Learning BOLD Response in fMRI by Reservoir Computing

Paolo Avesani*†, Hananel Hazan‡, Ester Koilis‡, Larry Manevitz‡, Diego Sona*†
*NeuroInformatics Laboratory (NILab), Fondazione Bruno Kessler, Trento, Italy
†Centro Interdipartimentale Mente e Cervello (CIMeC), Università di Trento, Italy
‡Department of Computer Science, University of Haifa, Israel

***Abstract*—This work proposes a model-free approach to *fMRI*-based brain mapping where the *BOLD* response is learnt from data rather than assumed in advance. For each voxel, a paired sequence of stimuli and *fMRI* recording is given to a supervised learning process. The result is a voxel-wise model of the expected *BOLD* response related to a set of stimuli. Differently from standard brain mapping techniques, where voxel relevance is assessed by fitting an hemodynamic response function, we argue that relevant voxels can be filtered according to the prediction accuracy of a learning model. In this work we present a computational architecture based on reservoir computing which combines a *Liquid State Machine* with a *Multi-Layer Perceptron*. An empirical analysis on synthetic data shows how the learning process can be robust with respect to noise artificially added to the signal. A similar investigation on real *fMRI* data provides a prediction of *BOLD* response whose accuracy allows for discriminating between relevant and irrelevant voxels.**

***Index Terms*—brain mapping, reservoir computing, model-free HRF.**

I. Introduction

In the framework of the *functional Magnetic Resonance Imaging* (*fMRI*), the main motivation of this work is *Brain Mapping*, a specific data analysis process concerned with the detection of brain areas relevant to specific cognitive or perceptual tasks. The *fMRI* experiments are usually designed by contrasting categories of stimuli, e.g., visual representation of faces versus houses. The working assumption at the basis of the analysis process is that the brain areas concerned with the corresponding cognitive or perceptual task are those allowing for a discrimination of the contrasting categories.

Analysis of *fMRI* data relies on prior assumptions on the *Blood Oxygenation Level Dependent* signal (*BOLD*) underlying the real brain activity signal. Given a stimulation protocol, prior knowledge allows for the definition of the expected *BOLD* response as a parametric *Hemodynamics Response Function* (*HRF*). Fitting of *HRF* parameters is commonly carried out by the solution of a *General Linear Model* (*GLM*). The voxels are therefore rated by uni-variate analysis of the correlation between the real signal and the estimated *HRF*.

There are recent and sophisticated *HRF* models trying to better capture the complex structure of the *BOLD* response (e.g., [1]). Nonetheless, these parametric models still encode the ideal *BOLD* not considering confounds in the design protocol and not including the dependencies due both to the brain structure (e.g., proximity of a big vessel, location) and to the cognitive/perceptual tasks under investigation.

We propose a method trying to tackle the above issues finding the unknown functional dependencies between the *BOLD* signal and the known set of stimuli, without making any preliminary assumption on the expected *HRF*. The proposal is to derive the correct *HRF* voxel-wise with a data-driven approach. We implemented a supervised learning schema that, emulating the brain, produces a prediction of the *BOLD* signal in response to a sequence of stimuli in input. A training process is required for each voxel and for each set of stimuli.

In this schema, the assumption is that the predictability of a voxel behaviour, given the sequence of stimuli, is related to its relevance for the perceptual/cognitive task. A non-predictable voxel is not relevant for the given set of stimuli because its behaviour is presumably driven by other factors. The procedure to build a brain map using the

prediction accuracy is out of the scope of this paper. The major contribution of this work is to provide evidence that the hemodynamic response function can be learnt from data.

The computational learning architecture is implemented using *Liquid State Machines* (*LSM*), a *reservoir computing* model exhibiting long term memory and high computational power [2]–[4]. In these models, a recurrent neural network is used to encode a non-linear transformation of the past input stream into a high-dimensional transient state. The information encoded in the network state may then be used to product various time-series by a simple readout model that in our case has been implemented with *feed-forward neural networks*.

II. Methods

The whole brain *BOLD* activity may be modeled by as many functions as the number of voxels. For each voxel i its *BOLD* activity $v_i(t)$ at time t, as recorded by an *fMRI* scan, is described by an unknown function $f_i(t)$, which encodes the dependency of the voxel behaviour from the sequence of stimuli S:

$$v_i(t) = f_i(S^{[t_0,t]}) + \epsilon_i(t) \tag{1}$$

where, $S^{[t_0,t]} = [S(t_0), \ldots, S(t)]$ is the stimuli time-series from the beginning of the experiment, and $\epsilon_i(t)$ is an unknown function describing the noise. This formalization describes a non-Markovian process, where the *BOLD* activity may in principle depends on the entire past sequence of stimuli, although it is commonly accepted that the hemodynamic response takes less than 20 seconds to recover to initial state. The function f_i may be implemented by a Markovian representation with a Moore system:

$$M(\Lambda, \Theta_i) = \begin{cases} X(t) &= h_\Lambda(X(t-1), S(t)) \\ v_i(t) &= g_{\Theta_i}(X(t)) + \epsilon_i(t) \end{cases} \tag{2}$$

where X is the internal state encoding the past inputs, $h_\Lambda()$ and $g_{\Theta_i}()$ are the transition and the output functions instantiated by the parameters Λ and Θ_i respectively. In this framework, the current state is a function of the previous state and the current stimulus, while the *BOLD* signal is a function of the current state alone. Notice that the transition function h_Λ and the state X are common to all voxels, while the output functions g_{Θ_i} are different for each individual voxel.

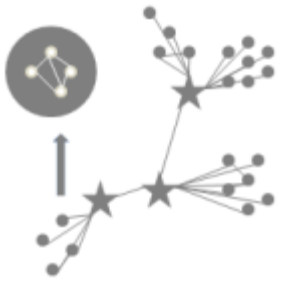

Fig. 1. Hub network architecture.

A requirement for the system $M(\Lambda, \cdot)$ to be Markovian is that the transition function $h_\Lambda()$ must preserve all the past information in $X(t-1)$. Since time-series in the neuroscience experiments are not infinite, we may implement the model with a finite memory dynamic system. The *reservoir computing* is a suitable methods for this purpose. As outlined above, we adopted in particular the *LSM* technology.

The system is made of two components: (i) the *LSM*, a "sufficiently" recurrent and interconnected neural network used to encode a nonlinear transformation of the input stream, capturing the past into a high-dimensional reverberating activity pattern; (ii) a sufficiently strong classifying detector that thanks to the persistence of the liquid trace retrieves the needed temporal information in the liquid state and produces the desired time-series.

In the above framework of eq. (2) the functions $h_\Lambda()$ and $g_{\Theta_i}()$ are the *LSM* and the detectors respectively. The power of this system is based on the separation between the network layer functions: $h_\Lambda()$ expands the input history into a rich enough reservoir state space, while $g_{\Theta_i}()$ combines the neuron signals X into the desired output signal v_i. A complete model is depicted in Figure 2.

Instead of a standard random topology of the *LSM* network as proposed in [2] we adopted a revised hub network architecture introduced in [5], [6] which has been proved to be more robust to noise and more effective in balancing between the input separation and the generalization abilities (See Figure 1). Each vertex of this network has a large number of local

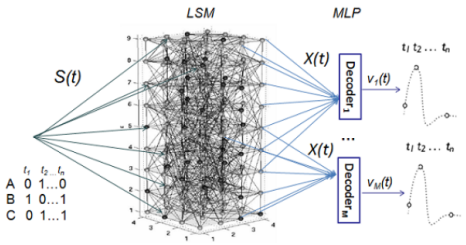

Fig. 2. Model used in the experiments: S(t) is the input stimulus, X(t) is the *LSM* internal state output, and v(t) are the signals decoded by the *MLP*s.

connections and a few global connections. In nature, this topography is commonly derived from networks with local connectivity in which connections have been randomly rewired with a small probability.

The *LSM* network was built using 400 neurons. According to the hub architecture, all neurons were randomly divided into minor sub-networks of 3-6 units. A small amount of sub-networks was selected to be hubs, i.e., the central connection points of the network. This structure created compression points between different network paths, with hubs acting as filters for information flow between groups, resulting in a more reliable model. See [5] for review of different architectural options, their strengths and drawbacks.

The decoder was implemented by a *Multi-Layer Perceptron* (*MLP*), with a linear output unit generating the *BOLD* signal. Such an architecture is slightly different from the usual *LSM* definition [2]. This approach conceptualizes the reconstruction of the *BOLD* response as a regression problem rather than a binary classification.

III. Materials

The analysis of the proposed approach was performed both on synthetic and real data. Synthetic data allowed for a controlled evaluation of the model's generalization property. The challenge is to extract the original dynamics from data corrupted by variable amounts of noise. The real *fMRI* data, on the contrary, was used to verify if the *LSM* model is effective in retrieving relevant voxels, where relevance was determined using the *GLM* method, which can be considered the state-of-the-art in *fMRI* analysis. Both datasets were made up of a temporal sequence of stimuli and a collection of voxels' *BOLD* signal time-course.

A. Synthetic Data

The synthetic data was generated in a framework replicating standard neuroscientific experimental protocols, where the *BOLD* signal of a subject exposed to a sequence of stimuli is recorded using magnetic resonance. The stimuli were generated according to a slow *event related* protocol, where contiguous stimuli are interleaved by a proper fixation time. Long fixation interleaving prevents non-linear overlapping effects on the generated *BOLD* signal. Then the ideal voxels behaviour, i.e. the expected *Hemodynamic Response Function* (*HRF*), was generated from the stimuli time-course using the *Balloon Model* [1], a functional model describing the ideal *BOLD* signal on a physiological basis. The parameters were fitted following the guidelines in [1].

In more detail, we assumed a *Repetition Time* (*TR*) of 2 seconds. Data was generated dividing the stimulation protocol into 4 chunks of stimuli, each one lasting for 15 minutes. The chunks were interleaved by 3 minutes of fixation. The entire experiment began and ended with 3 minutes of fixation lasting for a total of 75 minutes.

During the stimulation phase in each of the 4 chunks, the stimuli of 2 classes, namely A and B, were randomly presented every 20 seconds (10 *TR*s): a stimulus lasting for 2 seconds followed by 18 seconds of fixation, respectively 1 and 9 *TR*s. The prior distribution of the two classes was kept balanced within each chunk. The entire experiment, therefore, amounts to a total of 180 stimuli of the two classes.

Starting from this sequence of stimuli, 4 groups of voxels were generated using the aforementioned *Balloon Model*: (i) voxels reacting to stimuli of both classes A and B, (ii) voxels reacting to stimuli of class A only, (iii) voxels reacting to stimuli of class B only, (iv) voxels unrelated both to A and B. Unrelated voxels were generated with different random sequences of stimuli with a distribution similar to the first group of voxels. Each group of voxels was replicated 4 times corrupting the reference *HRF* time-course by an increasing amount

of *Auto-Regressive Gaussian* noise (*AR*) as follows:

$$\begin{cases} HRF(t) &= Balloon(S^{[t_0,t]}) \\ BOLD_i(t) &= HRF(t) + AR_i(t) \\ AR_i(t) &= 0.9 * AR_i(t-1) + \mathcal{N}(0,\sigma) \end{cases} \quad (3)$$

where $\mathcal{N}(0,\sigma)$ is Normally distributed with zero mean and variance $\sigma \in \{.1, .2, .3, .4\}$. The voxels in the group (iv) were created generating 4 random sequences of stimuli interleaved by random intervals of 16 to 44 seconds of fixation. The *BOLD* signals were then generated using the system equations (3) with $\sigma = .3$. Figure 3 shows a snapshot of some voxel time-courses over a short interval of time.

B. Real Data

The real world experiment was performed on a neuroimaging testbed dataset we acquired in our laboratories. It was specifically designed to evaluate multi-variate pattern analysis techniques on well understood cognitive tasks. During the *fMRI* acquisition, a volunteer was exposed to visual stimuli: *faces*, either plain or scrambled, and *fixation*. Recording was done with an event-related protocol: each *face* stimulus lasting for 2 seconds followed by a jittered interval of at least 9 seconds of *fixation* . The acquisition was performed using a Brucker 4T magnetic resonance, with a resolution of $3 \times 3 \times 3$ mm, with a *TR* of 2 seconds. The run included 516 volumes divided into 16 chunks.

Data was preprocessed computing slice-timing correction, head motion correction, chunk-wise linear detrending, z-scoring and a spatial smoothing with a Gaussian kernel of radius 5 millimeters. A standard analysis based on *GLM* was performed to compute a brain activity map for the *Visual* cognitive task. This activity map (see Figure 4) was then used to score the voxels and select a group of 25 voxels with different scores indicating various relationships to the time-course of the visual stimuli. In particular we selected the 5 most relevant voxels with positive correlation and the 5 most relevant with negative correlation, 5 voxels not relevant at all, and finally, 5 voxels partially relevant with positive correlation and 5 with negative correlation (see the first column of Table II for the group average relevance as indicated by *GLM* score coefficients). In Figure 3 are shown the time-courses of some

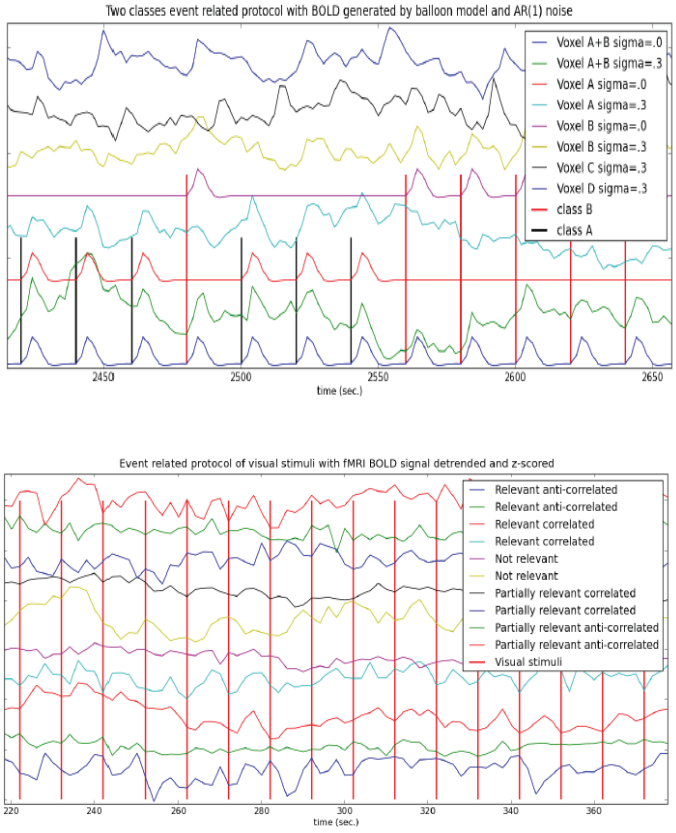

Fig. 3. Plots of voxels behaviour over a short interval of time for synthetic (top) and real (bottom) data. In the top plot are shown the *HRF* time-course (without noise, i.e., $\sigma = 0$) and the *BOLD* signal (with noise $\sigma = .3$) for each group of voxels. The vertical bars are the stimuli of class A (short black) and B (tall red) respectively. Plot at the bottom shows a pair of *BOLD* signal for each relevance group. The vertical bars are the visual stimuli presented to the subject.

voxels for each group over a short interval of time, and in Table II are shown the group average scores.

IV. Results

The empirical analysis aims to assess the ability of the chosen reservoir computing method to reconstruct the expected *BOLD* signal, despite the

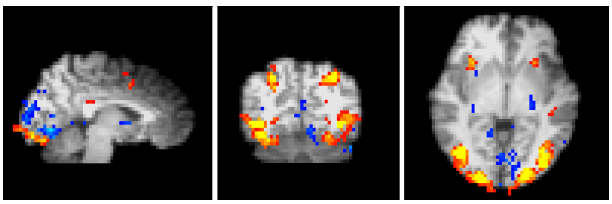

Fig. 4. Map of activity generated with a standard *GLM* analysis on the testbed dataset.

low signal to noise ratio in *fMRI* data. Since the domain doesn't allow to refer to a ground truth we organized the evaluation following a twofold strategy adopting both a synthetic and a real dataset. The former investigation on synthetic data focused on the assessment of model robustness versus the noise. The goal of latter analysis on real data was to evaluate whether accurate supervised learning of *BOLD* response allows for retrieving the relevant voxels in agreement with the scores obtained by *GLM*.

Following the scheme in equation 2, a unique *LSM* built according to the principles proposed in [5] was instantiated for all voxels. Then, different detectors (*MLP*s) were instantiated for different voxels. The synthetic data was divided into 4 folds following the inherent organization of data in chunks and a cross-validation process was iterated over all folds. A simple Pearson correlation was computed between the time series generated by the reservoir-based model on the test set and the original *HRF* before the noise corruption. Table I shows the average cross-validated correlations group-wise.

The encouraging results show that the model is able to generalize the underlying *HRF* with small sensitivity to noise (high correlation for group 1, 2, and 3), while it does not generalize when there is not any relationship between the *BOLD* time-course and the sequence of stimuli (low correlation for group 4). The implemented reservoir computing model appeared to be a very good denoising algorithm for the synthetic data. For a visual representation of the results, Figure 5 shows generalization examples for a voxel reacting to both classes (top graph), and for a voxel unrelated to any class of stimuli (bottom graph). In the graphs, together with the original *HRF* and the noisy *BOLD* time-course, are depicted the signals generated by the model on a test set. For the stimuli-related voxel (top graph) the model accurately generalize the original underlying *HRF* from the noisy signal (reconstructed signal overlaps the original voxel time-course). On the contrary, the model is not able to find any relationship for the unrelated voxel (bottom graph), and the generated signal is not significant.

TABLE I

[SYNTHETIC DATASET RESULTS] THE AVERAGE N-FOLD CV CORRELATIONS BETWEEN THE TIME-SERIES GENERATED BY *LSM* AND THE EXPECTED TIME-SERIES WITHOUT NOISE.

	Noise (σ)			
	0.1	0.2	0.3	0.4
Group 1 (class A&B)	.813	.535	.426	.383
Group 2 (class A only)	.762	.603	.452	.355
Group 3 (class B only)	.746	.587	.419	.363
Group 4 (not relevant)	.114	.083	.035	.095

TABLE II

[REAL DATASET RESULTS] THE AVERAGE N-FOLD CV CORRELATIONS BETWEEN THE TIME-SERIES GENERATED BY *LSM* AND THE REAL *BOLD* SIGNAL. THE FIRST COLUMN CONTAINS THE AVERAGE RELEVANCE DETERMINED USING THE *GLM*-BASED ANALYSIS. THE SIGN INDICATES THE CORRELATION.

	Mean Scores	**Mean Corr**	**St Dev**
Group 1 (relevant pos corr)	15.11	0.343	0.119
Group 2 (semi-relevant pos corr)	2.06	0.079	0.033
Group 3 (not relevant)	0.00	0.042	0.012
Group 4 (semi-relevant neg corr)	-1.78	0.070	0.016
Group 5 (relevant neg corr)	-5.75	0.074	0.027

Similarly to the experiments performed with the synthetic data, the real *fMRI* dataset was splitted into 4 folds and cross-validation was applied for an evaluation of the model. In this case, however, the time series generated by the reservoir-based model on the test dataset were compared with the corresponding real *BOLD* signal. The promising results in Table II show that for the group of strongly relevant voxels (group 1) the model obtains higher correlations than the other groups. On the other hand, little can be said for the groups of voxels with smaller relevance.

V. DISCUSSION

The results obtained on synthetic data showed that reservoir computing method instantiated with *LSM* for the reservoir and with *MLP*s for the detectors may be effective to analyze the *BOLD* time-sequences acquired during *fMRI* experiments. Actually, the model generalizes the *HRF* underlying the noisy *BOLD* signal for all the relevant voxels with a low sensitivity to increasing noise. On the contrary, when the voxels are unrelated to the sequence of stimuli, the model does not generalize any function.

As a consequence, when applying the method to real *fMRI* data, the reservoir-based model may be

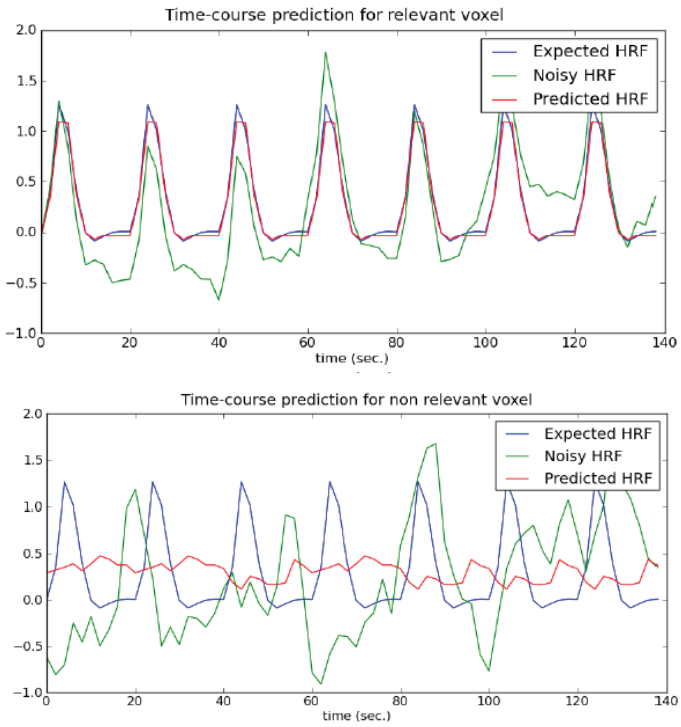

Fig. 5. Generated time-series on a test fold of the synthetic dataset. The upper graph depicts the prediction for voxel 6 of group 2 (class A). The lower graph depicts the prediction for voxel 16 of group 4 (unrelated to any class). A blue line is the expected *HRF*, a red line is the noisy time series, and a green line is the time sequence generated by the model on the test set.

able to extract the relevant information from the relevant voxels. Actually, the voxels that were retrieved as relevant by the standard *GLM*-based analysis were found relevant with the proposed approach as well. More difficult is to find a correspondence between the two approaches on the less relevant voxels and this is matter of further analysis.

In conclusion, we introduced a new promising approach for the analysis of *fMRI* data, based on a novel machine learning approach that exploits a neuro-inspired method to emulate the brain behaviour. This is an important step toward the adoption of model-free methods for the analysis of neuroimaging data.

ACKNOWLEDGMENT

The authors woul like to thank Emanuele Olivetti for the implementation of the Ballon model.

REFERENCES

[1] Y. Zheng, J. Martindale, D. Johnston, M. Jones, J. Berwick, and J. Mayhew, "A Model of the Hemodynamic Response and Oxygen Delivery to Brain," *NeuroImage*, vol. 16, no. 3, Part 1, pp. 617–637, 2002.

[2] W. Maass, T. Natschläger, and H. Markram, "Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations," *Neural Computation*, vol. 14, no. 11, pp. 2531–2560, 2010.

[3] H. Jaeger and H. Haas, "Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication," *Science*, vol. 304, no. 5667, pp. 78–80, 2004.

[4] P. Joshi and W. Maass, "Movement Generation and Control with Generic Neural Microcircuits," in *Biologically Inspired Approaches to Advanced Information Technology*, ser. Lecture Notes in Computer Science. Springer, 2004, vol. 3141, pp. 258–273.

[5] L. Manevitz and H. Hazan, "Stability and Topology in Reservoir Computing," in *Advances in Soft Computing*, ser. Lecture Notes in Computer Science. Springer, 2010, vol. 6438, pp. 245–256.

[6] H. Hazan and L. Manevitz, "The Liquid State Machine is not Robust to Problems in its Components but Topological Constraints Can Restore Robustness," *Proc. of the Int. Conf. on Fuzzy Computation and Int. Conf. on Neural Computation*, pp. 258–264, 2010.