Neural Networks 70 (2015) 61–73

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Non-parametric temporal modeling of the hemodynamic response function via a liquid state machine

Paolo Avesani [a,b], Hananel Hazan [c,e], Ester Koilis [c], Larry M. Manevitz [c,*], Diego Sona [a,d]

[a] NeuroInformatics Laboratory (NILab), Fondazione Bruno Kessler, Trento, Italy
[b] Centro Interdipartimentale Mente e Cervello (CIMeC), Università di Trento, Italy
[c] Department of Computer Science, University of Haifa, Israel
[d] Pattern Analysis and Computer Vision, Istituto Italiano di Tecnologia, Genova, Italy
[e] Network Biology Research Laboratory, Technion, Haifa, Israel

HIGHLIGHTS

- A model for HRF learned directly from data not based on a priori assumptions.
- This allows the BOLD signal to be personalized and voxel specific.
- Voxels can be filtered as "relevant" based on their predictive ability.
- Temporal info stored in reservoir computing; trained feed forward NN produces HRF.
- Learning process robust both to noise and underlying shape of the "true" HRF signal.

ARTICLE INFO

Article history:
Received 7 October 2014
Received in revised form 26 April 2015
Accepted 27 April 2015
Available online 21 May 2015

Keywords:
HRF modeling
Temporal modeling
fMRI
Machine learning
Neural networks
Reservoir computing
Liquid state machines

ABSTRACT

Standard methods for the analysis of functional MRI data strongly rely on prior implicit and explicit hypotheses made to simplify the analysis. In this work the attention is focused on two such commonly accepted hypotheses: (i) the hemodynamic response function (HRF) to be searched in the BOLD signal can be described by a specific parametric model e.g., double-gamma; (ii) the effect of stimuli on the signal is taken to be linearly additive. While these assumptions have been empirically proven to generate high sensitivity for statistical methods, they also limit the identification of relevant voxels to what is already postulated in the signal, thus not allowing the discovery of unknown correlates in the data due to the presence of unexpected hemodynamics. This paper tries to overcome these limitations by proposing a method wherein the HRF is learned directly from data rather than induced from its basic form assumed in advance. This approach produces a set of voxel-wise models of HRF and, as a result, relevant voxels are filterable according to the accuracy of their prediction in a machine learning framework.

This approach is instantiated using a temporal architecture based on the paradigm of Reservoir Computing wherein a Liquid State Machine is combined with a decoding Feed-Forward Neural Network. This splits the modeling into two parts: first a representation of the complex temporal reactivity of the hemodynamic response is determined by a universal global "reservoir" which is essentially temporal; second an interpretation of the encoded representation is determined by a standard feed-forward neural network, which is trained by the data. Thus the reservoir models the temporal state of information during and following temporal stimuli in a feed-back system, while the neural network "translates" this data to fit the specific HRF response as given, e.g. by BOLD signal measurements in fMRI.

An empirical analysis on synthetic datasets shows that the learning process can be robust both to noise and to the varying shape of the underlying HRF. A similar investigation on real fMRI datasets provides evidence that BOLD predictability allows for discrimination between relevant and irrelevant voxels for a given set of stimuli.

1. Introduction

Modeling the HRF signal is a crucial prerequisite for the analysis of fMRI data. One needs to know what is expected a priori from a

* Corresponding author.
E-mail address: manevitz@cs.haifa.ac.il (L.M. Manevitz).

http://dx.doi.org/10.1016/j.neunet.2015.04.009
0893-6080/

given stimulus that affects a voxel in order to, e.g., compare the actual readings and thereby be able to decide if the voxel is relevant to the stimulus or if it is pertinent to differentiate between contrasting stimuli. Most accepted techniques for accomplishing this task involve a mathematical model of the HRF signal. For example, this is the typical assumption in the basic GLM analysis wherein the effects of stimuli are thought to be linearly combined; or the somewhat more advanced methods wherein the time relationship is expanded by a temporal basis function, as in FIR (Goutte, Nielsen, & Hansen, 2000) or Fourier basis functions (Josephs, Turner, & Friston, 1997). In addition, the linear combination can be relaxed (Henson & Friston, 2007) by, e.g., using a Volterra expansion (Friston, Josephs, Rees, & Turner, 1998; Josephs & Henson, 1999) to allow higher order interactions.

In this study, building on previous work (Avesani, Hazan, Koilis, Manevitz, & Sona, 2011) a somewhat different direction is taken, using a temporal neural network approach. The basic idea is that a canonical (at least in the sense of brain modeling) recurrent model reflects the temporal interactions and the information retained over time; and then a second feed-forward network maps this into the readings of the BOLD signal. This method has several conceptual advantages: (i) by separating the feedback effects, the model acts as a Markovian system; (ii) the same recurrent framework can be used for all potential voxels and the difference in response of the voxel resides in a simple feed-forward network; (iii) the parameters of the model (parallel e.g. to "hyperparameters" (Henson & Friston, 2007)) are not related to a specific task but are chosen by general observations (Hazan & Manevitz, 2012). Thus the recurrence can in fact be considered a constant mechanism and is not learned; (iv) the parameters of the feed-forward model can be estimated directly from data.[1]

The application of this computational paradigm allows the use of the predictive success of the model to separate observed voxels into "relevant" and "non-relevant" which, in turn and, in principle, can be used to make studies of differentiation of voxels for specific stimulus related tasks.

1.1. Characterizing the information in the BOLD signal

Functional Magnetic Resonance Imaging (fMRI) is a widely used modality in the studies of brain perception and cognition, with applications to a broad variety of neuroscientific questions including brain mapping, which corresponds to the set of data analysis methods designed to detect and map 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) and analyzing the recorded data to find brain areas related to the classes of stimuli. The underlying assumption is that the brain areas allowing for discrimination between the contrasting categories are related to the corresponding cognitive or perceptual task.

The existing brain mapping methods can be divided into two categories: the hypothesis-driven methods and the data-driven methods. While the former methods use different prior assumptions on the basic characteristics of the BOLD (Blood Oxygenation Level Dependent) signal, the later class of methods, usually having an exploratory nature, tries to infer these characteristics directly from the data.

The hypothesis-driven methods for the fMRI data analysis are used in the majority of brain mapping studies. A survey (Grinbald, Wager, Lindquist, & Hirsch, 2008) of 170 fMRI studies shows that 96% of experiments were based on the hypothesis-driven analysis methods. These methods have a strong statistical framework for assessing the relevant regional activation areas. However, they rely on prior assumptions on the BOLD signal underlying the real brain activity. Given a stimulation protocol, prior knowledge enables the definition of the expected BOLD response as a parametric Hemodynamics Response Function (HRF), which is used in a General Linear Model (GLM) framework (Friston et al., 1994; Monti, 2011). The outcome is based on the univariate analysis of the correlation between the real signal and the estimated HRF. These methods require, therefore, an accurate definition of the expected HRF shape, although it is allowed to vary significantly in different populations, between subjects, and between different brain areas (Aguirre, Zarahn, & D'Esposito, 1998).

There are recent and sophisticated parametric HRF models that attempt to better capture the complex structure of the BOLD response (Zheng et al., 2002). Nonetheless, these models still encode some ideal BOLD shape without considering possible uncommon or irregular hemodynamics due both to the brain structure (e.g., proximity to a large vascular vessel or changes caused by different brain injuries) and to the kind of cognitive tasks under investigation.

To avoid these issues, one can try to obtain a correct HRF voxel by voxel with a data-driven approach, thereby finding the unknown functional dependencies between the BOLD signal and the known set of stimuli, without any prior assumption on the expected HRF. Some data-driven methods are reported in the literature, such as selective averaging with a long inter-stimulus interval assuming non-overlapping responses (Bandettini & Cox, 2000; Buckner et al., 1996) or the FIR methodology (Goutte et al., 2000). Other methods, like Bayesian approaches (Woolrich, Jenkinson, Brady, & Smith, 2004) or wavelet deconvolution (Wink, Hoogduin, & Roerdink, 2008) are computationally expensive or require some additional prior assumptions (e.g., separability of signal and noise in the frequency domain for the wavelet methods).

A promising approach in this direction is the FIR approach (Goutte et al., 2000; Henson & Friston, 2007). In FIR the HRF is modeled as a linear combination of "impulse basis functions", a set of adjacent boxcar functions, specified over the period of time fitting the expected duration of BOLD response. Then the corresponding coefficients are calculated from the data. The combination of FIR basis functions can capture any shape of response up to a given timescale (Henson & Friston, 2007). However, the precision of FIR model depends on a correct specification of the expected duration of hemodynamic response. Moreover, the FIR model is linear and it can suffer low sensitivity with experimental protocols generating non-linear effects in the HRF.

In another approach, Wang (2009) proposed a method for brain mapping based on machine learning techniques. This method is a combination of the HRF data-driven analysis and the hypothesis-driven GLM inference. In this method, the HRF profile extracted from the Support Vector Machine (SVM) classifier (Burges, 1998; Vapnik, 1995) is used as the data-driven regressor for the consequent GLM analysis. Note that SVM is a machine learning technique that has been demonstrated to be successful for the analysis of neuroimaging data in many applications (Atir-Sharon, Gilboa, Hazan, Koilis, & Manevitz, 2015; Boehm, Hardoon, & Manevitz, 2011; Cox & Savoy, 2003; Hardoon & Manevitz, 2005; Mitchell et al., 2004; Mourão-Miranda, Bokde, Born, Hampel, & Stetter, 2005) including, in particular, brain-decoding using a framework referred to as multivariate pattern analysis (MVPA). In this approach one tries to *classify the contrasted stimuli* from the existing BOLD signal. However, the HRF derivation under discussion here is a reverse task, compatible with a standard brain-mapping analysis, in which one wishes to *reconstruct the expected BOLD signal*based on the stimuli sequence.

[1] The architecture of the feed-forward model can be considered a separate parameter as in most neural network research. However, we neglect this issue because it is consistent with the model to replace the feed-forward network with any other machine learning scheme. See below.

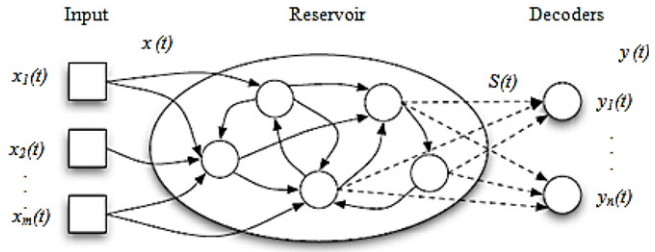

Fig. 1. Reservoir computing network. The reservoir processes a multi-dimensional input data stream $x(t)$ generating a series of high-dimensional internal state $S(t)$. At the same time the decoders produce the required multi-dimensional output function $y(t)$ based on the generated internal states.

Also note that the substantially temporal nature of both BOLD signal and the designed stimuli sequences makes it difficult to efficiently use machine learning methods such as SVMs without complex preprocessing procedures encoding the temporal dimension into a spatial representation. Such encoding schemes are subject to limitations in the total length of the experiment.

The current study shows how to address the HRF decoding task without incurring the above temporal issue using a method based on *Liquid State Machine* (Hazan & Manevitz, 2012; Jaeger & Haas, 2004; Maass, Natschlager, & Markram, 2002), a machine learning technique explicitly designed for complex temporal signal processing (Hazan & Manevitz, 2012; Jaeger & Haas, 2004; Maass et al., 2002). Using this technique, a supervised learning schema is implemented that produces a prediction of the BOLD signal in response to a sequence of stimuli in input (a training process is required for each voxel). The success of this trained predictor on test data serves as an indicator of the relevance of the voxel for the corresponding perceptual/cognitive task. Potentially, the proposed method allows for the discovery of unknown functional dependencies between the BOLD signal and the known set of stimuli with a completely data-driven approach, which does not make any prior assumption on the expected hemodynamic function.

1.2. Liquid state computing

The current study addresses the HRF decoding task using a method based on *Liquid State Machine* (Hazan & Manevitz, 2012; Jaeger & Haas, 2004; Maass et al., 2002). The main principle underlying such computational model can be explained via an analogy with a pebble thrown to a pond. When a small pebble is thrown to a pond full of water (a liquid reservoir), it generates waves representing a series of liquid states that can be measured at different time steps. Each liquid state represents a unique trace of the past activity. This way, the liquid reservoir's latest state contains the history of all previous events enabling the efficient encoding of temporal information. The obtained series of consequent states can be recorded with any suitable recording device and further exploited/analyzed. The device responsible for "reading" or "decoding" the information from the internal reservoir states is called a *readout*, or *decoder*. The readout typically identifies the desired series of events based on the internal reservoir states. The decoder is able to produce a required series of signals functionally related to the input (Maass et al., 2002).

More specifically, such a system (Fig. 1) is made of two components: (i) the *reservoir*, a nonlinear dynamical system based on a recurrently coupled network of artificial neural computational nodes; (ii) a sufficiently strong *decoder* or *readout* retrieving the temporal information in the reservoir's states and producing the desired output time-series. The reservoir is used to encode a nonlinear transformation of the input stream, capturing the past history into a high-dimensional reverberating internal activity state. From a machine learning perspective, a reservoir operates as a temporal kernel, projecting the input to a dynamic non-linear high-dimensional space. The reservoir states form a trajectory, which is dependent on both the current external input and the memory traces of previous stimuli. In this way, reservoir architectures can inherently process spatiotemporal patterns.

The power of this system is based on the separation between the network layer functions: the reservoir expands the input history into a rich state space, while the decoder combines the series of the internal states into the desired output signal.

The computations of the reservoir computer are driven by the underlying recurrent neural network dynamics and may be expressed as:

$$\begin{cases} \boldsymbol{S(t)} = \boldsymbol{L(x(t), S(t-1))} \\ \boldsymbol{y(t)} = \boldsymbol{D(S(t))}, \end{cases} \tag{1}$$

where the internal network state S at time t is generated by an operator L integrating the input value x to the network at current time t with the previous internal network state at time $t-1$. This transformation creates a new high-dimensional reservoir activity pattern carrying the information about the entire input sequence. The exact definition of L depends on the type of neurons and the topology used to construct the network. Based on a newly generated internal state $S(t)$, a memory-less decoder function D transforms the current liquid state into the machine output $y(t)$. The detector function D is commonly implemented using a classification or a regression algorithm trained with simple learning mechanisms.

The whole architecture is designed to make the learning as fast and robust as possible (Maas, 2010). It delegates the primary load of the goal-directed learning to a single and seemingly trivial stage — the output stage, which typically is a relatively simple computational component. The reservoir serves as a pre-processor for such a detector unit, creating the range of possible projection functions of the input streams $x(t)$ that the detector can learn. Such division of computational processing into reservoir and decoder is efficient since the same reservoir can serve a large number of different decoders, trained to extract different information from the same internal state sequence.

As described above, one may find a certain similarity between reservoir and kernel methods. The reservoir can be seen as a nonlinear transformer of temporal information into a higher-dimensional space. There are, however, some significant differences between the two methods. In the case of reservoirs, the mapping into the new space is *explicitly temporal as it encodes long-term and cross-stimuli dependencies in the input signal without the need to encode this information as part of the input*. In addition, reservoir computing manages both inputs and outputs as streams of data potentially in continuous time (Buonomano & Maass, 2009; Maas, 2010). This temporal dimension makes this model particularly suitable for fMRI data analysis, with both input stimuli and output BOLD signals representing relatively long data streams.

1.3. The bold signal

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

$$\boldsymbol{y_i(t)} = \boldsymbol{f_i(x^{[t_0;t]})} + \boldsymbol{\varepsilon_i(t)} \tag{2}$$

where $x^{[t_0;t]} = [x(t_0); \ldots; x(t)]$ is the stimuli time-series from the beginning of the experiment to the current time step t, and $\varepsilon_i(t)$ is an unknown function describing the noise and any unknown

component of the BOLD signal not related to the external stimuli. If the voxel behavior has some relationships with the sequence of stimuli, the function $f_i(t)$ in Eq. (2) is not null. On the contrary, if the voxel is not influenced by the external stimuli, the function $f_i(t)$ is a null component and the entire signal is described by $\varepsilon_i(t)$. This formalization describes a non-Markovian process, where the BOLD activity may in principle depend on the entire past sequence of stimuli, although it is commonly assumed that the hemodynamic response takes less than 20 s to recover to the initial state. The function $f_i(t)$ may be implemented by a Markovian representation as:

$$\boldsymbol{M}(\Lambda, \Theta_{\mathbf{i}}) = \begin{cases} \boldsymbol{S(t)} = \boldsymbol{h}_{\Lambda}(\boldsymbol{S(t-1)}, \boldsymbol{x(t)}) \\ \boldsymbol{y_i(t)} = \boldsymbol{g}_{\Theta_i}(\boldsymbol{S(t)}) + \boldsymbol{\varepsilon_i(t)} \end{cases} \tag{3}$$

where the output function g_{Θ_i} replacing $f_i(t)$ is now based on an internal state S only, determined by a transition function h_{Λ}, which encodes the entire input history as a function of the previous state and the current stimulus. The transition and the output functions h_{Λ} and g_{Θ_i} are instantiated by the parameters Λ and Θ respectively. Notice that the parameters for the transition function h and the state S are common to all voxels, while the parameters for the output functions g are different for each individual voxel.

In order for the system $M(\Lambda, \Theta_i)$ to be Markovian, the transition function h_{Λ} must preserve all the past information in $x(t-1)$. However, since the time-series in the neuroscience experiments are not infinite, h_{Λ} may be implemented with a finite memory dynamic system. *In particular, the system described in Eq. (3) may be implemented using the reservoir computer defined by Eq. (1).* In this mapping, the input signal $x(t)$ describes the stimuli sequence used in the experiment, the reservoir state $S(t)$ abstracts a representation of internal neural activity caused by the stimuli, and, finally, the decoder output $y(t)$ characterizes the model of the oxygenation producing the BOLD signal relevant for a given activity pattern. Functions h_{Λ} and g_{Θ_i} may be substituted by the reservoir operator L and the decoder function D respectively. The reservoir plays a central role in the proposed system by representing the temporal dimension of the presented stimuli sequence in its state. Moreover thanks to its "fading-out" property in the absence of external input, it may forget the preceding series of internal states/inputs imitating the final nature of the BOLD response where all blood parameters return back to the baseline level in the absence of a new stimulus.

The operational cycle of a reservoir computer as applied to the analysis of BOLD signal is depicted in Fig. 1. The binary stimuli vector $x(t)$ describing the time series of impulse-like stimuli $x(t) = [x_1(t), \ldots, x_m(t)]^{\mathrm{T}}$ (where m is a number of different classes of stimuli used in the experiment) is provided as input to the reservoir without any additional preprocessing steps. It is fed into the reservoir network, value by value, at each time step. The reservoir integrates the information in a parallel computation and produces a series of rich high-dimensional internal activity states $S(t)$ encoding all past input. At the same time, the decoder units $D_1 \ldots D_n$ (where n is a number of voxels in the brain) recreate the BOLD signal $y_1(t) \ldots y_n(t)$ based on the encoded internal state.

Using information metrics (e.g. correlation or mean squared error) it is therefore possible to evaluate the similarity between the voxel BOLD signal reconstructed by the reservoir method and the real BOLD signal detected by fMRI scanner, both generated as a response to the same stimuli sequence.

As a working assumption, a voxel is relevant for the particular perceptual/cognitive task if it is related in some way to the sequence of corresponding stimuli and irrelevant if no such relation exists. In other words, the ability of the model to act as a predictor for the BOLD signal time-course serves as an indicator for the relevance of a given voxel to the task.

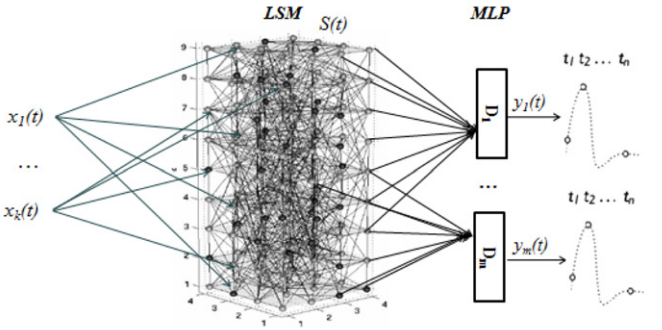

Fig. 2. Model used in the experiments: $\boldsymbol{x(t)}$ is an input stimuli (k — the number of different input stimuli), $S(t)$ is an internal LSM state output, and $y(t)$ are the synthetic BOLD signals decoded by the MLP (m — a number of brain volume voxels). Not shown here is the internal topology of the liquid which is described in the text.

2. Methods

2.1. The reservoir

The *Liquid State Machine* (*LSM*), a recurrent network based on spiking computational units, was used as the Markovian state ("reservoir") in all the following experiments. It is emphasized that a single Liquid State Machine instance was created for all voxels (Fig. 2).

There are many parameters to be specified in the choice of a reservoir, including the types of neurons, weights between neurons and topology of connectivity.[2]

The specific instance of this recurrent network was made of 235 biologically plausible leaky-integrate-and-fire neurons (Lapicque, 1907). These parameters were chosen after some experimentation, but they are roughly in line with previously reported results (Maass et al., 2002; Verstraeten, Schrauwen, D'Haene, & Stroobandt, 2007). 80% of the neurons had positive weights (excitatory neurons) and 20% of which had negative weights (inhibitory neurons). This ratio has its origin in modeling the rough proportions that occur in cortex, and from our experience this proportion of positive and negative weights also enjoys the property of reducing the variance between runs. Weight values were selected randomly from the interval 0.5 to 0.51, both for positive and negative weights. Intuitively the choice of the weights is a balance between requiring more than one signal for a neuron to fire thereby helping to screen out noise and preserving the memory with large enough weights to keep the network from "forgetting" too quickly.[3]

These weights are kept constant as only the decoder adapts its parameters within a training process. The initial voltage of neurons was set to −65 mV, with the actual spike level of 60 mV, corresponding to a scale reported by Maass (Maass et al., 2002). Some neurons (20%) chosen randomly throughout the reservoir were defined as input neurons, all of them received identical input signals. The remaining nodes were defined to be the output.

Instead of a standard random topology on the LSM network (Maass et al., 2002) or a uniform random network (Verstraeten

[2] In a nutshell, the most crucial parameter in all our experiments was the topology (a double power law generating a small world arrangement). In order to keep the information available for a designated period of time there was a clear trade-off between the number of neurons and the use of a sliding threshold (See Hazan, 2014).

[3] The choice of coupling strength has been empirically determined in several investigations (Hazan, 2014; Hazan & Manevitz, 2012). The main issue is a balance between sensitivity and over-reaction of the liquid to sensitivity. 0.5 has proven itself to be a good value. In addition, a variance in the coupling strength is necessary so that the feed-back based on the recurrence in the liquid preserves both the separation of signals and the specific information in a particular signal. Accordingly, random values around 0.5 were chosen.

et al., 2007) we adopted a revised Watts–Strogatz small-world network architecture (Watts & Strogatz, 1998) that has been shown to be more effective in balancing between the input separation and the generalization abilities (Hazan, 2014; Hazan & Manevitz, 2012). To create a small-world network the connections between pairs of neurons were chosen by selecting the pairs according to a power law. The remaining disconnected components (either from an input or output perspective) were connected to another neuron chosen randomly but preserving the power law proportionally to the existing connectivity level (Hazan, 2014; Hazan & Manevitz, 2012). Since small world topologies follow a power law connectivity, they produce hubs, i.e., vertexes having a large number of local connections and a few global ones. Such topologies are thought to "naturally" emerge and appear in the real neural systems (Albert & Barabasi, 2000). The network adopted in this paper has both input and output directions processed separately according to the small-world rules, thus, causing distinct neurons to serve as "hubs" for input or for output (Hazan, 2014; Hazan & Manevitz, 2012). This structure creates compression points between different network paths, with hubs acting as filters for information flow between dense groups of neurons, resulting in a more reliable model.

While the external (input/output) streams of data were sampled at a constant and synchronized time rate, in the current implementation the LSM was operating at a higher frequency (10 times faster than the data streams flow), where input stimulus injection is followed by 10 autonomous LSM iterations. This is related to the nature of the data we are modeling in this application (i.e. extremely low level of sampling in the BOLD signal given to us by the fMRI data in the order of every 2 s which is based on millisecond physiological changes).[4] The output state was defined as the last LSM state in each series of 10 autonomous runs.

2.2. The decoders

The behavior of each voxel was modeled by a different decoder instantiated and trained separately voxel by voxel. All decoders had the same input, corresponding to the *LSM* output. The decoders were implemented by *multi-layer perceptrons* (*MLP*). The hidden layer contained a limited amount of neurons (30% of the reservoir output size[5]) with a hyperbolic tangent activity function, and a single linear output unit generating the BOLD signal. The initial *MLP* weights were selected randomly using Nguyen–Widrow randomization method (Nguyen & Widrow, 1990) and modified using the resilient back-propagation algorithm (Riedmiller & Braun, 1993), both enabling a relatively quick and effective training phase. This can be particularly important for processing full brain volumes which might have thousands of MLPs to be trained (one for each voxel).

2.3. Simulation and assessment

The reconstruction of the BOLD signal was performed in two phases within a cross-validation framework to ensure the results were not incidental. During a training phase, a portion of data was given to a supervised learning process that fitted the decoders' parameters to produce the required BOLD signal voxel by voxel given the sequence of stimuli. In the testing phase, a remaining portion of data held-out from training was used to generate the expected BOLD response related to a set of stimuli, which was then compared to the real BOLD signal for relevance analysis. The initial assumption is that given a sequence of stimuli the *predictability* of a voxel behavior is related to its *relevance* for the perceptual/cognitive task. The purpose of the testing phase was, therefore, to evaluate the quality of the created model and thus to confirm/disprove the existence of any relationship between the stimuli and the recorded BOLD signal.

The assessment was based on two quality indices calculated between the original and the synthetic BOLD signals produced by the LSM: the root mean square deviation (RMSD) and the Pearson correlation. Good prediction accuracy (low RMSD and high correlation values) is associated with high model quality and indicates the relevance of a voxel for a given cognitive task. On the other hand, low prediction accuracy (high RMSD and low correlation values) indicates that the model does not match the data. As a result the unpredictable voxel can be considered to be not influenced by the given set of stimuli because its behavior is presumably driven by other factors, and therefore irrelevant in the context of the given experimental protocol.

This approach conceptualizes the reconstruction of the BOLD response as a regression problem rather than a binary classification. The reconstruction is data-driven and does not require any prior assumption on the expected hemodynamic response shape.

3. Materials

The proposed method was extensively evaluated on a variety of synthetic and real datasets. Synthetic data allowed for a controlled evaluation of the model's generalization property. In particular, we evaluated its ability to extract the original dynamics with various HRF shapes and with variable amounts of noise corruption. Real fMRI data, on the contrary, was used to verify the effectiveness of the reservoir computing model in retrieving relevant voxels, where the relevance was determined using the reference GLM-based analysis method (Friston et al., 1994).

The real dataset was generated in our laboratories with the aim to collect a neuroimaging reference dataset in order to test multivariate data analysis methods. Synthetic datasets were constructed generating a collection of BOLD signal time-course observations from a temporal sequence of stimuli simulating a real experiment.

To ensure a sound analysis of the method and to study its robustness to noise the synthetic datasets were generated corrupting the ground truth by different *amounts of structured noise*.

For each synthetic dataset, a sequence of stimuli denoted by $x(t)$ describing a virtual task was created with each of the design protocols outlined below. The ideal BOLD signal was then generated from the stimuli time-course using the Balloon Model (Zheng et al., 2002), a functional model describing the ideal Hemodynamic Response Function (HRF) on a physiological basis. The model parameters were fitted following the guidelines in Zheng et al. (Zheng et al., 2002). The precise choices are listed in the Appendix. Finally, a number of synthetic voxels were generated corrupting the ideal BOLD signal by various degrees of Auto-Regressive Gaussian noise (AR) as described below:

$$\begin{cases} \mathbf{HRF}(\boldsymbol{t}) = \mathbf{Balloon}(\boldsymbol{X}(\boldsymbol{t})) \\ \mathbf{BOLD}_{\boldsymbol{i}}(\boldsymbol{t}) = \mathbf{HRF}(\boldsymbol{t}) + \mathbf{AR}_{\boldsymbol{i}}(\boldsymbol{t}) \\ \mathbf{AR}_{\boldsymbol{i}}(\boldsymbol{t}) = \boldsymbol{\alpha} * \mathbf{AR}_{\boldsymbol{i}}(\boldsymbol{t} - \mathbf{1}) + \in_{\sigma}, \end{cases} \tag{4}$$

where α is a constant controlling the amount of structure in the noise (in our experiments we adopted $\alpha = 0.9$) and $\in_{\sigma} \sim N(0, \sigma)$ is the white noise normally distributed with zero mean and variance σ^2.

[4] This difference between "internal" and "external" time rates is related to the interaction between the refractory period, the sampling rate and the leak rate of neurons. For example, with a 1–1 time relationship and a non-trivial refractory period, a second successive input would always be ineffectual.

[5] The hidden layer size was selected experimentally, using pruning heuristics.

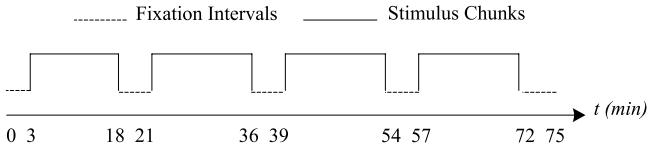

Fig. 3. The stimulation protocol. Chunks of stimuli lasting 15 min are interleaved by fixation intervals lasting 3 min.

3.1. Synthetic data with varying design protocols

3.1.1. Design protocols

The challenge of the proposed analysis method was to generalize various plausible hemodynamics; therefore both synthetic and real datasets were generated according to the most common *design protocols* adopted in fMRI experiments:

(i) *Slow event related design* — a set of stimuli of different classes (e.g. classes A and B) are randomly ordered in low-frequency sequences interleaved by long fixation intervals, preventing non-linear overlapping on the BOLD activity of consecutive stimuli;
(ii) *(Ultra)-fast event related design* — the stimuli of different classes are randomly ordered in high-frequency sequences with stimuli interleaved by randomly short fixation intervals (usually 0 to 10 s). These intervals allow for non-linear overlapping of effects for consecutive stimuli;
(iii) *Block design* — a set of stimuli of the same class are presented with high frequency (one every TR) in a block of 15 to 20 s. Each class of stimuli is presented in a different block. The blocks of stimuli are interleaved by blocks of fixation of the same length to prevent overlapping of BOLD activity of two consecutive blocks and maximize the sensitivity of statistical analysis methods.

3.1.2. Emulated fMRI

Emulating a typical fMRI experiment, the stimulation protocol was divided into 4 chunks of stimuli, each one lasting for 15 min (Fig. 3). Cross-validation was performed based on the created 4 folds. The chunks were interleaved by 3 min of fixation. The entire experiment began and ended with 3 min of fixation for a total length of 75 min. A Repetition Time (TR) of 2 s was used in all datasets, producing sequences made of 2250 samples ($75 * 60/2$).

The stimuli for each design were generated as follows:

(i) *Dataset 1 — Synthetic Event-Related Design:* During the stimulation phase in each of the 4 chunks, each stimulus was presented for one TR (2 s) followed by a fixation interval of 18 s (9 TRs). The entire experiment, therefore, amounted to a total of 180 stimuli;
(ii) *Dataset 2 — Synthetic Fast Event-Related Design:* This dataset was generated using the above schema with a random inter-stimuli fixation interval (2 to 14 s). This resulted in a dataset made of approximately 450 stimuli;
(iii) *Dataset 3 — Synthetic Block Design:* In this dataset, the stimuli were organized into homogeneous blocks, each one made of 8 consecutive stimuli lasting for a total of 16 s. The blocks were interleaved by 16 s of fixation.

From these sequences of stimuli, the aforementioned Balloon Model was then used to generate the following 4 groups of voxels for each dataset:

i. voxels with the hemodynamic influenced by both classes of stimuli A and B;
ii. voxels with the hemodynamic influenced by stimuli of class A only;
iii. voxels with the hemodynamic influenced by stimuli of class B only;
iv. voxels unrelated both to A and B.

For each of the first 3 above categories, 25 voxels were created in groups of 5. Each voxel in a group was generated with equal degree of *AR* noise with impulsive (Gaussian) component, controlled by different values of standard deviation $\sigma \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ for different groups.

The voxels in group (iv) (irrelevant for the stimuli) were created generating 25 random sequences of stimuli interleaved by random fixation intervals of 16 to 44 s. The BOLD signals were then generated with the Balloon model and corrupted with AR noise controlled by $\sigma = 0.3$ (Eq. (4)). For fast event related design, the stimuli for the group of irrelevant voxels were generated for each voxel separately with a distribution similar to the "relevant" task that is, with mean fixation interval of 8 s and jittering of 6 s. This group of voxels having the BOLD signals without any relationship to the original task was used as the baseline. The absence of generalization for this group of irrelevant signals was the expected behavior of the proposed model.

3.2. Synthetic data with varying HRF shape

The synthetic dataset with varying HRF was generated to explore the model capability to discover voxels having unusual HRFs hardly modeled by commonly used hemodynamic models. The stimulation protocol, with the same overall organization of above datasets, was generated using a slow event related design with only one class of stimuli. Each of the 4 chunks was composed of a sequence of stimuli presented every 30 s without jittering. The sequences were generated with a sampling rate (TR) of 0.1 s and were then re-sampled with a TR of 2 s. The entire experiment, therefore, amounts to a total of 120 stimuli.

From this sequence of stimuli, six groups of voxels were generated using the aforementioned Balloon Model, five of which were related to the stimuli. In particular, all groups but one were generated with a significant unconventional variation of the underlying HRF[6]:

(i) a standard HRF (the baseline reference group);
(ii) an HRF with oscillatory behavior,
(iii) a stretched HRF with a delayed undershoot;
(iv) a HRF with delayed time to peak;
(v) an HRF having two peaks heavily delayed;
(vi) a standard HRF (with random sets of stimuli — "irrelevant" signals).

Note that the sixth group of voxels was generated using the baseline HRF applied to a random sets of stimuli, producing hemodynamics not related to the corresponding task.

Fig. 4 depicts the HRF models that were used to generate the synthetic voxels related to the stimuli and the corresponding sampled curves with the sampling rate used in our experiments (2 s).

Similarly to experiment A, for each different HRF model a reference BOLD time-series was generated, and in turn it was corrupted by different degrees of AR noise with impulsive (Gaussian) component controlled by $\sigma \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$, generating a set of 25 voxels in groups of 5.

[6] The choice for this set of atypical HRFs was designed to allow the model to discover "unexpected" hemodynamics. Some of the responses (i.e., iii and iv) relate to known variations for different ROIs. On the other hand responses (ii) and (vi) are rather unconventional but they allow investigating the ability of the proposed model to discover "unexpected" hemodynamics.

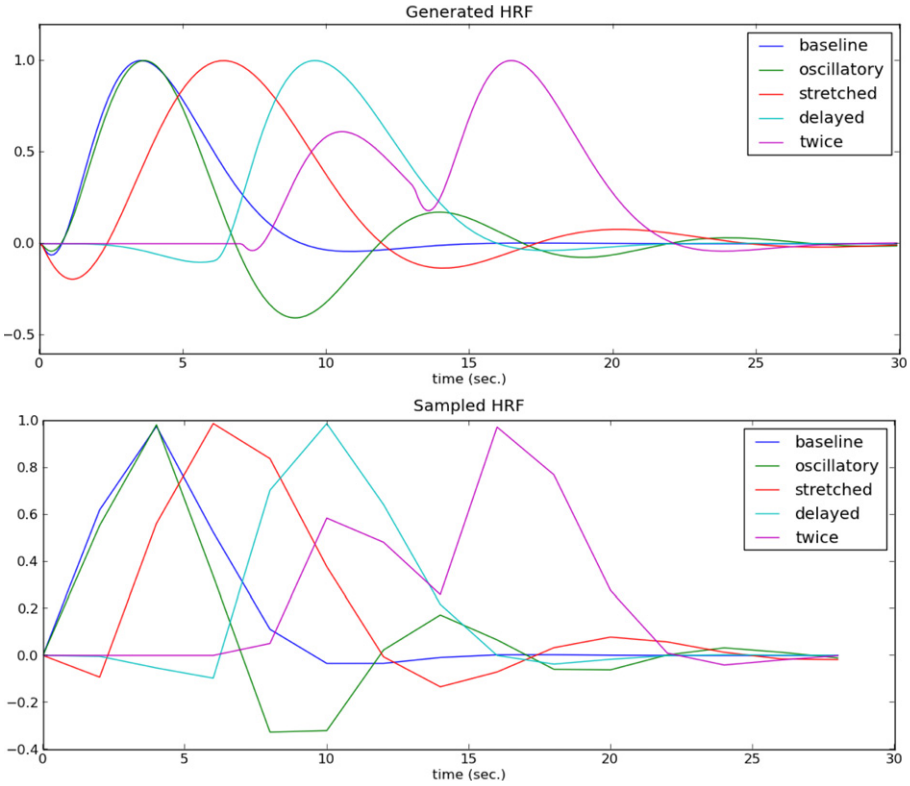

Fig. 4. Snapshot of the various shapes used to generate the synthetic data with an unusual HRF. In the top plot are depicted the expected HRFs while in the bottom plot are depicted the same curves sampled at 2 s.

The reference HRF for the "baseline" voxels was created adopting the Balloon model with parameters fitted following the guidelines in Zheng et al. (2002). The same parameters were used to generate the "irrelevant" signals, whose sequence of stimuli was a chunk-wise random permutation of the original sequence. For the "oscillatory" and the "stretched" HRFs small modifications of some Balloon model parameters (signal decay, auto-regulation, venous time constant, and neuronal efficacy) were applied. The "delayed" time series were generated modifying the sequence of stimuli, splitting each stimulus in a sequence lasting 6 s of exponentially increasing stimuli. Moreover, some parameters of the Balloon model (neuronal efficacy and venous time constant) were changed. Finally, the "twice peaked" time series, presenting a double peak, was generated shifting the stimuli twice by 7 and 13 s, with the first stimulus amplitude being 60% of the second one, and without any change to the Balloon model parameters. See Appendix for more details on the parameters setting. All voxels were then normalized to have unit maximum before adding the AR noise.

3.3. Real data

The real dataset was acquired in our laboratories aiming at collecting a neuroimaging dataset, specifically designed to evaluate multivariate 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. 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 s. The data was recorded using the same three stimulation protocol designs as used for the synthetic data generation: (i) Slow event related design; (ii) Fast event related design; (iii) Block design. The obtained data was preprocessed computing slice-time and head motion correction, chunk-wise linear de-trending, normalization, and spatial smoothing using isotropic Gaussian kernel with a radius sigma of 5 millimeters (i.e., a Gaussian blur done over a volumetric window of $11 \times 11 \times 11$ mm with a diagonal covariance matrix of 5 mm in all directions).

Table 1
Mean t-values (and mean standard deviations) produced by GLM analysis for each group of selected voxels and for each design protocol.

	Discriminatory voxels	Non-discriminatory voxels
Event related	14.7 ± 1.88	$3.4e{-6} \pm 8.1e{-7}$
Fast event related	16.2 ± 1.64	$-6.5e{-5} \pm 5.3e{-7}$
Block	26.4 ± 7.35	$-1.6e{-5} \pm 5.5e{-7}$

The ground truth for relevant and irrelevant voxels was determined by selecting representative time-series from voxels chosen from an activation map generated with a standard GLM-based analysis. Thus we identify the 50 most "relevant" voxels with the most "discriminatory" voxels as determined by the t-values in the GLM setting. Similarly, we identify the 50 "irrelevant" voxels as the most non-discriminatory voxels as determined by the t-values in the GLM setting. Table 1 gives the t-values of the two chosen sets for each experiment. Note that the non-discriminatory and thus our "irrelevant" voxels have essentially a t-value of 0.[7]

Snapshots of two voxel time-courses for each group and for each protocol over a short interval of time are shown in Fig. 5.

4. Results

The empirical analysis aims at assessing the ability of the suggested computational approach to filter the expected BOLD signal,

[7] The chosen non-discriminatory voxels turn out to be well (randomly) distributed throughout the brain. Since the task was a visual one, the discriminatory voxels were in fact all chosen within the occipital lobes. They were indeed distributed reasonably broadly within the lobes; and at least four or five distinct clusters exist, with substantial variance within each cluster.

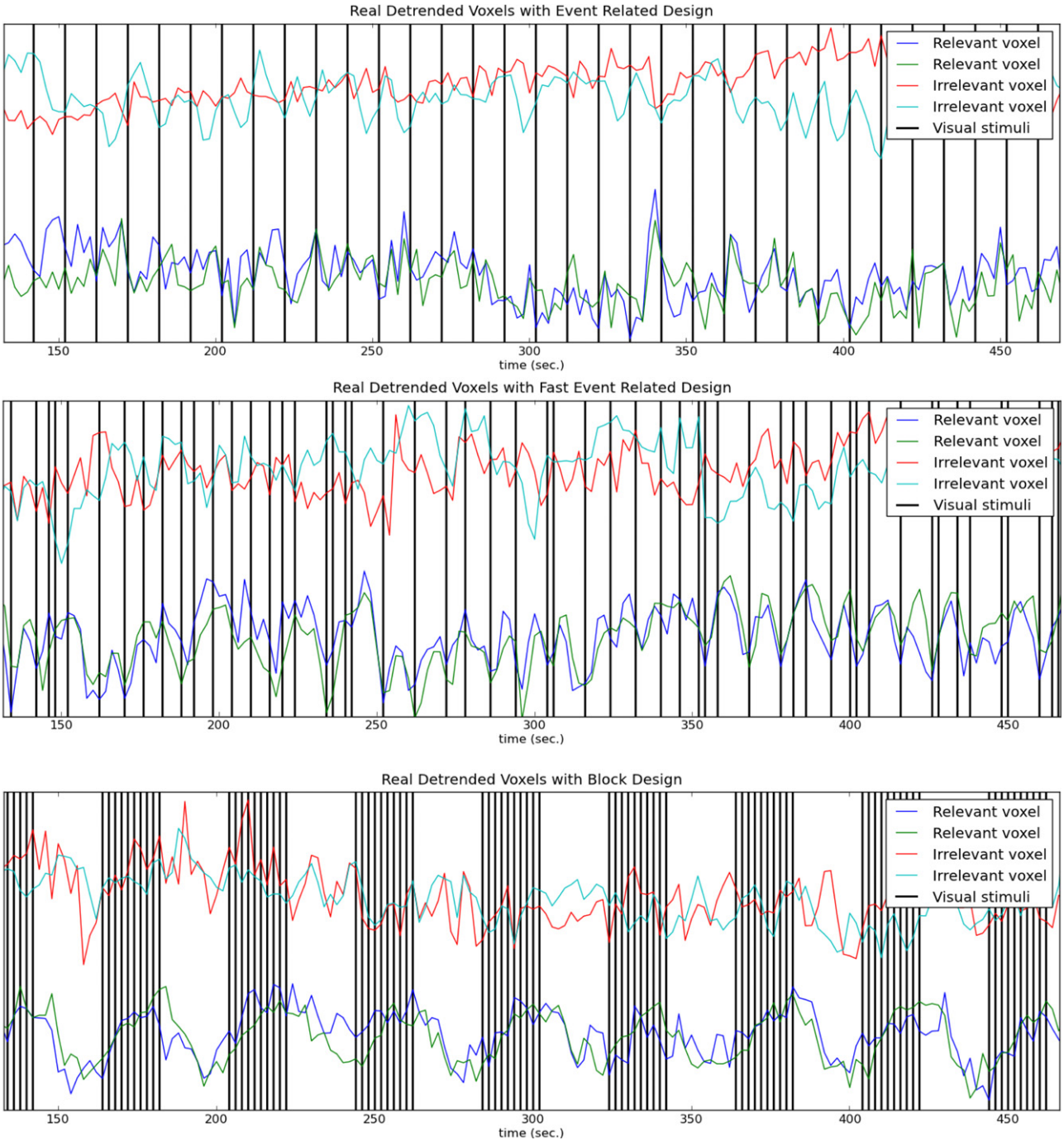

Fig. 5. Snapshot of some real voxels over a short interval of time. Two irrelevant and two correlated voxels are depicted. The vertical bars indicate the times when the visual stimuli were presented to the subject. The two groups of voxels are y-shifted to make the visualization more comprehensible. (In fact, the absolute mean signal cannot be used to discriminate between the correlated and the irrelevant voxels.)

despite the typical low signal-to-noise ratio. The investigation on synthetic data focused on the assessment of model robustness versus both the noise and the variations of the HRF shape. The goal on the real data was, on the other hand, to evaluate whether accurate supervised learning of BOLD response allows for the recognition of relevant voxels in agreement with the discriminatory scores obtained by standard GLM analysis.

4.1. Synthetic data with varying stimulation protocols

The experiment performed on synthetic data with varying stimulation protocols explores the ability of the model to process data with different levels of noise corrupting the BOLD signal. The results of this experiment, as seen in Tables 2 and 3, demonstrated a good tolerance to noise of the proposed system.

Tables 2 and 3 present respectively the average Pearson correlation and the average RMSD between the BOLD activities generated by the system on the test set and the original noisy BOLD activities. These results are shown for the two groups of relevant and irrelevant voxels according to the protocol designs and for two levels of noise (low noise, $\sigma = 0.1$, and large noise, $\sigma = 0.5$).

Independently of the voxel type, either reacting to class A or B or both, the results appear to be similar. Indeed, independently of noise level, higher correlation values are associated with

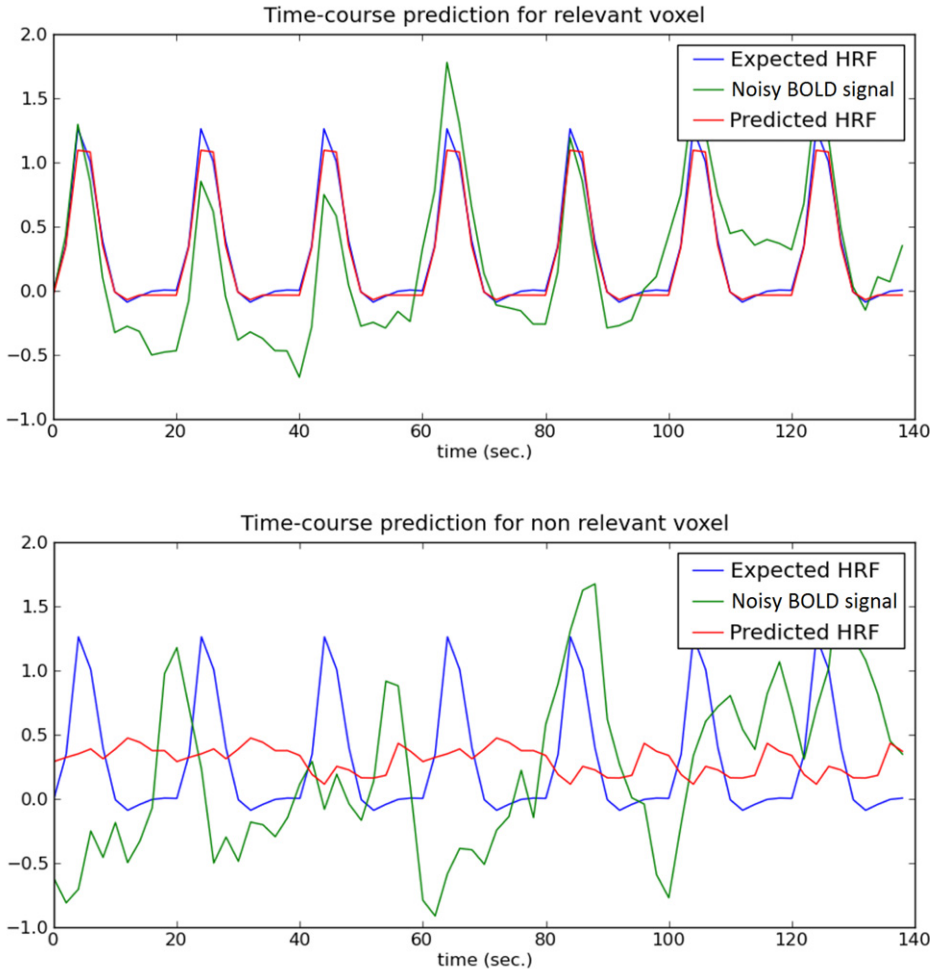

Fig. 6. An example of a BOLD time-series on a test fold of a synthetic event related dataset. The upper graph depicts the noisy (green line), the ground truth (blue line), and the predicted (red line) BOLD signal for a relevant voxel. The lower graph depicts the same time-series for an irrelevant voxel. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Correlation values for synthetic data with the various stimulation protocols. High correlations are associated with voxels relevant for the stimuli; low correlations are associated with voxels irrelevant for the stimuli.

Design	Correlation			
	$\sigma = 0.1$		$\sigma = 0.5$	
	Relevant	Irrelevant	Relevant	Irrelevant
Event related	0.71 ± 0.04	0.06 ± 0.03	0.29 ± 0.04	0.08 ± 0.06
Fast event related	0.42 ± 0.07	0.08 ± 0.04	0.17 ± 0.04	0.11 ± 0.08
Block	0.72 ± 0.05	0.05 ± 0.04	0.38 ± 0.05	0.09 ± 0.03

Table 3
RMSD values for synthetic dataset with the various stimulation protocols. RMSD values are lower for voxels relevant for the stimuli than for voxels irrelevant for the stimuli.

Design	RMSD			
	$\sigma = 0.1$		$\sigma = 0.5$	
	Relevant	Irrelevant	Relevant	Irrelevant
Event related	0.64 ± 0.18	1.37 ± 0.10	1.13 ± 0.12	1.41 ± 0.10
Fast event related	1.12 ± 0.12	1.31 ± 0.10	1.29 ± 0.10	1.37 ± 0.09
Block	0.38 ± 0.12	0.77 ± 0.06	0.64 ± 0.09	1.03 ± 0.04

voxels relevant for stimuli; conversely, RMSD values are lower for relevant voxels. In any case, it is possible to discriminate between relevant and irrelevant voxels under either measure.

Fig. 6 shows the time course for typical examples of response to a stimulus of the proposed model, one for a relevant and one for an irrelevant voxel. The graph depicts the signals generated by the model as well as the original HRF and the noisy BOLD time-course on a test set. It can be seen that the model accurately predicts the underlying HRF of the stimuli-related voxel (the reconstructed signal almost overlaps the clean voxel time-course). In contrast, the model was not able to find any relationship between stimuli and the BOLD signal of irrelevant activity, where the signal was generated so to be unrelated to the underlying HRF.

The encouraging results show that the model generalizes the underlying HRF with robustness to noise, while it behaves as a null model in the absence of any relationship between the BOLD time-course and the sequence of stimuli. Thus, the implemented reservoir computing model appeared to be a very good de-noising algorithm for the synthetic data.

4.2. Synthetic data with varying HRF shapes

The experiments performed on synthetic data with varying HRFs (see Fig. 4 and Table 4) investigated the ability of the proposed model to adapt to unusual HRF shapes. The results revealed good ability to model all the shapes we challenged it with.

A decrease in prediction accuracy occurred as the noise level increases, irrespective of the particular HRF shape. Despite this variability, values produced with voxels relevant for the stimuli (as measured either by the correlation or the RMSD measure) still differed significantly from those produced with voxels irrelevant for the stimuli. The detailed results of this experiment are shown

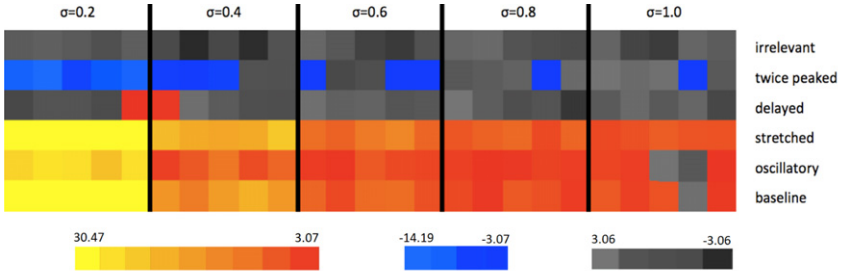

Fig. 7. t-values map for synthetic data with HRF variations determined by a standard GLM analysis. Each row corresponds to one HRF variation. (The ground truth has all rows except the first positively correlated to the stimulus.) Columns are organized in groups of five, with increasing AR noise. The colored t-map was thresholded at the highest t-value appearing in the top row (irrelevant voxels). This value was 3.07. Warm colors identify high significant t-values while cold colors identify the features with a significant but negative t-value. The gray pixels are those whose t-value has absolute value below 3.07. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Correlation and RMSD values for synthetic data with varying HRF shape. All relevant signals are reconstructed successfully; high correlation values are associated with voxels relevant for the stimuli even in presence of the substantial noise; low correlation values are associated with voxels irrelevant for the stimuli.

HRF shape	Correlation		RMSD	
	$\sigma = 0.2$	$\sigma = 1.0$	$\sigma = 0.2$	$\sigma = 1.0$
Baseline	0.56 ± 0.02	0.15 ± 0.02	0.84 ± 0.03	0.99 ± 0.11
Oscillatory	0.54 ± 0.03	0.16 ± 0.04	0.87 ± 0.07	0.96 ± 0.05
Stretched	0.58 ± 0.01	0.15 ± 0.11	0.83 ± 0.05	0.99 ± 0.08
Delayed	0.57 ± 0.04	0.14 ± 0.07	0.84 ± 0.08	0.97 ± 0.06
Twice peaked	0.55 ± 0.04	0.14 ± 0.02	0.85 ± 0.06	0.99 ± 0.07
Irrelevant	$\mathbf{0.01 \pm 0.07}$	$\mathbf{0.05 \pm 0.07}$	$\mathbf{1.39 \pm 0.24}$	$\mathbf{1.43 \pm 0.20}$

in Table 4 for the extreme ranges only, i.e., the lowest ($\sigma = 0.2$) and the highest ($\sigma = 1$) noise levels. For all intermediate noise levels the results are, as expected, in between the extremes.

These experimental results reveal that the proposed system is able to identify the underlying hemodynamics even when it is substantially different from the classical HRF used in the GLM analysis method and affected by high degrees of noise.

This result is even more striking when we compare with that of a standard GLM analysis performed on the data which substantially failed to retrieve both the voxels with the "delayed" HRF and the voxels with the "twice peaked" HRF, as well as having problems for other voxels with high noise.[8]

Fig. 7 shows a graphical depiction of the results of the GLM analysis via a map of the voxels' significance (as is typically used in brain activity maps). Each square represents the result of the GLM analysis on a generated time-series; each row refers to a specific underlying type of HRF; the noise levels are indicated by σ levels and the squares are organized into groups of similar noise (indicated by σ values). Gray levels indicate t-values below a threshold (determined as the maximum absolute t-value of the irrelevant voxel (the maximum in the first row)). Color values indicate signals with absolute t-values exceeding this threshold: red to yellow for positives values and blue levels for negative ones. Note that the GLM analysis wrongly produces negative t-values for "twice peaked" variables although they are positively correlated to the stimuli.

Table 5 gives the detailed numerical results comparing the proposed method and the standard GLM analysis for the two extreme noise levels. LSM successfully reconstructs all relevant signals (high correlation and low RMSD), even in presence of substantial noise ($\sigma = 1.0$). Low correlation and high RMSD values are instead associated with voxels irrelevant for the stimuli. GLM on the contrary has problems with relevant signals with "delayed" and "twice peaked" variables; moreover as the noise increases the other variables become more difficult to discriminate.[9] In short, the GLM method is biased by the assumption of the underlying model of the HRF. Our system avoids this assumption and this is perhaps its greatest advantage. While a good fitting of the model can be achieved by GLM taking advantage of the matched filter theorem, that method requires knowing in advance some parameters describing the underlying HRF. However the dependency of such parameters from the design protocol and from the specific brain area very often is unknown. These results indicate that the liquid state machine system can be successfully used for discovering unexpected relations between the stimuli and the underlying BOLD signal. Using this method, additional voxels that present uncommon relations with the stimuli can be identified, and thus this method has the potential of improving brain mapping technologies.

4.3. Real data

The above results hold for our real fMRI data as well. Table 6 presents the average Pearson correlation and the average RMSD between the BOLD activities generated by the system given the time-series of stimuli on the test set and the corresponding original BOLD activity. Indeed, the correlation between the time series generated by the reservoir-based model on the test dataset and the corresponding real BOLD signal (Table 6) shows the model's capability to distinguish the relevant voxels (as detected by GLM) from the irrelevant ones. These results provide evidence that the proposed method does not lose important information.

On the other hand, our results on real data do not (yet) say anything one way or the other about the abilities (as revealed on the synthetic data) to better discriminate when the underlying HRF is non-standard. It is interesting to note that in the experiment with the event related protocol, a few voxels marked irrelevant with the GLM method were identified as relevant by the proposed method. Further analysis of the reconstructed HRFs showed a delay to time peak of around 0.5–2 s. The discovery of such voxels should, in principle, enable more accurate brain maps.

Fig. 8 shows, in parallel to Fig. 6, two generalization examples for a relevant and an irrelevant voxel in a block design experiment, depicting for a small portion of a test set the real BOLD signal

[8] The GLM analysis was performed with the FMRIB Software Library, FSL 4.1 (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012), using FEAT and setting the shape of the expected HRF for convolution as a double-gamma. To increase the robustness to shape shifts in GLM analysis we also allowed temporal first derivatives in the design matrix, as suggested by Friston et al. (1998).

[9] While it might be possible to extend the GLM analysis to handle these specific cases (at least without the noise) by extending the basis or using a higher derivative expansion, this is not a general methodology (Handwerker, Gonzalez-Castillo, D'Esposito, & Bandettini, 2012), and there could be other signals missed.

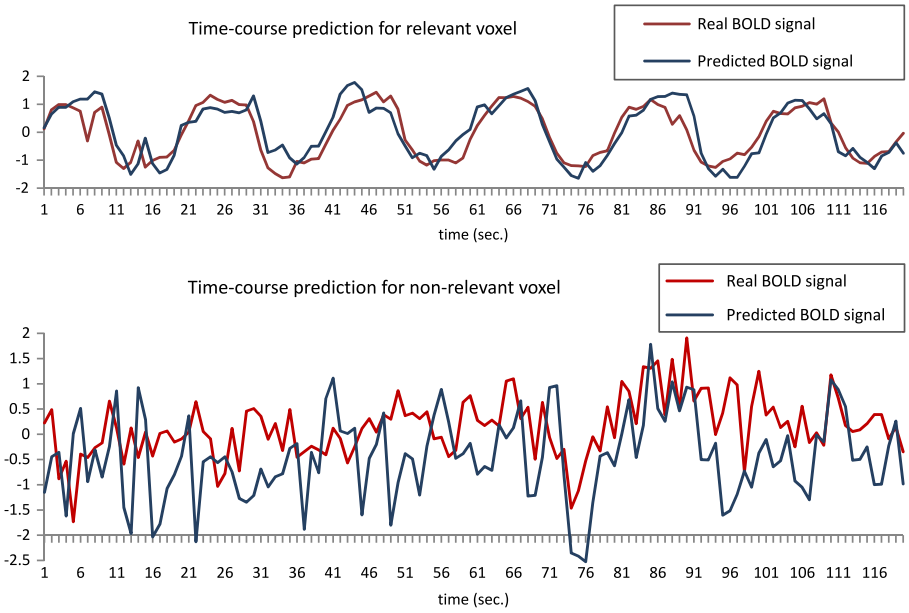

Fig. 8. Reconstructed BOLD time-series for a real dataset. Voxel hemodynamic response obtained with Block design, for relevant voxels (top) and irrelevant voxels (bottom). The real BOLD signal is shown in red; the generated BOLD signal is in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Correlation and root mean squared deviation indicating the quality of reconstruction obtained with LSM, and the t-values determined with GLM indicating the relevance of the variables for the task. The values are shown for the two σ levels (0.2 and 1.0). Red color is for voxels recognized as irrelevant or non-discriminatory by the corresponding method.

HRF Shape	$\sigma = 0.2$			$\sigma = 1.0$		
	LSM		GLM	LSM		GLM
	Correlation	RMSD	t-values	Correlation	RMSD	t-values
Baseline	0.56 ± 0.02	0.84 ± 0.03	22.89 ± 1.60	0.28 ± 0.04	0.99 ± 0.11	4.81 ± 1.79
Oscillatory	0.54 ± 0.03	0.87 ± 0.07	16.95 ± 1.10	0.35 ± 0.01	0.96 ± 0.05	3.08 ± 1.84
Stretched	0.58 ± 0.01	0.83 ± 0.05	28.57 ± 1.59	0.23 ± 0.01	0.99 ± 0.08	6.38 ± 0.53
Delayed	0.57 ± 0.04	0.84 ± 0.08	$\mathbf{2.15 \pm 0.89}$	0.22 ± 0.04	0.97 ± 0.06	$\mathbf{0.95 \pm 1.10}$
Twice peaked	0.55 ± 0.04	0.85 ± 0.06	$\mathbf{-6.83 \pm 1.24}$	0.27 ± 0.14	0.99 ± 0.07	-2.18 ± 0.57
Irrelevant	$\mathbf{0.01 \pm 0.07}$	$\mathbf{1.39 \pm 0.24}$	$\mathbf{-0.70 \pm 1.55}$	$\mathbf{0.05 \pm 0.07}$	$\mathbf{1.43 \pm 0.20}$	$\mathbf{-0.72 \pm 0.73}$

Table 6
Results for a real dataset. For all protocols, the correlation values associated with voxels relevant for the stimuli are easily distinguishable from those associated with voxels irrelevant for the stimuli.

Design	Correlation		RMSD	
	Relevant	Irrelevant	Relevant	Irrelevant
Event related	0.35 ± 0.06	0.06 ± 0.04	0.97 ± 0.8	1.16 ± 0.18
Fast event related	0.28 ± 0.04	0.10 ± 0.04	0.98 ± 0.6	1.20 ± 0.22
Block	0.54 ± 0.06	0.09 ± 0.04	0.84 ± 0.5	1.34 ± 0.08

and the one generated by the model. The two curves show a good matching only in the case of a voxel being relevant for the set of stimuli.

The relevance or irrelevance of voxels can be detected by simply thresholding the correlation between the predicted and observed behavior. Doing this resulted in a very high performance both for real data (100% of voxels correctly determined) and for synthetic BOLD signal (97% of voxels). These results, obtained for all three experiments, suggest that brain maps could be generated with data-driven approaches, without requiring any prior knowledge on the expected HRF.

5. Conclusions and potential future work

The results obtained on the synthetic data show that reservoir computing method instantiated with a LSM for the reservoir and with MLPs for the detectors is an effective tool for the analysis of BOLD time-sequences acquired during fMRI experiments. Actually, the model generalizes the HRF underlying the BOLD signal for all relevant voxels with a low sensitivity to noise. On the other hand, when the voxels are unrelated to the sequence of stimuli, the model does not generalize to the test data. This dichotomy allows the discrimination between relevant and irrelevant voxels.

Thus, when applying the method to real fMRI data, the reservoir-based model seems to be able to extract the relevant information from the relevant voxels. Indeed, virtually all voxels that were retrieved as relevant by the standard GLM-based analysis were found relevant with the proposed approach as well. Moreover, some voxels that were undetected as relevant by the standard GLM method were chosen as relevant by the presented method. Fig. 9 illustrates an example where the reproduced HRF is somehow similar to a standard HRF usually adopted, with a further delayed time to peak of around 2 s. Such delays, even if not necessarily related to neural activity, were shown to be physiologically plausible (Handwerker et al., 2012). For example, such delay could be a result of neurovascular variations caused by astrocytic processes (Schummers, Yu, & Sur, 2008).

Thus our hypothesis is that these voxels as well should not be neglected. Even if at this time we cannot verify this hypothesis as our work did not supply us with absolute "ground truths" on the real experiments our successful reconstruction of the artificial HRF helps to empirically support the hypothesis.

Table 7
The Balloon model parameters used to generate various HRF shapes. Empty cells are equivalent to standard values.

	Standard	Oscillatory	Stretched	Delayed	Twice peaked
Signal decay (τ_s)	1.5	3.0	3.0		
Autoregulation (τ_f)	2.4		4.0		
Venous time constant ($mb\tau_0$)	1.0		4.0	2.0	
Neuronal efficacy (ε)	0.5		0.1	0.13	

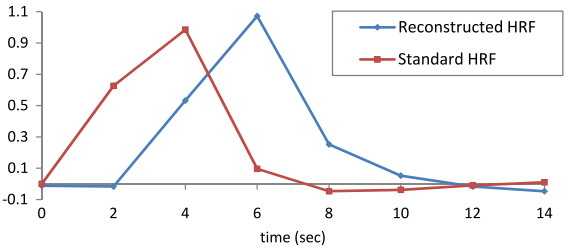

Fig. 9. An example of a real voxel as (red) reproduced by a standard HRF and as (blue) using the LSM mechanism as described in this paper. Note that the peak is off-set to the right similarly to examples on synthetic data.

The experiments on synthetic data also demonstrated that the proposed method discovers the relational effects between the stimuli and the BOLD signals even when the underlying HRF shape varied significantly from the canonical HRF models. This was done autonomously by the system, without any additional effort required for fitting the parameters and selecting the right HRF shape. This characterizes the method as potentially useful for creating more accurate brain maps compared to the conventional hypothesis-driven methods, considering the areas with the underlying hemodynamics that are different from those covered by the conventional HRF shapes.

The next step is to include the entire brain volume in the analysis. The datasets used in the current study were limited in a number of aspects, e.g., they were constructed from a small number of voxels (compared to the entire volume) and contained relatively short sequences. The goal is to apply these methods for creating contrast brain maps, where, in addition to finding voxels relevant for the specific stimulus, one wants to find the voxels pertinent to the contrast between stimuli. Thus, for a sequence including stimuli A and B, one wants to find the voxels related to the stimulus A and not related to the stimulus B and vice versa. The proposed method can in principle be applied to these complex fMRI data collected for multiple stimuli, without the necessity to split the entire input sequence into several sub-series, each composed of two contrasting states of interest. This is reflected by the results obtained with synthetic dataset presenting stimuli of two classes, where all voxels related to at least one of the two stimuli were determined as relevant using the same liquid.

An important issue that needs further development involves the definition of an appropriate threshold for deciding if the voxel is relevant or not relevant. Several possible measures ("Correlation between predicted and observed", "RMSD between prediction and observed") have been presented here. In addition, we used cross-validation as an effective precaution against both false-positive and false-negative results. Using this schema, we showed that over the 100 sampled examples (chosen with a ground truth from GLM methodologies, or chosen based on the synthetic actual ground truth) the classification results were very good. Additional measures, might include parametric techniques following the same rules adopted for standard GLM-based analysis (e.g., t-statistics on correlation values, including a posterior correction for multiply comparison). Most of these additions, however, are pertinent only when analyzing a large number of voxels (e.g. a full-brain analysis) and will be appropriate for future work.

Amongst the known limitations of the proposed method, we may mention its inability to distinguish automatically a physiologically plausible BOLD signal from noisy BOLD signal changing periodically with frequency equal to the stimulus frequency. This issue may be resolved, for example, by a comparison of BOLD signal predicted by LSM model for two different types of stimuli and discarding voxels with similar predicted signals.

As mentioned above, the *LSM* configuration used in the current study was based on random constant weights that did not undergo any training phase. It might be useful to apply one of the plasticity related training methods (Legenstein, 2005; Wade, McDaid, Santos, & Sayers, 2010) tuning the *LSM* for dealing with the BOLD-related kind of data. A predefined set of pre-collected fMRI data could be used for such training.

In conclusion, the current study introduced a new approach for the analysis of fMRI data, based on a machine learning technique that exploits a neuro-inspired method to emulate the brain behavior.

6. Software

The software is developed in C# programming language and based on existing LSM (Hazan & Manevitz, 2012) and feed-forward (Encog, 2009) software libraries.

The software to generate the synthetic data is partially based on Python.

Acknowledgments

The authors would like to thank Emanuele Olivetti (NILab, Fondazione Bruno Kessler, Trento, Italy) for the implementation of Balloon model used in synthetic data production. We thank the Caesarea Rothschild Institute of the University of Haifa and the Fondazione Bruno Kessler, Trento, for partial support of this work. The authors are listed in alphabetical order. Some of this work serves as part of the M.Sc. thesis in computer science of Ester Koilis under the advisorship of Larry Manevitz at the Neurocomputation Laboratory located at the Caesarea Rothschild Institute at the University of Haifa.

Appendix

The parameters used to generate standard HRF with the Balloon model were selected following the guidelines in Zheng et al. (2002). For all other uncommon HRF some parameters have been changed. Table 7 shows only the parameters that have been used to synthetically generate the uncommon hemodynamics. The empty cells indicate a parameter assignment identical to the standard HRF.

References

Aguirre, G., Zarahn, E., & D'Esposito, M. (1998). The variability of human BOLD hemodynamic responses. *Neuroimage*, *8*, 360–369.
Albert, R., & Barabasi, A. (2000). Topology of evolving networks: local events and universality. *Physical Review Letters*, *85*(24), 5234–5237.
Atir-Sharon, T., Gilboa, A., Hazan, H., Koilis, E., & Manevitz, L. (2015). Decoding the formation of new semantics: MVPA investigation of rapid neocortical plasticity during associative encoding through Fast Mapping. *Neural Plasticity* (to appear).

Avesani, P., Hazan, H., Koilis, E., Manevitz, L., & Sona, D. (2011). Learning BOLD response in fMRI by reservoir computing. In *PRNI, IEEE international workshop on pattern recognition in neuroimaging* (pp. 57–60).

Bandettini, P., & Cox, R. (2000). Event-related fMRI contrast when using constant interstimulus interval: Theory and experiment. *Magnetic Resonance in Medicine*, *43*, 540–548.

Boehm, O., Hardoon, D. R., & Manevitz, L. (2011). Classifying cognitive states of brain activity via one-class neural networks with feature selection by genetic algorithms. *International Journal of Machine Learning and Cybernetics*, *2*(3), 125–134.

Buckner, R., Bandettini, P., Craven, K., Savoy, R., Petersen, S., Raichle, M., et al. (1996). Detection of cortical activation during. *Proceedings of the National Academy of Sciences*, *93*, 14878–14883.

Buonomano, D., & Maass, W. (2009). State-dependent computations: Spatiotemporal processing in cortical networks. *Nature Reviews in Neuroscience*, *10*(2), 113–125.

Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, *2*, 121–167.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) 'brain reading': detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, *19*, 261–270.

Encog (2009). Encog artificial intelligence framework for Java and DotNet. Retrieved from http://www.heatonresearch.com/.

Friston, K. J., Josephs, O., Rees, G., & Turner, R. (1998). Non-linear event-related responses in fMRI. *Magnetic Resonance in Medicine*, *39*, 41–52.

Friston, K., Holmes, A. P., Worsley, K. J., Poline, J., Frith, C. D., & Frackowiak, R. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*(4), 189–210.

Goutte, C. A., Nielsen, F. A., & Hansen, L. K. (2000). Modeling the haemodynamic response in fMRI using smooth FIR filters. *IEEE Transactions on Medical Imaging*, *19*(12), 1188–1201.

Grinbald, J., Wager, T., Lindquist, M. F., & Hirsch, J. (2008). Detection of time-varying signals in event-related fMRI designs. *Neuroimage*, *43*, 509–520.

Handwerker, D., Gonzalez-Castillo, J., D'Esposito, M., & Bandettini, P. (2012). The continuing challenge of understanding and modeling hemodynamic variation in fMRI. *NeuroImage*, *62*, 1017–1023.

Hardoon, D., & Manevitz, L. (2005). fMRI analysis and compression neural networks. In *International joint conference in artificial intelligence* (pp. 1604–1606).

Hazan, H. (2014). *Temporal classification and computation tools inspired by biological neurons*. (Ph.D. thesis), Unpublished.

Hazan, H., & Manevitz, L. M. (2012). Topological constraints and robustness in liquid state machines. *Expert Systems with Applications*, *39*(2), 1597–1606.

Henson, R., & Friston, K. (2007). Convolution models for fMRI. In K. Friston, J. Ashbumer, & S. Kiebel (Eds.), *Statistical parametric mapping: The analysis of functional brain images* (pp. 178–192). London: Elsevier.

Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, *304*(5667), 78–80.

Jenkinson, M., Beckmann, C., Behrens, T., Woolrich, M., & Smith, S. (2012). FSL. *NeuroImage*, *62*, 782–790.

Josephs, O., & Henson, R. (1999). Event-related fMRI: modelling, inference and optimisation. *Philosophical Transactions of the Royal Society, London*, *354*, 1215–1228.

Josephs, O., Turner, R., & Friston, K. (1997). Event-related fMRI. *Human Brain Mapping*, *5*, 243–248.

Lapicque, L. (1907). Recherches quantitatives sur l'excitation e'lectrique des nerfs traite'e comme une polarization. *Journal de Physiologie et de Pathologie Générale*, *9*, 620–635.

Legenstein, R. (2005). What can a neuron learn with spike-timing-dependent plasticity? *Neural Computation*, *17*(11), 2337–2382.

Maas, W. (2010). Liquid state machines: Motivation, theory, and applications. In B. Cooper, & A. Sorbi (Eds.), *Computability in context: computation and logic in the real world* (pp. 275–296). Imperial College Press.

Maass, W., Natschlager, T., & Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Computation*, *14*(11), 2531–2560.

Mitchell, T., Hutchinson, R., Niculescu, R., Pereira, F., Wang, X., Just, M., et al. (2004). Learning to decode cognitive states from brain images. *Machine Learning*, *57*, 145–175.

Monti, M. (2011). Statistical analysis of fMRI time-series: A critical review of the GLM approach. *Frontiers in Human Neuroscience*, *5*(28), http://dx.doi.org/10.3389/fnhum.2011.00028.

Mourão-Miranda, J., Bokde, A., Born, C., Hampel, H., & Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data. *NeuroImage*, *28*(4), 980–995.

Nguyen, D., & Widrow, B. (1990). Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In *Proc. of the int. joint conference on neural networks*, 3 (pp. 21–26).

Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In *IEEE international conference on neural networks* (pp. 586–591).

Schummers, J., Yu, H., & Sur, M. (2008). Tuned responses of astrocytes and their influence on hemodynamic signals in the visual cortex. *Science*, *320*(5883), 1638–1643.

Vapnik, V. N. (1995). *The nature of statistical learning theory. Statistics for engineering and information science* (2nd Ed.). New York: Springer-Verlag.

Verstraeten, D., Schrauwen, B., D'Haene, M., & Stroobandt, D. (2007). An experimental unification of reservoir computing methods. *Neural Networks*, *20*, 391–403.

Wade, J. J., McDaid, L. J., Santos, J. A., & Sayers, H. (2010). SWAT: A spiking neural network training algorithm for classification problems. *IEEE Transactions on Neural Networks*, *21*(11), 1817–1830.

Wang, Z. (2009). A hybrid SVM-GLM approach for fMRI data analysis. *Neuroimage*, *46*(3), 608–615.

Watts, D., & Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440–442.

Wink, A., Hoogduin, H., & Roerdink, J. B. (2008). Data-driven haemodynamic response function extraction using Fourier-wavelet regularised deconvolution. *BMC Medical Imaging*, *8*(7).

Woolrich, M., Jenkinson, M., Brady, J., & Smith, S. (2004). Fully Bayesian spatio-temporal modeling of fMRI data. *IEEE Transactions on Medical Imaging*, *23*(2), 213–231.

Zheng, Y., Martindale, J., Johnston, D., Jones, M., Berwick, J., & Mayhew, J. (2002). A model of the hemodynamic response and oxygen delivery to brain. *NeuroImage*, *16*(3), 617–637.