C., Petridou, N., Hoogduin, H., Luijten, P.R., Ramsey, N.F., 2011. Cortical depth-dependent temporal dynamics of the BOLD response in the human brain. *J. Cereb. Bloodflow Metab.* 31 (10), 1999–2008.
- Silva, A.C., Koretsky, A.P., 2002. Laminar specificity of functional MRI onset times during somatosensory stimulation in rat. *PNAS* 99 (23), 15182–15187.
- Smith, S.M., Miller, K.L., Moeller, S., Xu, J., Auerbach, E.J., Woolrich, M.W., 2012. Temporally-independent functional modes of spontaneous brain activity. *PNAS* 109 (8), 3131–3136. <https://doi.org/10.1073/pnas.1121329109>.
- Song, S., Zhan, Z., Long, Z., Zhang, J., Yao, L., 2011. Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data. *PLoS One* 6 (2). <https://doi.org/10.1371/journal.pone.0017191>.
- Strother, S.C., Anderson, J., Hansen, L.K., Kjems, U., Kustra, R., Sidtis, J., et al., 2002. The quantitative evaluation of functional neuroimaging experiments : the NPAIRS data analysis framework. *Neuroimage* 771, 747–771. <https://doi.org/10.1006/nimg.2001.1034>.
- Talairach, J., Tournoux, P., 1988. *Co-Planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System: An Approach to Cerebral Imaging*.
- Troje, N.F., 2002. Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. *J. Vis.* 2 (5), 371–387. <https://doi.org/10.1167/2.5.2>.
- Ugurbil, K., 2014. Magnetic resonance imaging at ultrahigh fields. *IEEE Trans. Biomed. Eng.* 61 (5), 1364–1379.
- Vu, A.T., Phillips, J.S., Kay, K., Phillips, M.E., Johnson, M.R., Shinkareva, S.V., et al., 2016. Using precise word timing information improves decoding accuracy in a multiband-accelerated multimodal reading experiment. *Cognit. Neuropsychol.* 33, 265–275.
- Welvaert, M., Rosseel, Y., 2013. On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. *PLoS One* 8 (11). <https://doi.org/10.1371/journal.pone.0077089>.
- Westfall, P.H., Young, S.S., 1993. *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment* Vol. 279 John Wiley & Sons.
- Westfall, P.H., Young, S.S., Wright, S.P., 1993. On adjusting P-values for multiplicity. *Biometrics* 49 (3), 941–945. Published by: International Biometric Society, Stable URL: <http://www.jstor.org/stable/2532216>.
- Wilcox, R.R., 2005. *Introduction to Robust Estimation and Hypothesis Testing*, 2nd ed. Academic Press, San Diego.
- Yen, C.C., Papoti, D., Silva, A.C., 2018. Investigating the spatiotemporal characteristics of the deoxyhemoglobin-related and deoxyhemoglobin-unrelated functional hemodynamic response across cortical layers in awake marmosets. *NeuroImage* 164, 121–130. <https://doi.org/10.1016/j.neuroimage.2017.03.005>. March 2017.