Haemodynamic response modelling using General Linear Model

Antea Ceko, Chiara Billi, Irene Vardabasso, Ludek Cizinsky and Lou Houngbedji

October 2023

1 Introduction

In this project we aim at modelling the haemodynamic response of the brain when the area of the cortex functionally correlated to face recognition is stimulated. Naturally, when exposing the subject to an image of a face, we expect this brain region to temporarily increase its activity. Consequently, this should lead to a localized increase in oxygen consumption, that would characterize the expected haemodynamic response. To assess this, we separated our project into 2 parts. The first part consists of generating synthetic data resembling the response to a sequence of visual stimulations. In order to obtain this, we set up the events sequence, and convolute it with the haemodynamic response function (HRF). We then repeat these steps to select the HRF parameters in a comparative fashion and finally, we add noise. In the second part, we then model the generated data using a General Linear Model (GLM) through the definition of a design matrix. Finally, we assess the significance of the regressors via statistical analyses.

2 Methods

Simulation of stimuli events. For this part, we create 2 different sequences, in an ordered event-related manner, where each sequence includes 2 phases: rest phase and image (face or object) exposition. We decided to use an event-related design experiment because the time between the stimuli of the faces and objects is considerably longer than the duration of the event itself. We made this decision so we could be sure the model had time to fit the data. We are aware that for event-related experiments a randomly-designed onset sequence of the stimuli is the standard, however, we believe that it is not necessary to design our experiment in such a way, given the scope of this project. We set the total duration of the experiment to be 350s, which allowed us to observe 5 trials for every stimulus. Each of the trials lasted for 2s and was repeated every 60s. To introduce variability, the amplitude of the individual stimuli was randomly selected from a predetermined sampling range.

Generation of the BOLD signal. In the brain, the response to the exposition of the stimuli sequences takes the form of haemodynamic responses. In order to mimic this biological process, we compute the Blood-Oxygen-Level-Dependent (BOLD) response given by the exposition to face and to object events by convoluting each sequence to the HRF function. The suggested HRF was based on the double-gamma definition.

$$hrf(t) = A\left(\frac{t^{a_1-1}e^{-t}}{\Gamma(a_1)} - c\frac{t^{a_2-1}e^{-t}}{\Gamma(a_2)}\right)$$

Adjustment of HRF parameters. We plot our BOLD signals with different sets of parameters (a_1, a_2) to see for which ones the curve reproduces a more physiologically plausible shape of the HRF, i.e., the initial dip, primary response, and then negative undershoot. This requirement was satisfied for $a_1 = 8$ and $a_2 = 16$.

Addition of noise. Real-life signals acquired with fMRI scans are usually noisy, because of the many interferences caused by other working areas of the brain, motion artefacts, etc. To simulate a realistic observed activity of a voxel of interest, we thus need to add to the BOLD signal a Gaussian noise with a standard deviation of 0.15.

Model fitting. To approximate the theoretical GLM model $Y = X\beta + \epsilon$, we first need to define our design matrix X. The design matrix consists of 3 regressors. The first two regressors correspond to the BOLD signal from the face and object visual stimuli and the last one represents a bias term which is independent of the input data. Our ground truth was given by the synthetic fMRI signal y, which we generated and enriched with noise as described above.

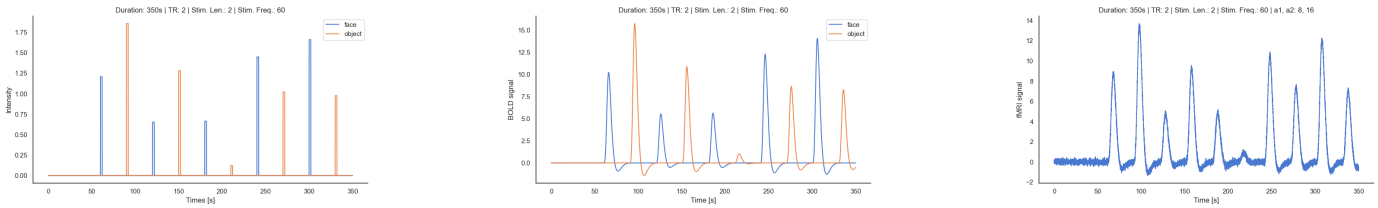
Finally, to find the best $\hat{\beta}$ parameters we used the Ordinary Least Squares (OLS) optimization method. To assess model fit, we made a residual plot as well as a histogram of residuals. To test the significance of the

computed regressors, we decided to compute the one-sided p-value of our model using a t-test, with a threshold set at 0.05. Finally, we showed a graphical representation of the design matrix to visualise the contrast between the two conditions.

BONUS: repetition of the model fitting on openneuroi data To test our methods on real-life data, we loaded an experimental setting from the OpenfMRI project (run 08 on subject 2), consisting of a block-design sequence of visual stimuli. In particular, 8 different objects' images were shown in blocks of 12 repetitions lasting 0.5s each. We then generated a BOLD signal by convolution of the events sequence with an HRF function (after parameters' selection), and we added Gaussian noise (with 0.15 standard deviation). We then fitted a GLM, with a design matrix X consisting this time of 9 columns (one constant and the others for each different type of visual stimulation). For the rest of the steps, we followed the exact same procedure as described in the previous section.

3 Results

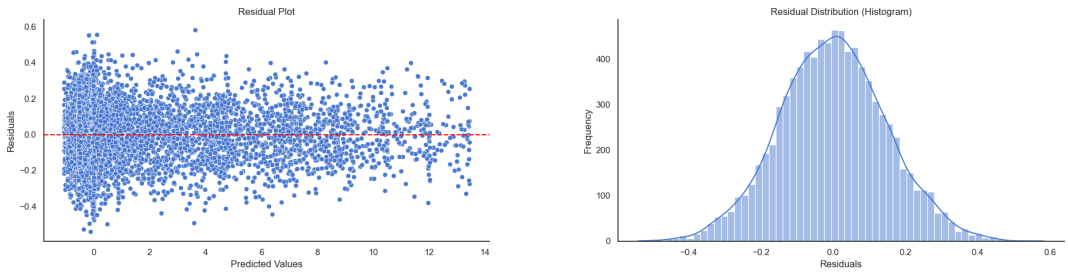
Stimuli. *Figure 1* shows the graphical representation of the stimuli sequences. On *Figure 1.a*, we can see the 2 sequences of presentation of face and object. In the *Figure 1.b*, we can see the BOLD function of both of our signals. We observed that they have a physiologically plausible form of haemodynamic response. Last but not least, the fMRI signal with the added Gaussian noise is shown on *Figure 1.c*.



(a) Stimuli intensity (b) BOLD signal (c) fMRI signal

Figure 1: Visualization of the stimuli

Residual plot and distribution. *Figure 2.a* represents the predicted values against the residuals. By doing so we are able to verify that the residuals are randomly scattered (and not dependent on the estimates in any systematic way). In addition, *Figure 2.b* shows the histogram of the residuals. The plot of the distribution allowed us to verify that they followed a normal distribution, with a standard deviation of ≈ 0.15. We expected the distribution of the residuals to show these properties since the ground truth of our GLM was deterministically assessed via convolution of the stimuli events with the HRF, thus the residuals must reflect the stochastic features introduced in our model by the addition of a Gaussian noise. Finally, we conducted a statistical significance test for each of the regressors associated with a particular condition. Under our null hypothesis, we assume that the regressor is equal to zero, which would mean that it is not important for the prediction of the fMRI signal. For both of our regressors, we obtained a p-value which was less than our threshold, i.e., $\alpha = 0.05$. Therefore, in both cases, we reject the null hypothesis in favour of the alternative one, i.e., the parameter estimates are different from zero.

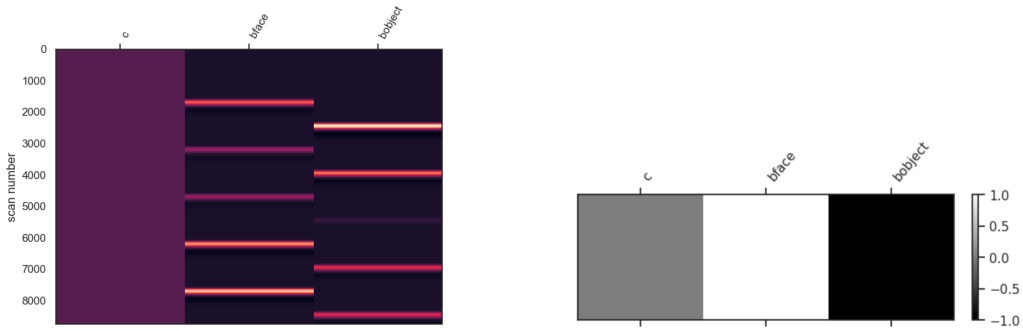


(a) Residual plot (b) Residual distribution

Figure 2: Residual plot and distribution

Design matrix. *Figure 3.a* showed the content of the design matrix. The purple colour represented the constant column. The stripped columns represented the BOLD signal of either face or object stimuli sequences. The light

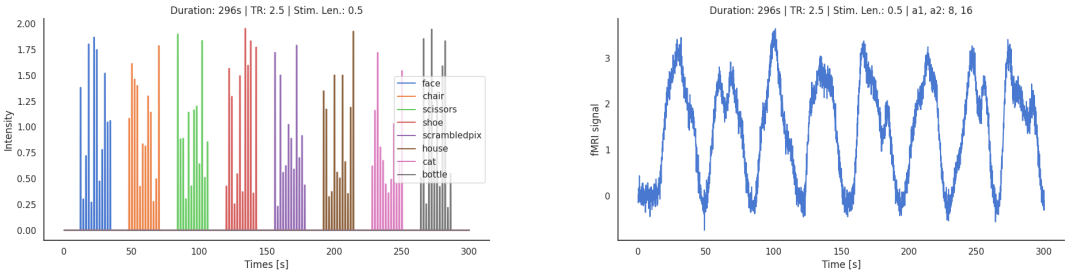
strips represent the stimuli events and the dark ones the rest events. In addition, *Figure 3.b* shows the contrast between the face and object stimuli condition.



(a) Design matrix (b) Contrast matrix

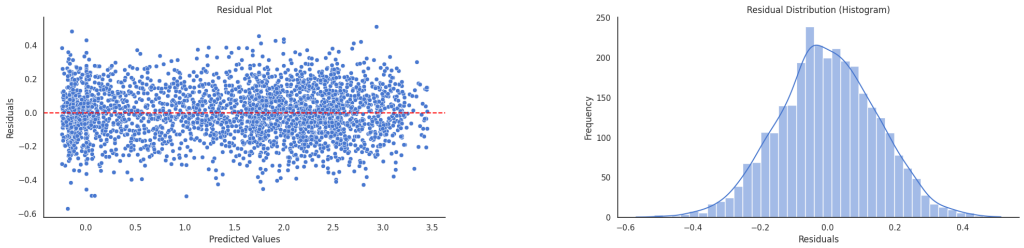
Figure 3: Visualization of the design matrix

Repetition on real-life sequence. Based on the real-life stimuli sequence shown in *Figure 4.a* and on a selection of HRF parameters that led us to the same choice of a_1 and a_2, we generated the BOLD signal enriched with Gaussian noise (*Figure 4.b*). Following model fitting, the reproducibility of our pipeline was finally assessed via statistical evaluation of our predictions, which once again proved to be reliable, since the residuals plot (*Figure 5.a*) and histogram (*Figure 5.b*) satisfy the requirements illustrated above in terms of independence from the values of the predictions, normal distribution and standard deviation.



(a) Real-life stimuli sequence (b) Simulated BOLD signal

Figure 4: Visualization of the design matrix



(a) Residuals plot (b) Residuals histogram

Figure 5: Visualization of the design matrix

4 Discussion

What happens when we vary the signal amplitude? How does the model fit change when the signal amplitudes are stronger and weaker? Increasing the amplitude of the signal, while keeping the noise at a fixed standard deviation, will cause the Signal to Noise Ratio (SNR) to increase, thus providing the GLM with less noisier ground truth and leading to better estimates. In fact, a signal with a small amplitude is more susceptible

to the noise and as a consequence, the artifact will be higher. For instance, for values of $A = [0.05, 0.5, 1, 8]$ the corresponding $R_2 = [0.372, 0.984, 0.996, 1]$ were computed.

What happens when we vary the noise? How does the amount of noise in the data impact our model fit? Following what we stated above, changing the standard deviation of the Gaussian noise will affect the SNR. In particular, a bigger standard deviation will lead to a smaller SNR, hence the worse estimates of the model parameters and overall fit of the model (e.g. indicated by R_2), and vice-versa. Increasing the noise makes it harder to distinguish the desired signal from the background noise and impacts statistical significance. For instance, for values of standard deviations $= [0.05, 0.5, 1, 3]$ the corresponding $R_2 = [1.000, 0.956, 0.839, 0.383]$ were computed.

What type of regressors can you add in the design matrix to control for low-frequency drift? How can you verify that these added regressors do not compete with other regressors in the design matrix? In order to control for low-frequency drift, we could add low-frequency basis vectors as regressors, for example, polynomial or sinusoidal, into the design matrix. To assess whether those regressors compete with the others, one should compute the correlation matrix PxP between regressors, to eventually rule out any correlation between them.

A common question in experimental design is determining the optimal number of trials since it has a direct effect on the likelihood that the design is able to detect a real effect. Which parameter of your model can allow you to determine the optimal number of trials? To determine the optimal number of trials (proportional to the duration of the experiment, if repetition time and sampling time are fixed), we have to take into account the fact that we want to obtain a Gaussian distribution of the data, or at least be able to approximate our data to that, which would be possible if our Nt allows for the law of large numbers to apply. At the same time, we don't aim at a large dataset since the data acquisition process is resource-consuming. This leads to a trade-off that needs to be evaluated based on the outcome.

What if one condition simply results in processes that systematically take longer than the other condition? What do you think would happen to the β estimates? The β estimates that we would obtain would be collinear, thus we would have not-independent regressors. That would be because a process that systematically takes longer would likely result in an overlapping of the HRF responses, thus not satisfying the assumption of independence we imposed on the individual trials that are necessary for the GLM to result in an estimation of independent β regressors.