

# Functional Magnetic Resonance Imaging Signal Modelling and Contrasts: an Example of Manual Praxis Tasks

**M. Buchwald**

*Poznan Supercomputing and Networking Center  
ul. Jana Pawła II 10, 61-139 Poznań, Poland  
E-mail: mikolaj.buchwald@man.poznan.pl*

Received: 2 December 2021; revised: 14 December 2021; accepted: 16 December 2021; published online: 29 December 2021

**Abstract:** The goal of neuroscience as a discipline is to understand how the neural system is organized in the brain, giving rise to mental processes and the control of behavior. One of the most frequently utilized methods in neuroscientific studies is the functional magnetic resonance imaging (fMRI), which is a non-invasive technique for quantifying brain processes dynamics. In a standard fMRI procedure, the hypothesis of the correlation between a cognitive task and the observed physiological signal is tested. This way, a certain computational model of a given brain mechanism can be validated. The procedure of modelling fMRI signal time course will be explained in this article as exemplified by planning functional grasps of tools. Subsequently, the results of contrasting model parameter estimates will be presented for a different experiment on manual praxis skills, i.e., bimanual tool grasps and manipulations.

**Key words:** fMRI, signal modelling, cognitive subtraction, GLM, praxis

---

## I. INTRODUCTION

Computational methods are an essential part of neuroscience. Since the beginnings of the neural paradigm, the major goal of the neuroscientific approach was to provide a detailed description of what the structure of the brain is and explain, as clearly as possible, how brain functions are implemented in the underlying neural tissue [1, 2]. More recently, also the connectionist approaches are gaining popularity, emphasizing the role of brain structural and functional networks in the way the neural signals are being processed, enabling emergence of cognitive phenomena [3–5].

There is one neuroscientific method that is particularly useful for studying brain functions, such as language or vision: the functional magnetic resonance imaging (fMRI). The popularity of fMRI can be explained by the fact that while being a non-invasive approach it still allows for studying neural processes with relatively high temporal resolution – in the scale of seconds – while providing an excel-

lent localization of the activity within the brain, i.e., with a millimeter-scale precision, even for the subcortical structures. If we compare fMRI to other non-invasive brain imaging methods, it provides better spatial resolution than electroencephalography (EEG) [6] (with EEG being superior in terms of temporal resolution), and the reliability of fMRI is relatively better described in scientific literature than some of the other approaches, such as functional near-infrared spectroscopy (fNIRS) [7] or functional photoacoustic computerized tomography (fPACT) [8].

Computational neuroscience also brings many challenges in the domain of high performance computing (HPC; see, e.g., [9, 10]). Recent works and tools developed by the leading laboratories around the world underline the importance of incorporating supercomputing methods to neuroscientific data analysis workflows [11–13]. With higher computing power, as well as increased memory capacity, new research perspectives are possible, such as genetic programming [14] or simulation and reconstruction of neural cir-

cuity [15]. Moreover, it is also easier and faster to perform some kinds of analyses, as the larger batches of data can be analyzed at once, without the need of conducting separate calculations for particular subjects or experimental sessions.

In the present study, the procedure of computational modelling of brain functions will be described as exemplified by fMRI tasks involving manual praxis skills [16]. Praxis is an ability to prepare and perform complex manual sequences in the communication and/or tool-related context, which is often linked (but not confined) to language skills [17, 18]. In order to provide reliable results, the fMRI procedure requires performing an experiment with at least ten to twenty participants, each being exposed to tens or hundreds of repetitions (trials) within a particular experimental condition, with pseudo-random intertwining occurrences of at least two such conditions [19].

The goal of this paper is to provide a technical insight into the computational aspects of fMRI and the specific cognitive domain of praxis skills. The example presented herein is an arbitrary choice based on the author's experiences and field of expertise. More specifically, the concept of fitting the general linear model (GLM) in fMRI data will be described here with the standard multi-trial fMRI dataset being restricted to only one experimental trial, for a single participant, from one of the study runs that were performed for this specific participant.

Importantly, similar descriptions of fMRI modelling, usually found in handbooks, various course materials, and in the web, involve modelling the entire experimental run [20, 21]. Presenting a model fit for a series of trials at once can be misleading for someone less familiar with particular types of experimental designs utilized in fMRI research or with the empirical methods in social sciences in general. Hence, simplifying the example to a single trial can provide a new perspective for students gaining knowledge on neuroscience methods and for specialists from other scientific domains, such as educational research and informatics, willing to get an insight into computational and statistical aspects of the method itself.

Here, the computational approach in a cognitive neuroscience study will be briefly illustrated by an example of a single trial, providing a relevant mathematical description as well as an appropriate visualization. Subsequently, the paper will elaborate on how the fitted models are used to compare different experimental conditions with each other. Such comparisons result in statistical parametric maps, containing information on the statistic calculated for all voxels (i.e., 3D pixels) in the brain, which are subsequently presented in a volumetric space or on the three-dimensional brain cortical surfaces. This provides the researches with the means to determine the exact location as well as the extent of the brain activity which occurred while performing a given cognitive task, such as grasp planning in praxis research, which is considered in this article. Finally, examples of such para-

metric maps obtained in an fMRI study on the task of planning functional grasps of bimanual tools will be presented and discussed [22].

## II. METHODS

The modelled hemodynamic response function (HRF) is typically fit for a series of trials from an fMRI experiment. The example described herein will be restricted to a single trial of a specific cognitive task, namely, planning functional grasps, and the data used here are from a study by Przybylski and Króliczak [16]. This single-trial data sample consists of 17 volumes of blood-oxygen-level-dependent (BOLD) fMRI signal, i.e., the signals were measured at 17 time points, evenly distributed over the span of 34 seconds (one volume was acquired for 2 seconds). The BOLD contrast is a natural phenomenon, which allows estimating the difference in blood-oxygenation level, based on the magnetic characteristics of the oxygenated and deoxygenated red blood cells. In other words, the assumption behind the fMRI is that the local increase in demand for oxygenated blood is related to the performance of a given cognitive task: at the moment in which this increased demand was observed.

### II. 1. Reference Signal Waveform

The signal used in the current example was acquired in one of the experimental runs for one of the participants taking part in the study. The motivation and complete results for this experiment are described in detail elsewhere (see [16]). The expected signal time course was pre-modeled with FMRIB Software Library (FSL<sup>1</sup> [23], Oxford, UK), and the modelled hemodynamic neural response was stored in a standard "design.mat" FSL file. This file usually contains a matrix of values of the modelled signal for all experimental conditions (e.g., the reference signal waveform), and in the case of the current single-condition example it is a single-column matrix (i.e., a vector of 17 values):  $[-8.74, 1.53, 3.56, 4.91, 3.00, 7.70, -5.93, -1.22, -1.42, -1.37, -1.22, -1.08, -9.90, -9.33, -9.04, -8.91, -8.85]$ .

### II. 2. General Linear Model (GLM)

In the considered case, a general linear model was used to fit the observed signal to the modelled (theoretical) time course of the brain response. In other words, first an assumption is made about how the brain should respond to a given task (a reference waveform), and then the observed signal is fitted (compared) to the pre-modelled time course. Here, for the purpose of fitting the model, a single voxel was selected from the whole volume of  $64 \times 64 \times 36$  voxels (around  $1.47 \cdot 10^5$  points in the stereotactic space). The signal time course was extracted from this voxel and the values obtained that way were used as a vector of a response variable in a re-

<sup>1</sup> Source: <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>

gression model (red line in the top panel of Fig. 1) with a predictor variable being the designed (theoretical) time course (blue line in the top panel of Fig. 1). The parameter estimate calculated for particular conditions is the slope of the regression line, i.e., the beta weight of the regression, representing the fit to the expected signal time course (see Eq. (1); and the blue regression line in the bottom panel of Fig. 1). Usually, in the case of fMRI data analysis there are multiple parameters being fit, one per each experimental condition. Here only one parameter, i.e., a single experimental condition, was used for the sake of simplicity of the example. Such a single parameter fit is a special case of the more general linear model which is described mathematically in Eq. (1):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i, \quad (1)$$

where:

- $Y_i$  is a response variable, i.e., the values of the observed fMRI signal,
- $X_{i1}, \dots, X_{i,p-1}$  are explanatory variables, which in this case are the expected brain responses to the experimental conditions,
- $\beta_0, \beta_1, \dots, \beta_{p-1}$  are regression parameters to be fit,
- $\varepsilon_i$  is an error of the model fit,
- $p$  is a number of explanatory variables (conditions),
- $i = 1, \dots, n$  is an index of the observation (number of a brain signal volume).

Although the location of this best-fit voxel in the analyzed fMRI data sample was known prior to the analysis described in this article, as it was calculated beforehand with the FSL software (similarly to the research conducted by Przybylski and Króliczak [16]), the signal from this particular voxel was re-analyzed in order to present the idea of fitting the modelled time course to the real fMRI signal. The location of the analyzed voxel in the 3D space of the brain was:  $x = 19, y = 43, \text{ and } z = 25$ . The time course for this particular voxel for the 17 volumes of fMRI BOLD activity is presented in Fig. 1, top graph, as the red line.

One of the most popular goodness-of-fit measures for the general linear model is the coefficient of determination, often denoted as  $R^2$  (R-squared). This measure is based on the ratio of the sum of squares of residuals to the total sum of squares. The better the model predicts the observed values, given some features (determinants), the higher the coefficient is, up to the value of 1.0, which means a very good fit. If the model is constant (its beta weights are equal to 0.0), and it always perfectly predicts the values, then the R-squared is 0.0, i.e., the model does not explain the variability in the observed data (see, e.g., the [Scikit-learn documentation](#)<sup>2</sup>).

### II. 3. Subtraction Contrasts

Calculating the model fit for all voxels in a series of fMRI signal volumes is just the first stage of the complete

computational workflow. The subsequent step is to compare different conditions to each other, which is performed with one of the two statistical approaches, either univariate (see, e.g., [16, 24]) or multivariate analysis (e.g., [25–27]). This article will focus on the former, i.e., the univariate analysis, which utilizes the method of a cognitive subtraction.

The cognitive subtraction procedure refers to comparing different experimental conditions to each other, e.g., comparing the signal related to a performance of a given cognitive task vs. the *resting* state condition, i.e., lying idle in the scanner. In terms of calculations to be performed, the subtraction comes down to executing mathematically defined procedures on the beta weights obtained for the modelled conditions, hence the statistical term for the product of this procedure: contrast of parameter estimates (COPEs).

However, in practice, such contrasts are not just a simple arithmetic subtraction of the parameters averaged across experimental runs, but a more advanced statistical operation has to be performed in order to compare the conditions in a reliable manner. Most often, a t-test is performed to validate a null hypothesis of no difference between two conditions, as represented in the formula below (Eq. (2)):

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}, \quad (2)$$

where the denominator is

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}, \quad (3)$$

and the numerators in Eq. (3) ( $s_p^2$ ) refer to a pooled variance

$$s_p^2 = \frac{\sum (X_1 - \bar{X}_1)^2 + \sum (X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}. \quad (4)$$

Additionally, the suffix “1” indicates a signal acquired for the first condition, and “2” stands for the data from the second condition. The “n” character represents images (signal volumes) collected for each of these conditions. For a detailed description of the subtraction procedure see [13] and [14]; equations presented herein are adapted from the [FSL documentation](#)<sup>3</sup>.

### II. 4. Repeated Measures Analysis of Variance (rmANOVA)

A comparison between three or more conditions requires a slightly different statistical approach. Namely, instead of performing a simple contrast (comparing just two conditions), a repeated-measures analysis of variance (rmANOVA) is performed with the additional post-hoc tests between the pairs of all considered conditions. This way,

<sup>2</sup> Source: [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LinearRegression.html#sklearn.linear\\_model.LinearRegression.score](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html#sklearn.linear_model.LinearRegression.score)

<sup>3</sup> Source: [https://users.fmrib.ox.ac.uk/~stuart/thesis/chapter\\_6/section6\\_3.html](https://users.fmrib.ox.ac.uk/~stuart/thesis/chapter_6/section6_3.html)

three or more experimental factors can be accounted for in a single analysis. Consider the following example from a slightly different domain: assessing treatment efficacy (e.g., as a part of drug testing research). In such case, rmANOVA conditions could be, e.g., signal values acquired: 1) before the treatment, 2) shortly after the treatment (e.g., a day or two), and 3) some longer time after the second measurement was taken (e.g., weeks, or even months after the treatment). As a matter of fact, this statistical procedure is also often utilized in longitudinal studies.

Additionally, with statistical procedures such as a two-way ANOVA, also the potential interaction effect between

conditions can be revealed. For instance, there can be a combined effect of type of the treatment and patient group with some therapies being effective, e.g., only for patients without comorbidities. In such scenarios the following two conditions can be considered with two levels per condition each: the first condition 1) treatment type with levels “treatment A”, and “treatment B”; and the second condition 2) patient group with levels “without comorbidities” and “with comorbidities”. In the present study, this more advanced statistical approach based on the analysis of variance will be explained in terms of fMRI analyses regarding planning bimanual grasps of tools vs. planning grasps of unimanual

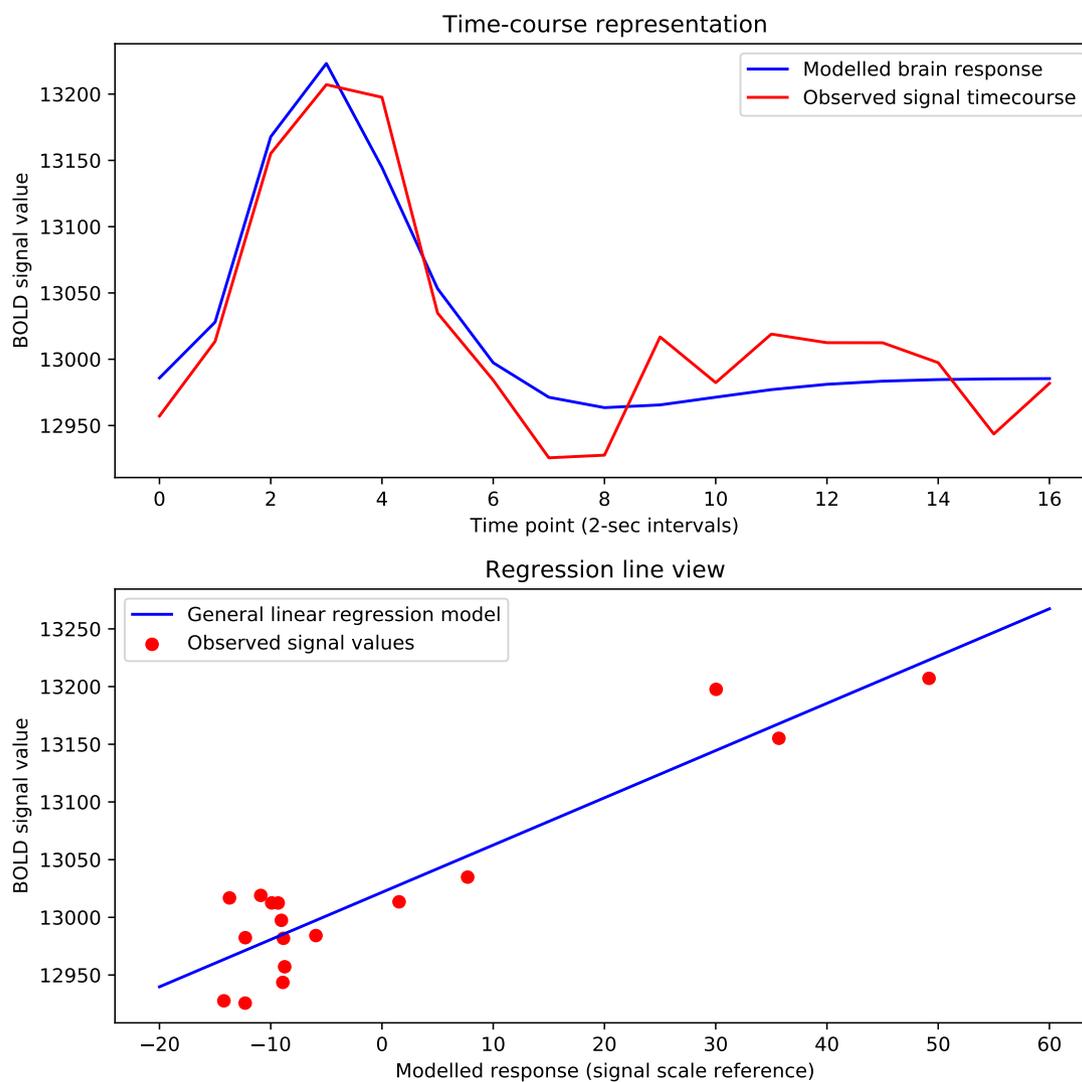


Fig. 1. Modeled and actual signal time courses and a general linear model fit for one voxel from a single trial of an experiment. In the top panel, an overlay of signal time courses is presented with the modelled signal time course (blue line plot) scaled to the actual observed fMRI signal values (red line). In the bottom panel, the modelled signal values (not scaled this time) were plotted on the abscissa ( $x$ -axis) against the actual observed signal values (ordinate,  $y$ -axis) as red dots. The straight blue line represents the regression model. Notice the linear trend for the data that was confirmed with the  $R^2$  statistic of 0.86. These results were obtained for a single voxel with the best fit within the analyzed series of fMRI volumes

tools; with conditions 1) object type – bimanual or unimanual objects, and 2) the hand to be used to perform the grasp – the right, or the left hand [22].

### III. RESULTS

The estimated parameter of the model fit calculated for a single trial of planning grasps of tools for the analyzed voxel was 4.095 (i.e., the beta weight) and the constant intercept of this model was 13 021.69. The coefficient of determination  $R^2$  for the calculated model was 0.86 (see Fig. 1). In other words, the model fit for the data from the analyzed voxel was quite good. The decent quality of the model fit was not surprising, given the fact that this exact voxel was revealed by a prior FSL analysis as the most crucial voxel for this condition (see the Methods section). However, it still could be the case that, despite being the best fit spatial location in an entire brain, the coefficient of determination for the model in this voxel would be, e.g., below 0.5. That would mean that in the whole brain volume there was not a single voxel with a satisfying fit for the analyzed trial. In other words, signal variability in no single location in the brain would have correlated with the analyzed experimental condition (at least for this participant in this specific trial). Nevertheless, it was not the case here, as the observed hemodynamic response was predicted by the model relatively good, as depicted graphically in Fig. 1.

If there were more conditions analyzed in this example, e.g., 1) planning functional grasps and 2) planning grasps of control objects, it would be possible to compare these two conditions with a t-test, hence performing a kind of a cognitive subtraction (see the description in the Methods section). The result of such comparison would have revealed whether or not there was a statistically significant difference between these conditions, thus indicating brain regions with significantly different activity for these two conditions. In fact, such comparison will be described here as exemplified by a different experiment, regarding planning bimanual grasps of tools [22].

An fMRI experiment on bimanual grasps of tools involved the following two conditions: 1) the to-be-used hand, i.e., the leading hand (right or left); and 2) manuality of the object (bimanual or unimanual) [22]. The inputs to this analysis were contrasts of tool-related conditions with non-tool reference objects: it was meant to account for the factor of the functionality of the object. The main research question in the study was whether the information related to the bimanual tools is processed within the praxis representation network in a different manner, as compared to signal fluctuations in this network for unimanual tools. Three separate rmANOVAs were performed, one for each of the phases of the interaction with tools: grasp planning, performing the grasps, and using tools. The results for the main effect of manuality in a form of a post-hoc test (a direct compari-

son) between bimanual-tool-related effects and unimanual-tool parameters are presented in Fig. 2.

The results presented in panel A of Fig. 2 show that during the phase of planning functional grasps the three activated clusters of voxels were located exclusively in the right hemisphere, more specifically in the superior parietal lobule, in somatosensory cortex, and in the motor cortices. Performing the functional grasps of bimanual tools, in addition to the activity observed during planning, also involved processing within the left hemisphere (sensorimotor cortex and superior parietal lobule), and in the dorsal parts of the occipitoparietal cortices within the right hemisphere. There were also some clusters of activity in the medial motor cortex bilaterally, and some processing also took place within the visual cortex, although it was restricted almost exclusively to the right hemisphere (see Fig. 2B). Bimanual tool use engaged exclusively the left hemisphere, with activity located in the somatosensory and early dorsal visual cortices (Fig. 2C). An overlay of all three stages of the interaction with tools is presented in Fig. 2D.

The more complex example of fMRI rmANOVA analysis described above shows that the processing pipeline representing the whole workflow can be quite complex, as not only the signal time course models have to be fit for all brain voxels, but also the resulting representations of experimental conditions have to be compared to each other using different statistical methods. Apart from the computational issues with the analysis, the interpretation of such outcomes requires knowledge of brain anatomy and the relation between brain networks and the functions they implement.

### IV. DISCUSSION

Studying human cognition with the functional magnetic resonance imaging is a computationally demanding task. In order to successfully perform model fitting, one has to process extensive amounts of data and that part is just the beginning of fMRI data analysis, with subtraction contrasts, adjustment for multiple comparisons, and group-level analyses being the subsequent stages.

The computational challenge in fMRI research can be explained in terms of the following example: a standard fMRI image is a multi-dimensional array of a usual size of approximately  $64 \times 64 \times 36$  voxels (i.e., spatial cubes,  $2 \text{ mm}^3$  each). The exact size of a single fMRI signal volume depends on the specific settings of the magnetic field of view during the procedure, and scanner's spatial resolution (its accuracy). An fMRI experimental run is a series of approximately 150 of such signal volumes: i.e., images (scans or frames), each of which is acquired over the time period of 2 seconds (the time interval for the acquisition of a single volume is referred to as a repetition time, or time-to-repetition, TR). Therefore, it takes roughly 5 minutes to perform a single experimental run, which, from the point of

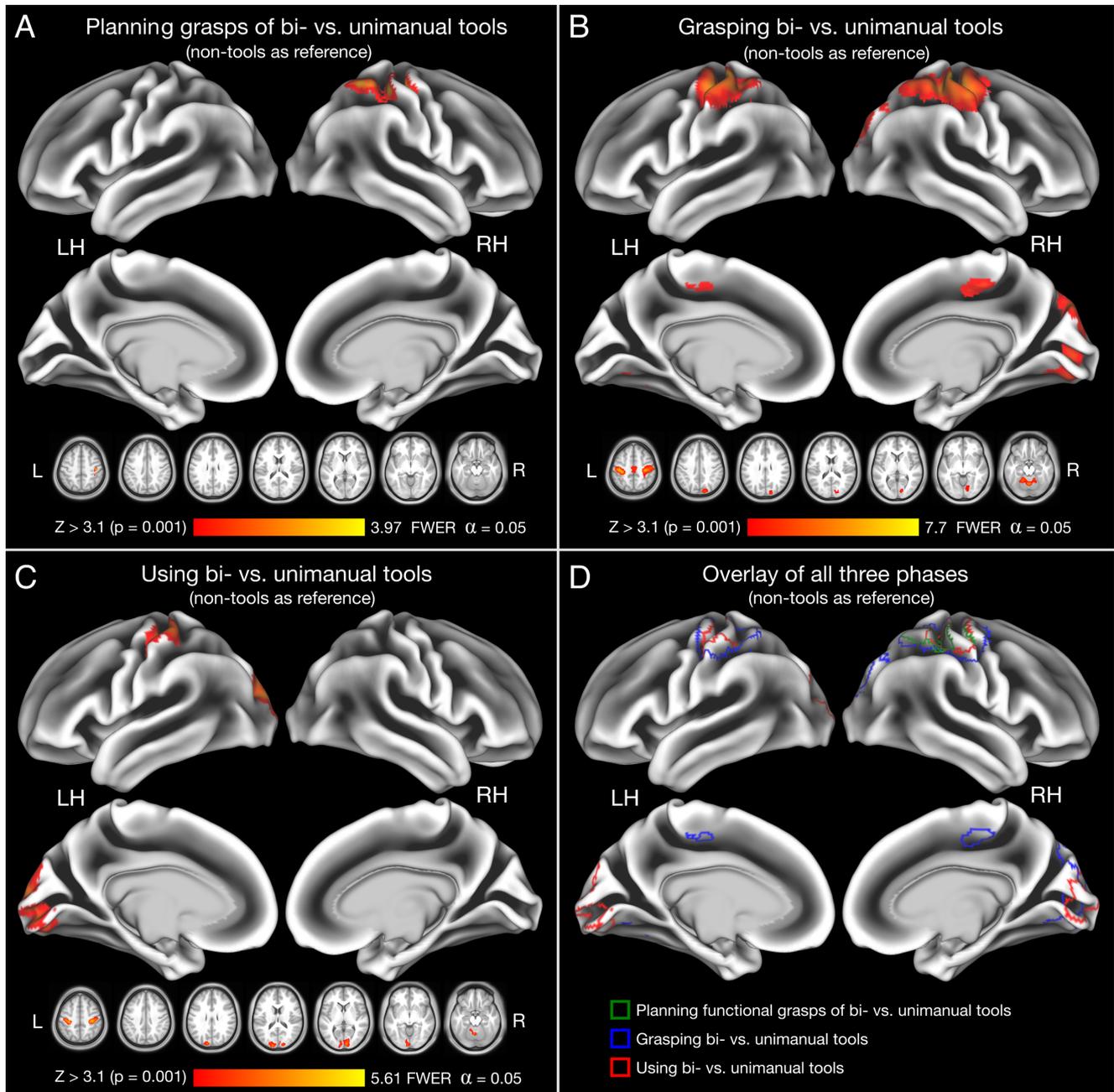


Fig. 2. A post-hoc comparison for the main effect of tool type (bimanual vs. unimanual) from an rmANOVA, with control objects (non-tools) as reference. There are three phases of interaction with tools presented herein: planning functional grasps – a grasp preparation stage (A), performing the grasps of the objects (B), and the tool-use phase (C). Panel D presents an overlay of the results from phases depicted in panels (A), (B), and (C). The results are mapped to partially inflated (midthickness, lateral and medial views) brain surfaces, as well as 7 brain slices across the axial plane. Color maps and bars represent standardized ( $Z$ -scored)  $t$  statistics from the follow up (post-hoc) direct comparisons between the bimanual and unimanual tool conditions, thresholded above the 3.1  $Z$  value. Family-wise error rate (FWER) was calculated for an alpha ( $\alpha$ ) level of 0.05, i.e., a correction for the number of clusters. The figure is adapted from [22]

view of the design of the experiment, is a series of trials, also called an order of trials. Most probably, each participant will undergo not just one but at least two such experimental runs, because splitting the excessive number of trials into separate runs is a better practice from the ethical as well as technical

point of view. Moreover, separating sets of trials into separate runs can not only provide more reliable results, but is also more convenient for both the researcher, and for the participant. If the experimental plan assumes performing a study with as many as 200 participants with two runs per a partici-

part, then the number of values to be analyzed grows significantly. To sum up this example so far, by performing such fMRI experiment we obtain the information representing the particular measurements that were taken, and these data can be conceptualized as a six-dimensional array with the following information stored in particular dimensions: 1) width of a scan  $\times$  2) its length  $\times$  3) height  $\times$  4) number of scans in a single run  $\times$  5) number of participants  $\times$  6) number of experimental runs. Based on this example, we can estimate that in such a study we would have to potentially store and analyze  $64 \times 64 \times 36 \times 150 \times 200 \times 2 = 8\,847\,360\,000$  signal values (i.e., over  $8.847 \cdot 10^9$  data points). In addition to that, there is also an  $150 \times 200 \times 2 \times 4$  array containing information about the modeled time course for each person's run from two sessions, for, e.g., 4 modeled experimental conditions (i.e., the number of scans/volumes  $\times$  number of participants  $\times$  number of experimental runs  $\times$  number of conditions). In terms of computations to be performed, the statistical procedure requires fitting as many as  $64 \times 64 \times 36 \times 200 \times 2 = 58\,982\,400$  (over  $5.898 \cdot 10^7$ ) 4-parameter models with 150 values in each of these models representing the discrete moments in the time during which the data were acquired.

Because constructing such arrays for the purpose of performing calculations is virtually impossible using a single work station or a laptop, the common practice is to divide the calculation into separate processing procedures, e.g., analyze data for a single participant at a time, using dedicated software, such as FSL [23], or SPM [28] packages. In the subsequent step of the analysis, the results for particular participants are averaged in order to reveal a group-level effect.

The only informatics improvement for the standardization and parallelization of the fMRI analysis workflow are the tools for describing, combining, and running particular processing pipelines, e.g., nipype [29]. In such tools, different stages of the analysis workflow can be represented as the procedures with a convenient graphical visualization of the whole analytical process. Therefore, it is not surprising that most of the neuroscience PhD programs around the world require the candidates to demonstrate a certain level of proficiency in programming languages, such as Python, and/or statistical computing frameworks, e.g., MATLAB, or R.

What is desperately needed in the field of neuroscience is strict, standardized protocols of data acquisition and analysis that a user who is less familiar with programming methods can comprehend and perform. Having such protocols implemented in a form of a coherent system – a platform and/or an analytical framework – can greatly facilitate the usage of the already available, advanced tools for defining and running the processing workflows [29, 30]. Interestingly, the tools that are available now, in order to define even the simplest automated pipeline, require the user to have basic knowledge on both of the commonly utilized programming approaches: object-oriented, as well as functional programming. Due to that complexity and diversity of the required skills, the learn-

ing curve for the modern fMRI data analysis frameworks is very steep. With an easy-to-use, unified neuroscientific statistical framework not only the results could be obtained easier and with less effort, but also the outcomes from different laboratories could be promptly compared to each other, additionally allowing for a replication and/or modification of the analytical procedures that were originally applied to the data acquired in a single laboratory. At the moment, due to the plethora of software methods and statistical frameworks used across the world, it is so extremely complex and time-consuming to exactly replicate even a single fMRI study that practically no one is making the effort to do so [31].

## V. SUMMARY AND CONCLUSIONS

Analyzing fMRI data is a computationally demanding task which requires utilizing a complex, domain-specific information processing workflow. Neuroscientific data-related procedures can be standardized and unified by creating a single, common statistical framework. While computational methods, such as fitting a general linear model or hypothesis testing, are a cornerstone of cognitive neuroscience, the ever-increasing amount of the acquired data brings new challenges in the field of informatics. With the availability of high performance computing resources, increased memory and storage, utilizing novel computational approaches in neuroscience is now more feasible than ever before. Moreover, in order to efficiently apply these quantitative methods to fMRI data, also a convenient user interface is needed for specialists who are first and foremost experts in their non-technical domains of knowledge, i.e.: neuropsychology, medicine, or biology. By providing a link between the most advanced informatics and mathematics techniques on the one hand, and cognitive science, psychology, and medicine on the other, the idea of an interdisciplinary approach to the science as a holistic human activity can be successfully realized.

## Acknowledgment

Part of the work described in this article was carried out when the author was a PhD student at the Faculty of Psychology and Cognitive Science of The Adam Mickiewicz University in Poznań, Poland. The main text of the manuscript was prepared and edited by the author in November, 2021. Almost all of the data from the praxis-related exemplary studies described in this paper were funded with the Polish National Science Center Maestro grant no. 2011/02/A/HS6/00174 awarded to Professor G. Króliczak from The Adam Mickiewicz University in Poznań, Poland. The author would like to thank Professor Króliczak for his expertise and mentorship when the author was a PhD student at The AMU, and for the opportunity to conduct the unique research with the most advanced equipment used in cognitive science studies. The author would

also like to thank dr C. Mazurek, head of the Poznan Supercomputing and Networking Center, for the invitation to write this paper, which motivated the author and gave him a chance to gather and organize his notes on the computational aspects of the fMRI research.

### Additional materials

- Supplementary material: a “Supplementary materials – fMRI signal modelling on the example of praxis tasks.zip” file, published alongside this article, containing a HTML report from the Jupyter IPython Notebook which was used to perform the model fit and generate Fig. 1. These materials are also available at the author’s GitHub repository: [link](#)<sup>4</sup>.
- Basics of fMRI Analysis: Preprocessing, First Level Analysis, and Group Analysis; MGH, Harvard Medical School: [link](#)<sup>5</sup>.
- Convolution with the hemodynamic response function: [link](#)<sup>6</sup>.
- Two-way ANOVA example: [link](#)<sup>7</sup>.
- Functional MRI techniques and applications: [link](#)<sup>8</sup>.
- fMRI subtraction technique explained: [link](#)<sup>9</sup>.
- Multiple Linear Regression – notes from the Yale’s course: [link](#)<sup>10</sup>.

### References

- [1] S. Ogawa, T.M. Lee, A.R. Kay, D.W. Tank, *Brain magnetic resonance imaging with contrast dependent on blood oxygenation*, Proc. Natl. Acad. Sci. U. S. A. **87**(24), 9868–9872 (1990).
- [2] M.A. Lindquist, J. Meng Loh, L.Y. Atlas, T.D. Wager, *Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling*, Neuroimage **45**(1), S187–S198 (2009).
- [3] M.F. Glasser, S.M. Smith, D.S. Marcus, J.L.R. Andersson, E.J. Auerbach, T.E.J. Behrens, T.S. Coalson, M.P. Harms, M. Jenkinson, S. Moeller, E.C. Robinson, S.N. Sotiropoulos, J. Xu, E. Yacoub, K. Ugurbil, D.C. Van Essen, *The Human Connectome Project’s neuroimaging approach*, Nat. Neurosci. **19**(9), 1175–1187 (2016).
- [4] D.S. Marcus, J. Harwell, T. Olsen, M. Hodge, M.F. Glasser, F. Prior, M. Jenkinson, T. Laumann, S.W. Curtiss, D.C. Van Essen, *Informatics and Data Mining Tools and Strategies for the Human Connectome Project*, Front. Neuroinform. **5**, 1–12 (2011).
- [5] L.E. Suárez, R.D. Markello, R.F. Betzel, B. Misic, *Linking Structure and Function in Macroscale Brain Networks*, Trends Cogn. Sci. **24**(4), 302–315 (2020).
- [6] A. Gramfort, M. Luessi, E. Larson, D.A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, M. Hämäläinen, *MEG and EEG data analysis with MNE-Python*, Front. Neurosci. **7**, 1–13 (2013).
- [7] S. Tak, J.C. Ye, *Statistical analysis of fNIRS data: A comprehensive review*, Neuroimage **85**, 72–91 (2014).
- [8] S. Na, J.J. Russin, L. Lin, X. Yuan, P. Hu, K.B. Jann, L. Yan, K. Maslov, J. Shi, D.J. Wang, C.Y. Liu, L.V. Wang, *Massively parallel functional photoacoustic computed tomography of the human brain*, Nat. Biomed. Eng. (2021).
- [9] K.E. Bouchard, J.B. Aimone, M. Chun, T. Dean, M. Denker, M. Diesmann, D.D. Donofrio, L.M. Frank, N. Kasthuri, C. Koch, O. Ruebel, H.D. Simon, F.T. Sommer, Prabhat, *High-Performance Computing in Neuroscience for Data-Driven Discovery, Integration, and Dissemination*, Neuron **92**(3), 628–631 (2016).
- [10] A.S. Shatil, S. Younas, H. Pourreza, C.R. Figley, *Heads in the Cloud: A Primer on Neuroimaging Applications of High Performance Computing*, Magn. Reson. Insights **8**, MRI.S23558 (2015).
- [11] K.E. Bouchard, J.B. Aimone, M. Chun, T. Dean, M. Denker, M. Diesmann, D.D. Donofrio, L.M. Frank, N. Kasthuri, C. Koch, O. Rübél, H.D. Simon, F.T. Sommer, Prabhat, *International Neuroscience Initiatives through the Lens of High-Performance Computing*, Computer (Long. Beach. Calif.) **51**(4), 50–59 (2018).
- [12] M. Kumar, C.T. Ellis, Q. Lu, H. Zhang, M. Capotă, T.L. Willke, P.J. Ramadge, N.B. Turk-Browne, K.A. Norman, *BrainIAK tutorials: User-friendly learning materials for advanced fMRI analysis*, PLoS Comput. Biol. **16**(1), e1007549 (2020).
- [13] K. Amunts, T. Lippert, *Brain research challenges supercomputing*, Science **374**(6571), 1054–1055 (2021).
- [14] A. Tahmassebi, A.H. Gandomi, A. Meyer-Bäse, *High Performance GP-Based Approach for fMRI Big Data Classification*, Proc. Pract. Exp. Adv. Res. Comput. **2017** Sustain. Success Impact – PEARC17, 1–4 (2017).
- [15] H. Markram, E. Muller, S. Ramaswamy et al., *Reconstruction and Simulation of Neocortical Microcircuitry*, Cell **163**(2), 456–492 (2015).
- [16] Ł. Przybylski, G. Kroliczak, *Planning Functional Grasps of Simple Tools Invokes the Hand-independent Praxis Representation Network: An fMRI Study*, J. Int. Neuropsychol. Soc. **23**(2), 108–120 (2017).
- [17] G. Kroliczak, M. Buchwald, P. Kleka, M. Klichowski, W. Potok, A.M. Nowik, J. Randerath, B.J. Piper, *Manual praxis and language-production networks, and their links to handedness*, Cortex **140**, 110–127 (2021).
- [18] S.H. Johnson-Frey, R. Newman-Norlund, S.T. Grafton, *A distributed left hemisphere network active during planning of everyday tool use skills*, Cereb. Cortex **15**(6), 681–695 (2005).
- [19] G.A. Orban, S. Ferri, *fMRI Techniques and Protocols*, **41**(1) (2009).
- [20] J.H. Zar, *Biostatistical Analysis: Data transformation*, Pearson Prentice Hall, USA (1996).
- [21] R.A. Johnson, I. Miller, J.E. Freund, *Probability and statistics for engineers*, Pearson Education London (2000).
- [22] M. Buchwald, *Neural substrates underlying planning interactions with bimanual tools: a functional magnetic resonance imaging study*, PhD Thesis, Adam Mickiewicz University in Poznań, Poland (2021).

<sup>4</sup> Source: <https://github.com/mikbuch/fmri-data-analysis>

<sup>5</sup> Source: <https://ftp.nmr.mgh.harvard.edu/pub/docs/SavoyfMRI2014/fmri.april2011.pdf>

<sup>6</sup> Source: [http://www.jarrodmillman.com/rcsds/lectures/convolution\\_background.html](http://www.jarrodmillman.com/rcsds/lectures/convolution_background.html)

<sup>7</sup> Source: <https://tutorials.methodsconsultants.com/posts/two-way-anova/>

<sup>8</sup> Source: [https://www.ctsi.ucla.edu/education/files/download/functional-mri-bookheimer.pdf?file\\_id=6661](https://www.ctsi.ucla.edu/education/files/download/functional-mri-bookheimer.pdf?file_id=6661)

<sup>9</sup> Source: [https://users.fmrib.ox.ac.uk/~stuart/thesis/chapter\\_6/section6\\_3.html](https://users.fmrib.ox.ac.uk/~stuart/thesis/chapter_6/section6_3.html)

<sup>10</sup> Source: <http://www.stat.yale.edu/Courses/1997-98/101/linmult.htm>

- [23] M. Jenkinson, C.F. Beckmann, T.E.J. Behrens, M.W. Woolrich, S.M. Smith, *FSL*, *Neuroimage* **62**(2), 782–790 (2012).
- [24] M.W. Woolrich, B.D. Ripley, M. Brady, S.M. Smith, *Temporal autocorrelation in univariate linear modeling of FMRI data*, *Neuroimage* **14**(6), 1370–1386 (2001).
- [25] M. Buchwald, Ł. Przybylski, G. Kroliczak, *Decoding Brain States for Planning Functional Grasps of Tools: A Functional Magnetic Resonance Imaging Multivoxel Pattern Analysis Study*, *J. Int. Neuropsychol. Soc.* **24**(10), 1013–1025 (2018).
- [26] S. Anzellotti, M.N. Coutanche, *Beyond Functional Connectivity: Investigating Networks of Multivariate Representations*, *Trends Cogn. Sci.* **22**(3), 258–269 (2018).
- [27] J.V. Haxby, A.C. Connolly, J.S. Guntupalli, *Decoding Neural Representational Spaces Using Multivariate Pattern Analysis*, *Annu. Rev. Neurosci.* **37**(1), 435–456 (2014).
- [28] W.D. Penny, K.J. Friston, J.T. Ashburner, S.J. Kiebel, T.E. Nichols, *Statistical parametric mapping: the analysis of functional brain images*, Elsevier (2011).
- [29] K. Gorgolewski, C.D. Burns, C. Madison, D. Clark, Y.O. Halchenko, M.L. Waskom, S.S. Ghosh, *Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python*, *Front. Neuroinform.* **5** (2011).
- [30] A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion, G. Varoquaux, *Machine Learning for Neuroimaging with Scikit-Learn*, *Front. Neuroinform.* **8**(14), 1–15 (2014).
- [31] R. Botvinik-Nezer, F. Holzmeister, C.F. Camerer et al., *Variability in the analysis of a single neuroimaging dataset by many teams*, *Nature* **582**(7810), 84–88 (2020).



**Mikołaj Buchwald** is a cognitive neuroscience specialist and psychophysiology researcher at Poznan Supercomputing and Networking Center, Polish Academy of Sciences. He received his doctoral degree in Cognitive Science at The Adam Mickiewicz University in Poznań in 2021. Since 2018 he has been affiliated with Poznan Supercomputing and Networking Center, where he is involved in R&D projects related to the development of information and communication systems for data analysis and visualization. He has also been actively participating in application of technologies and algorithms in psychological and medical sciences scenarios, as well as in commercial user experience research. He has authored and co-authored research papers in international, SCOPUS-ranked journals covering domains such as neuropsychology, human-computer interaction, signal analysis, biomedical data science, and higher education.