PyHRF: A Python Library for the Analysis of fMRI Data Based on Local Estimation of the Hemodynamic Response Function
Jaime Arias, Philippe Ciuciu, Michel Dojat, Florence Forbes, Aina Frau-Pascual, Thomas Perret, Jan Warnking

To cite this version:
Jaime Arias, Philippe Ciuciu, Michel Dojat, Florence Forbes, Aina Frau-Pascual, et al.. PyHRF: A Python Library for the Analysis of fMRI Data Based on Local Estimation of the Hemodynamic Response Function. 16th Python in Science Conference (SciPy 2017), Jul 2017, Austin, TX, United States. pp.34-40, 10.25080/shinma-7f4c6e7-006 hal-01566457

HAL Id: hal-01566457
https://hal.archives-ouvertes.fr/hal-01566457
Submitted on 21 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
PyHRF: A Python Library for the Analysis of fMRI Data Based on Local Estimation of the Hemodynamic Response Function

Jaime Arias‡‡∗, Philippe Ciuciu‡†, Michel Dojat¶∥†, Florence Forbes‡†, Aina Frau-Pascual‡‡, Thomas Perret‡†, Jan M. Warnking¶∥†

Abstract—Functional Magnetic Resonance Imaging (fMRI) is a neuroimaging technique that allows the non-invasive study of brain function. It is based on the hemodynamic variations induced by changes in cerebral synaptic activity following sensory or cognitive stimulation. The measured signal depends on the variation of blood oxygenation level (BOLD signal) which is related to brain activity: a decrease in deoxyhemoglobin concentration induces an increase in BOLD signal. The BOLD signal is delayed with respect to changes in synaptic activity, which can be modeled as a convolution with the Hemodynamic Response Function (HRF) whose exact form is unknown and fluctuates with various parameters such as age, brain region or physiological conditions.

In this paper we present PyHRF, a software to analyze fMRI data using a Joint Detection-Estimation (JDE) approach. It jointly detects cortical activation and estimates the HRF. In contrast to existing tools, PyHRF estimates the HRF instead of considering it as a given constant in the entire brain. Here, we present an overview of the package and showcase its performance with a real case in order to demonstrate that PyHRF is a suitable tool for clinical applications.

Index Terms—BOLD response, fMRI, hemodynamic response function

Introduction

Neuroimaging techniques, such as functional Magnetic Resonance Imaging (fMRI), allow the in vivo study of brain function by measuring the cerebral response to sensory or cognitive stimulation. For more than 20 years, the Blood-Oxygen-Level-Dependent (BOLD) fMRI modality has being the technique most used by neuroscientists to map the main functional regions and their links in the healthy and diseased brain.

The BOLD signal [OLKT90] reflects the changes in deoxyhemoglobin concentration in the brain. Briefly, when brain activity increases, local oxygen consumption in brain tissue increases, slightly increasing the concentration of deoxyhemoglobin in blood (see Fig. 1). Subsequently, cerebral blood flow is strongly upregulated locally by arteriolar vasodilation to replenish the tissue, increasing local blood oxygen saturation in veins and capillaries above the initial level. Oxygenated and deoxygenated blood has different magnetic properties. As a result, the above causes a BOLD signal increase. Thus, the BOLD signal is an indirect measure of cerebral activity based on physiological changes in oxygen consumption, cerebral blood flow and cerebral blood volume.

Fig. 1: fMRI BOLD signal [OLKT90]. The BOLD signal measures the local changes in blood oxygenation occurring during brain activity.

fMRI data is acquired by repeated imaging of the brain while the subject or patient executes a task or receives a sensory stimulus during repeated epochs separated by periods of rest. This data is analyzed by correlating the measured time-varying BOLD signal in each image location with a predicted BOLD signal, obtained by convolving the known function representing the stimulus with a Hemodynamic Response Function (HRF) modeling the delay in the vascular response. Locations in the brain where this correlation is statistically significant are considered to exhibit a neuronal response to the task or stimulus, and thus to be involved in its cognitive processing.

BOLD fMRI is non-invasive, non-ionizing, and gives access in vivo to brain activity with a relatively high spatial resolution. However, it is highly dependent of the HRF of the brain. The BOLD signal does not give access to true physiological parameters such as cerebral blood flow or cerebral blood volume, but rather measures a mixture of these quantities that is difficult to untangle.

In this regard, BOLD is a very interesting tool in neuroscience, but in general it is not widely used for clinical applications because the impact of physiopathological situation on the HRF and the response amplitude are unknown, hampering the BOLD signal interpretation. For instance, the vascular response giving rise to the BOLD signal is altered in normal ageing [FGM+14]...
and pathologies like Alzheimer’s disease [CVM+11] or Stroke [AVT+14].

Most used open source libraries for the analysis of fMRI data (e.g., SPM¹, FSL²) consider the HRF as constant in all the brain and the same for all subjects. However, several works (see [BVC13] for a survey) show that the HRF changes across different regions of the brain and across individuals, increasing thus the possibility of obtaining false negatives and decreasing the reliability of the results. The software PyHRF [VBR+14] was developed to overcome the above limitation by analyzing fMRI data using a Joint Detection-Estimation (JDE) approach.

In the JDE approach, the detection of the cortical activation is achieved together with the estimation of the unknown HRF response by analyzing non-smoothed data. This detection-estimation is calculated on different parcels of interest pAVING the cerebral volume. Therefore, PyHRF allows to navigate throughout the brain and to focus on regions of interest during the experiment in order to visualize the activations and their temporal behavior through the estimated HRF. Over the last years, efforts have been made in terms of image processing, user-friendliness and usability of the PyHRF package to make it more easy to use by non experts and clinicians.

Next, we present the PyHRF package. Then we illustrate its use on real fMRI data. Finally, we conclude by discussing directions of current/future work. An online Jupyter notebook containing the results presented here can be found at http://www.pyhrf.org/scipy2017_notebook.

PyHRF

PyHRF (http://www.pyhrf.org) is an open source tool implemented in Python that allows to jointly detect activation and estimate (JDE) the hemodynamic response function (HRF) [MIV+08], which gives the temporal changes in the BOLD effect induced by brain activity. This estimation is not easy in a voxel-wise manner [PJM+03], and a spatial structure was added to JDE [VRC10] in order to get reliable estimates. In this regard, HRF estimation in JDE is parcel-wise and an input parcellation is required. However, the use of the Markov Chain Monte Carlo (MCMC) method for estimation added a huge computational load to the solution, leading to the development of a faster method to deal with the parameter estimation. Thus, a Variational Expectation Maximization (VEM) solution [CVF+13] was implemented.

JDE aims at improving activation detection by capturing the correct hemodynamics, since using the wrong HRF function could hide existing activations. The use of a canonical HRF is usually sufficient for activation detection. However, HRF functions have been found to have different shapes in different regions [HOM04], and to have different delays in specific populations [BVC13]. They change depending on pathologies such as stenosis.

Fig. 2 shows some HRF functions estimated by PyHRF from BOLD data on a healthy adult. This data was acquired in a block-design setting with visual, auditory and motor experimental conditions. The parcels correspond to regions of the brain that are known induce evoked activity in response to these experimental conditions. Observe that the HRF estimated in the visual and motor regions (first and third figure from left to right) are well approximated by the canonical HRF whereas in the auditory area (second figure from left to right), the recovered hemodynamic profiles peak earlier than the canonical shape.

Standard methods (e.g., GLM), with the posterior classical statistics applied, give Statistical Parametric Maps (SPM) that describe the significance of the activation in each region. JDE is a Bayesian approach and estimates, for each parameter, posterior probability functions. For this reason, we can compute Posterior Probability Maps (PPMs) from the output of PyHRF. These PPMs are not directly comparable to the classical SPM maps, but give a similar measure of significance of activation. For instance, in Fig. 4 we show the SPM and PPM maps for a visual experimental condition in the same data used for Fig. 2. We used the package Nilearn (http://nilearn.github.io) to generate the brain maps presented in this document.

In Fig. 3 we present the inputs and the outputs of PyHRF for the analysis of BOLD data. It needs as inputs the data volume (BOLD), the experimental paradigm, and a parcellation of the brain. After running the JDE algorithm, the output will consist of HRF functions per parcel, BOLD effect maps per experimental condition, and PPMs per condition. In the next section, we will describe in more details these elements and how to use PyHRF.

Fig. 2: HRF computed using PyHRF from BOLD data in several parcels belonging, respectively from left to right, to visual (yellow, dark blue and green parcels), auditory (cyan and light green parcels) and motor (red and purple parcels) regions.

1. SPM official website: http://www.fil.ion.ucl.ac.uk/spm/software/
2. FSL official website: https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/
Example of Use

To illustrate the use of PyHRF, we will describe the steps for the analysis of BOLD data. A Jupyter notebook containing the complete code is available at http://www.pyhrf.org/scipy2017_notebook.

Getting fMRI BOLD Data

First of all, we need to get our fMRI BOLD data. In our running example, we will analyze the dataset used in [GSB+13]. This dataset (ds000114) is open shared and it can be downloaded from https://openfmri.org/dataset/ds000114/. For convenience, we implemented the method get_from_openfmri that uses the library fetchopenfmri (https://github.com/wiheto/fetchopenfmri) to download datasets from the site openfmri. For instance, the following code downloads the dataset ds000114 to the folder ~/data.

```python
>>> dataset_path = get_from_openfmri('114', '~/data')
Dataset ds000114 already exists
/home/jariasal/data/openfmri/ds000114_R2.0.1
```

Briefly, in this dataset ten healthy subjects in their fifties were scanned twice using an identical experimental paradigm. This paradigm consists of five task-related fMRI time series: finger, foot and lip movement; overt verb generation; covert verb generation; overt word repetition; and landmark tasks. For the sake of simplicity, we will focus our analysis only on motor tasks (i.e., finger, foot and lips movement). Fig. 5 shows the paradigm containing only the three tasks mentioned above. As we can see in the experimental paradigm, tasks do not overlap each other and stimuli are presented to the subject during a certain time (i.e., block paradigm).

**Fig. 4:** A) PPM and B) SPM maps computed with JDE and GLM, respectively. Scale is logarithmic.

**Fig. 5:** Experimental paradigm of the dataset ds000114. We show only the motor tasks of the dataset (finger, foot and lips movement).
a BOLD volume (e.g., a whole brain) is usually not built at once but using a series of successively measured 2D slices. Each slice takes some time to acquire, so slices are observed at different time points, leading to suboptimal statistical analysis.

We used the library Nipype (https://github.com/nipy/nipype) to define and apply our preprocessing pipeline. This library allows to use robust tools, such as SPM and FSL, in an easy manner. The proposed workflow (see Fig. 6) starts by uncompressing the images since they are in a .nii.gz format (gunzip_func and gunzip_anat nodes). After, it applies a slice timing operation in order to make apparent that all voxels of the BOLD volume have been acquired at the same time. We then apply a realignment in order to correct for head movements. Moreover, we apply a coregistration operation in order to have the anatomical image (high spatial resolution) in the same space as the BOLD images. Finally, we normalize our images in order to transform them into a standard space (a template).

The pipeline described above was run for the images of all subjects from the dataset (i.e., 10 subjects) on multiple processors, since Nipype uses the library joblib (https://github.com/joblib/joblib). We used the acquisition parameters presented in [GSB+13] to parameterize each preprocessing operation. For instance, the number of slices for the volume, the time for acquiring all slices (TR), and the order in which they were acquired (e.g., interleaved).

In the following snippet, we show a portion of the code to define the slice timing task with Nipype.

```python
>>> import nipype.interfaces.spm as spm
>>> import nipype.pipeline.engine as npe

# Acquisition parameters
>>> TR = 2.5
>>> TA = TR - (TR / NUM_SLICES)
>>> REF_SLICE = 1

# interleaved slice order
>>> SLICE_ORDER = list(range(1, NUM_SLICES+1, 2) + range(2, NUM_SLICES+1, 2))

# slice timing with SPM
>>> slice_timing = npe.Node(spm.SliceTiming(num_slices=NUM_SLICES, time_repetition=TR, time_acquisition=TA, slice_order=SLICE_ORDER, ref_slice=REF_SLICE), name='slice_timing_node')
```

PyHRF Analysis (Inputs)

So far, we have prepared our functional and structural images for BOLD analysis. It is important to note that PyHRF receives non-smoothed images as input, thus we excluded this operation from our preprocessing pipeline.

For the sake of simplicity, in our running example we only analyze the 4th subject from our dataset. Additionally, we use the package Nilearn to load and visualize neuroimaging volumes. Fig. 7 shows the mean of the functional images of the 4th subject after preprocessing.

![Fig. 7: Mean of all preprocessed functional images (over time) of the 4th subject of the dataset ds000114.](image)

As we explained before, the JDE framework estimates HRF parcel-wise. This means that PyHRF needs a parcellation mask to perform the estimation-detection. The package provides a Willard atlas [RAM+15] (see Fig. 8) created from the files distributed by Stanford (http://findlab.stanford.edu/functional_ROIs.html). This atlas has a voxel resolution of 3x3x3 mm and a volume shape of 53x63x52 voxels.

![Fig. 8: Willard atlas [RAM+15].](image)

We used the method get_willard_mask to resize the original atlas to match the shape of the BOLD images to be
analyzed. In addition, this method saves the resampled mask in a specified path. For instance, Fig. 8 shows the Willard atlas resized to the shape of the functional image in Fig. 7. The following code illustrates how to resize the Willard atlas provided by PyHRF to match the shape of the image ~/data/bold.nii, and saves it in the folder ~/pyhrf.

```python
>>> willard = get_willard_mask('~/pyhrf', /home/jariasal/pyhrf/mask_parcellation/willard_3mm.nii

PyHRF also needs the experimental paradigm as input. It must be given as a csv file following the convention described in the documentation (https://pyhrf.github.io/manual/paradigm.html). For the sake of convenience, we used the method convert_to_pyhrf_csv to read the paradigm file provided by the dataset (a tsv file) and rewrite it using the PyHRF format. Since each dataset has its own format for the paradigm, we give it as an input to our method.

```python
>>> columns_tsv = ['onset', 'duration', 'weight', 'trial_type']
>>> paradigm = convert_to_pyhrf_csv('~/data/paradigm.tsv', 0, columns_tsv)
```

Table 1 shows the experimental paradigm of the dataset ds000114 written using the PyHRF format. Note that it only contains motor stimuli since we are only interested in them for our BOLD analysis.

**PyHRF Analysis (Run)**

Now we are ready to run our BOLD analysis with PyHRF. For that, we need to define some important parameters of the underlying JDE model (e.g., beta, hrf-hyperprior, sigma-h, drifts-type) and a folder to save the output (output).

Moreover, we need to specify if we want to estimate the HRF response or use, for example, its canonical form. In our running example, we will estimate the HRF (estimate-hrf) with a time resolution (dt) of 1.25 s, a duration (hrf-duration) of 25.0 s, and we force to zero the beginning and ending of the response (zero-constraint).

Once the parameters of the model have been defined, we run our analysis by using the command-line tool pyhrf_jde_vem_analysis provided by PyHRF. We can execute the analysis using several processors (parallel) because PyHRF uses the library joblib. The reader can find more details about this command and its parameters in the PyHRF documentation.

**PyHRF Analysis (Output)**

We show in Fig. 9 the active parcels (A), the PPMs (B), and the estimated HRFs (C) generated by PyHRF for the motor task Finger. Reading the description given in [GSB13], this task corresponds to finger tapping. Recall that PyHRF estimates a HRF for each parcel and generates a PPM for each condition.

We compared the output of PyHRF with the thresholded statistical maps shared on the site Neurovault (http://www.neurovault.org/images/307/) for the same dataset and same task (see Fig. 9). While the experimental paradigm is not optimized for JDE (standard block paradigm is not ideal to estimate different points of the HRF course), we obtained similar results to standard statistical analysis additionally providing the form of the HRF. As we can observe, at cut z=60 both results (Fig. 9 B and D) are quite similar, showing an activation in the supplementary motor area and the left primary sensorimotor cortex.

**Concluding Remarks**

In this paper we presented PyHRF, a software to analyze fMRI data using a Joint Detection-Estimation (JDE) approach of the cerebral activity. It jointly detects cortical activation and estimates the Hemodynamic Response Function (HRF), in contrast to existing tools, that consider the HRF as constant over the brain and over subjects, thus aiming to improve the reliability of the results.

PyHRF is an open source software that has evolved rapidly over the last few years. As we showed, it allows to generate Posterior Probability Maps (PPMs) to describe the significance of the activation in each region of the brain. Moreover, PyHRF uses efficient estimation methods in order to provide a fast tool. Currently, the package does not provide finely tuned values for the parameters of the JDE model, leaving the user the cumbersome task of finding the best values for the estimation.

Since 2013, PyHRF has started to evolve to deal with functional Arterial Spin Labelling (fASL) [VWV*13] data, including a physiological prior to make the perfusion estimation more robust [FPV*14] [FFPC15b]. A fast solution for fASL based on VEM was proposed in [FFPC15c], with similar results to the solution...
based on stochastic simulation techniques [FPFC15a]. Moreover, many efforts have been made in terms of image processing, user-friendliness and usability of the PyHRF tool to make it more easy to use by non experts and clinicians.

In the years to come, we plan to develop a light viewer to explore the results of PyHRF interactively. Moreover, we aspire to make the package compatible with Python 3 and extend its use to explore the results of PyHRF interactively. Moreover, we aspire to make the package compatible with Python 3 and extend its use to

**REFERENCES**


Correlated gene expression supports synchronous activity in brain networks. 


