MEASURES OF GOODNESS OF FIT TO CONVOLUTION MODEL FOR ANALYSIS OF FMRI DATA

Wakako Nakamura, Yujiro Inouye

Faculty of Science and Engineering, Shimane University
Nishikawatsu 1060, Matsue, Shimane 6908504, Japan
phone, fax: +81-852-32-6065, email: wakako@riko.shimane-u.ac.jp

Lab. for Advanced Brain Signal Processing, Brain Science Institute, RIKEN
Hirosawa 2-1, Wako 3510198, Saitama, Japan

ABSTRACT

On an fMRI data analysis, it is common to assume that we know when stimuli were presented or when subjects performed a task. However, for mental tasks such as memory retrieval, we cannot obtain an exact time of the task execution. When we use complex stimuli or natural stimuli such as a movie in experiments, then sometimes we cannot define the presentation time of stimuli straightforwardly. For these cases, we propose measures of a neural activity that we can obtain without a time series of stimuli presentations or task executions. We apply a blind deconvolution algorithm to an fMRI data set and separate it into a Hemodynamic Response Function (HRF) and a series of presentation times of stimuli. We propose to use values of the cost function for this separation algorithm as measures of a neural activity. The cost function is consisted of two terms. One is an error term representing discrepancy from a conventional convolution model of fMRI. The other term represents statistical characteristics of the estimated stimuli presentation time series.

1. INTRODUCTION

In fMRI data analyses, model-based methods are most commonly used. For natural stimuli [1] or complex stimuli, often time series of stimuli cannot be obtained or there are many components of interest in stimuli. It is pointed out that a data-driven method is necessary for these experiments. Some data-driven methods have been proposed, such as, Independent Component Analysis (ICA) [2], clustering based on time series [3] and a canonical correlation analysis approach [4]. Here, we propose a novel data-driven analysis method for fMRI data.

A series of fMRI data is often modeled by the convolution of the hemodynamic response function (HRF) and a time series of presentations of stimuli. Usually we assume that we can obtain both the HRF and the time series of stimuli. If the fMRI data from the voxel is close to the convolution of them, we conclude that the voxel is related to the processing of the stimuli. When we cannot obtain a true time series of stimuli, we can use a blind signal processing technique to estimate the time series from data. In that case, in addition to the convolution model, we put an assumption about some statistics of the time series of stimuli and constitute a generative model of the fMRI data. A measure of goodness of fit to this generative model is used as a cost function to estimate the time series. In our method, we propose to use this measure to judge whether some neural activity occurres in the voxel or not. For an assumption about statistics of the time series of stimuli, we put two assumptions, a sparse distribution and a Gaussian distribution.

Since we use a blind signal separation algorithm, an HRF either can be estimated from the data or can be assumed a priori. Considering the difference of HRFs from area to area or from subject to subject [5], we estimate them from the data.

We applied the proposed algorithm to fMRI data when visual stimuli are presented to a subject. We used the standard analysis of variance to classify voxels into activated regions and deactivated regions. Then we examine whether the proposed measures take different values between data from voxels with neural activities and data from voxels without them.

Blind deconvolution of fMRI data was proposed by Hansen (2003) [6]. They separated multi-voxel data into neural activities, noise and artifacts and separation into an HRF and a stimuli time series was not examined in detail.

Hutchinson et al. (2009) [7] proposed a method called Hidden Process Models which estimates both an HRF and a time series of stimuli from a set of fMRI data. In most cases, they put stronger assumptions on presentation times of stimuli than our method and mainly studied multivariate analyses.

2. BLIND DECONVOLUTION ALGORITHM

We use a blind deconvolution algorithm derived in the same way as Olshausen and Field (1996) [8]. $x(t)$ denotes fMRI data measured at a certain voxel where an index for time is t. $s(t)$ denotes a time series of stimuli which typically takes 1 when a stimulus is presented and takes 0 at other times. A waveform induced on fMRI data by a single stimuli, that is an HRF is represented by a_τ. Then, fMRI data is assumed to be generated as follows.

$$x(t) = \sum_{\tau=1}^{\tau_{max}} a_\tau s(t-\tau+1) + \varepsilon(t) \qquad (1)$$

$\varepsilon(t)$ denotes noise and it is assumed to be normally distributed and white.

In our analyses, we allow $s(t)$ to take any real values instead of binary values of 0 and 1. Positive values can represent strengths of activations induced by stimuli. Negative values cannot be interpreted as stimuli presentations and unnatual. One way to avoid this is to add a constant value to $x(t)$ for making it positive and to assume that both $s(t)$ and a_τ are positive. Our analysis of fMRI data follows this method

partly. We make most parts of $x(t)$ positive and allow $s(t)$ and a_τ to take negative values. This can decrease effects of small outliers in $x(t)$ compared with an analysis with perfectly positive $x(t)$.

Here we consider only one waveform of HRF a_τ per a voxel. In this paper, we mainly describe voxelwise analysis methods. Therefore, $s(t)$ can be defined voxelwise also. If one voxel is activated by only one kind of stimulus or task, eq. (1) can be interpreted simply. The time series $s(t)$ can be a mixture of different kinds of stimuli or tasks. In this model, still, a_τ is assumed to be the same single waveform for the voxel. Differences in magnitudes, durations and time lags of the HRF can be modeled in $s(t)$. The additive noise on fMRI data is known not to be white and the whiteness assumption of the model is not accurate. However, the whiteness assumption appears in our proposed algorithm only as a minimum square error principle and it dose not force the residual noise to be white. Different HRF waveforms for different kinds of stimuli or an additive noise with a time correlation should be modeled in future studies. Here we study behaviors of the simplified model.

The distribution of $s(t)$ was assumed to be sparse in Olshausen and Field (1996) [8]. Here we employ two assumptions, a sparse distribution and a normal distribution.

We assume eq. (1) and the presentation times of stimuli distribute sparsely or normally. We estimate a_τ and $s(t)$ from the data by minimizing the following cost functions to retrieve the time series of stimuli.

$$L_s = \sum_t (x(t) - \sum_\tau a_\tau s(t-\tau+1))^2 + \lambda_s \sum_t \log(1+\beta s(t)^2) \tag{2}$$

$$L_n = \sum_t (x(t) - \sum_\tau a_\tau s(t-\tau+1))^2 + \lambda_n \sum_t s(t)^2 \tag{3}$$

For L_s, the cost function can be interpreted as a likelihood given data $x(t)$. The probabilistic distribution of $s(t)$ is a Cauchy distribution [8] which is a sparse distribution and the distribution of $x(t) - \sum_\tau a_\tau s(t-\tau+1)$ is a normal distribution. This is a kind of the maximum a posteriori estimation. For L_n, the distribution of $s(t)$ is a normal distribution and other properties are the same as L_s.

First we determine initial values of a_τ and $s(t)$ either randomly or based on a priori information of characteristics of these variables. The processes for updating a_τ and updating $s(t)$ are iterated. When we estimate both a_τ and $s(t)$ from data, we cannot determine magnitudes of these variables. Therefore, we assume that $\sum_{\tau=1}^{\tau_{\max}} a_\tau^2 = 1$.

First, we fix $s(t)$ to tentative values and update a_τ. This process is common to L_s and L_n. a_τ that minimizes the cost function L_s or L_n under the constraint $\sum_{\tau=1}^{\tau_{\max}} a_\tau^2 = 1$ can be obtained from the following equation obtained by using the Laglange multiplier method,

$$\frac{\partial L}{\partial a_u} + 2\mu a_u = 2\sum_t (\sum_\tau a_\tau s(t-\tau+1) - x(t)) s(t-u+1) + 2\mu a_u = 0, \qquad u = 1,2,\cdots \tau_{\max} \tag{4}$$

The equation can be written by using vectors and a matrix. Here, $\boldsymbol{a} = (a_1, a_2, \cdots, a_{\tau_{\max}})^T$ and the k-th component of the $\tau_{\max}$ dimensional vector $\boldsymbol{c}_{sx}$ is

$$\{\boldsymbol{c}_{sx}\}_k = \sum_t x(t) s(t-k+1) \tag{5}$$

and the kl-component of the $\tau_{\max}$-by-$\tau_{\max}$ matrix C_{ss} is

$$\{C_{ss}\}_{kl} = \sum_t s(t-k+1) s(t-l+1). \tag{6}$$

Then eq. (4) can be written as follows.

$$\frac{\partial L}{\partial \boldsymbol{a}} + 2\mu \boldsymbol{a} = 2C_{ss}\boldsymbol{a} - 2\boldsymbol{c}_{sx} + 2\mu \boldsymbol{a} = 0$$

Multiplying both sides by $\boldsymbol{a}^T$, we obtain

$$\mu = -\boldsymbol{a}^T C_{ss} \boldsymbol{a} + \boldsymbol{a}^T \boldsymbol{c}_{sx} \tag{7}$$

The second derivative is also calculated to derive an update rule from the Newton method.

$$\frac{\partial}{\partial a_v}\left(\frac{\partial L}{\partial a_u} + 2\mu a_u\right) = 2\sum_t s(t-v+1)s(t-u+1) + 2\mu\delta_{uv} \tag{8}$$

where δ_{uv} is a Kronecker delta. It can be written as

$$\frac{\partial}{\partial \boldsymbol{a}}\left(\frac{\partial L}{\partial \boldsymbol{a}} + 2\mu \boldsymbol{a}\right) = 2C_{ss} + 2\mu I \tag{9}$$

where I denotes an identity matrix.

Finally the update rule is derived as follows,

$$\begin{aligned} \boldsymbol{a} &\leftarrow \boldsymbol{a} - \left(\frac{\partial}{\partial \boldsymbol{a}}\left(\frac{\partial L}{\partial \boldsymbol{a}} + 2\mu\boldsymbol{a}\right)\right)^{-1}\left(\frac{\partial L}{\partial \boldsymbol{a}} + 2\mu\boldsymbol{a}\right) \\ &= (C_{ss} + (-\boldsymbol{a}^T C_{ss}\boldsymbol{a} + \boldsymbol{a}^T\boldsymbol{c}_{sx})I)^{-1}\boldsymbol{c}_{sx} \end{aligned} \tag{10}$$

$$\boldsymbol{a} \leftarrow \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|} \tag{11}$$

Next, for L_s, we fix a_τ and update $s(t)$ to decrease L simply by the gradient method,

$$s(t) \leftarrow s(t) - c\left(\frac{dL_s}{ds(t)}\right) \tag{12}$$

$$\begin{aligned} \frac{dL_s}{ds(t)} =& 2\sum_u (\sum_\tau a_\tau s(t+u+1-\tau) - x(t+u)) a_u \\ &+ \lambda_s \frac{2\beta s(t)}{1+\beta s(t).^2} \end{aligned} \tag{13}$$

For L_n, we fix a_τ and update $s(t)$ by solving $dL_n/ds(t) = 0$ where

$$\begin{aligned} \frac{dL_n}{ds(t)} =& 2\sum_u (\sum_\tau a_\tau s(t+u+1-\tau) - x(t+u)) a_u \\ &+ \lambda_n 2 s(t) \end{aligned} \tag{14}$$

We iterate a step to update a_τ with fixed $s(t)$ and a step to update $s(t)$ with fixed a_τ until changes of values by update processes are small enough. In this way, we estimate a_τ and $s(t)$ that minimizes L_s or L_n.

3. ANALYSES OF FMRI DATA

We used data from Hasson et al. (2001) [9]. The data set is obtained through the fMRI data center(www.fmridc.org). The accession # is 2-2001-111P8. The original paper is on

processes of local features of objects in vision. Though they showed various kinds of pictures to subjects, here we analyze the data focusing only on whether a subject was seeing an object or a blank screen. We used a data set from a single subject in which activated regions for visual objects compared to a blank screen are clearly identified. TR was 3 sec. In the most part of the experimental period, a pair of a presentation of an object for 9 sec (3 scans) and a presentation of a blank screen for 6 sec (2scans) was repeated. A total of 167 scans was obtained and the first presentation of an object started at the 5-th scan.

The data is analyzed by a standard method using the analysis of variance and the regions activated by presentations of objects were identified. For this analysis and necessary preprocessing for it, we used the SPM software, which can be downloaded from http://www.fil.ion.ucl.ac.uk/spm/ . We chose twenty voxels randomly from activated and deactivated regions and applied the proposed blind deconvolution algorithm with a sparse or a normal prior to the data. A voxel is classified as activated where t value is more than 7.00. It corresponds to uncorrected p-value less than 3.45×10^{-11}. A voxel is classified as deactivated where t value is less than 3.00 which corresponds to uncorrected p-value more than 1.57×10^{-3}.

Before applying the algorithm, we low-pass filtered the data at 0.0167 Hz to remove slow drifting components. Then we normalized each data so that their mean values are 1.3 and their variances are one. With this mean value, about 90% of values in a series are nonnegative. The proposed deconvolution algorithm does not assume explicitly that $s(t)$ is nonnegative. However, it is natural to assume that time series of stimuli take positive values and that HRFs were added to form fMRI data. This is the reason for the positive mean value.

Initial values of a_τ are decided from a common shape of an HRF[10].

$$a_\tau = k(\tau - 1)^{8.60} e^{-\frac{\tau - 1}{0.547}} \tag{15}$$

Initial values of $s(t)$ are decided to minimize $\|\sum_\tau a_\tau s(t-\tau+1) - x(t)\|$ with the initial values of a_τ. We fixed $\tau_{\max} = 5$, thus $s(t)$ is defined for $t = 1, 2, \cdots, 163$. We fixed $\lambda = 0.002 \times 163$ and $\beta = 25$.

First, we show examples of the original fMRI data after we apply a high-pass filter to them and normalize their means and variances. A signal from an activated region is shown in Figure 1 (a). It reflects a period of 5 scans of the experimental design, that is, a presentation of an object for 3 scans and a break for 2 scans. A signal from a deactivated voxel is shown in Figure 1 (b). It does not have this property.

We show values of the cost function L_s and L_n after the convergence of a_τ and $s(t)$. In Figure 2 (a) , a value of the cost function L_s for time series from each voxel is shown by a circle and their mean value is shown by a cross. The left side of Figure 2 (a) shows values for the data from an activated region and the right side shows those for the data from a deactivated region. We found a clear difference between these two groups. Values of the cost function for data from activated regions are mostly lower than those for data from a deactivated region. Since these values can be calculated without a true time series of stimuli, the cost function could be an index of the existence of some neural activity. Values of the first term of L, $L_1 = \|\sum_\tau a_\tau s(t-\tau+1) - s(t)\|^2$ and

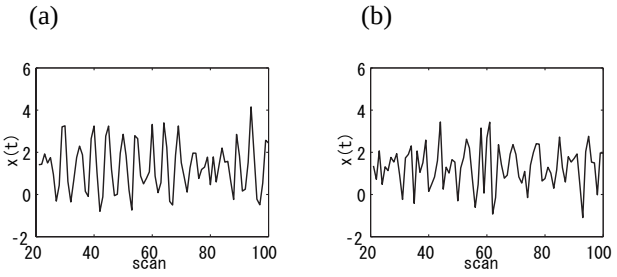

Figure 1: Original fMRI data after preprocessing. (a) A signal from a region activated by presentations of objects. (b) A signal from a deactivated region

the second term of L, $L_2 = \lambda_s \sum_t \log(1 + \beta s(t)^2)$ are shown in Figures 2 (b) and (c) in a similar way. In both indices, the difference between activated regions and deactivated regions is not as clear as in L_s. Note that we normalized the means and the variances of $x(t)$ and the norm of a_τ so that we could compare the values of the cost function between different data.

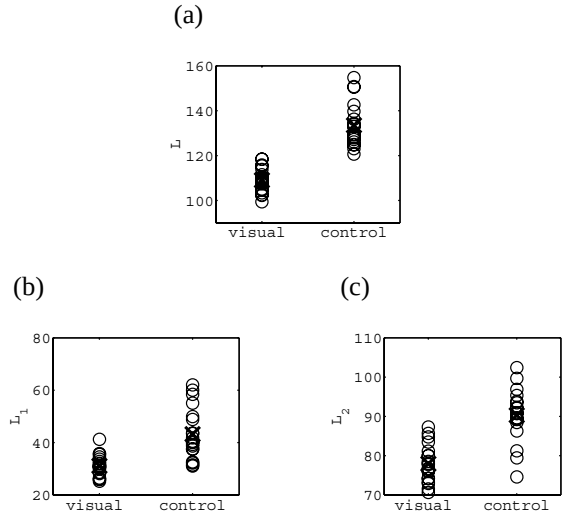

Figure 2: (a) Values of the cost function L_s after convergence. A value for time series from each voxel is represented by a circle and their mean value is represented by a cross. visual: values by data from an activated region. control: values by data from a deactivated region. (b) Values of the first term of cost function, L_1. Other conventions are the same as in (a). (c) Values of the second term of cost function, L_2. Other conventions are the same as in (a).

In Figure 3 (a) , a value of the cost function L_n for time series from each voxel is shown by a circle and their mean value is shown by a cross. The left side of Figure 3 (a) shows values for data from an activated region and the right side shows those for data from a deactivated region. Values of L_n also showed difference between activated voxels and deactivated voxels. However, the difference is less clear than that of L_s. Values of the first term of L_n, $L_1 = \|\sum_\tau a_\tau s(t-\tau+1) - s(t)\|^2$ and the second term of L, $L_2 = \lambda_n \sum_t s(t)^2$ are shown in Figures 3 (b) and (c) in a similar

way. For L_n, the values of the first term tend to be lower for data from activated regions. The values of the second term tend to be lower for data from deactivated regions. In both cases, the differences are not as clear as L_s.

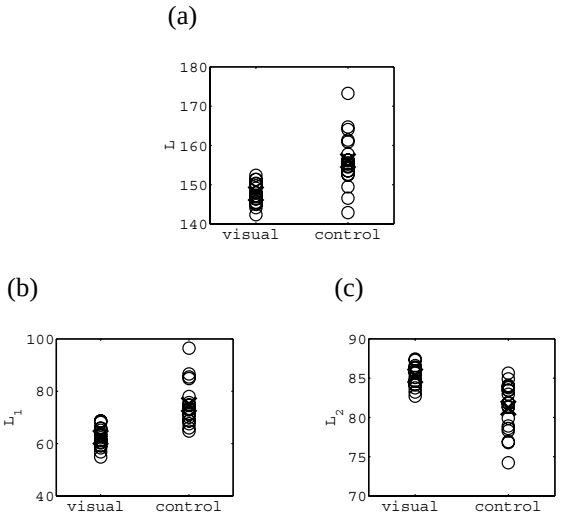

Figure 3: (a) Values of the cost function L_n after convergence. A value for time series from each voxel is represented by a circle and their mean value is represented by a cross. visual: values by data from an activated region. control: values by data from a deactivated region. (b) Values of the first term of cost function, L_1. Other conventions are the same as in (a). (c) Values of the second term of cost function, L_2. Other conventions are the same as in (a).

We examine how $x(t)$ is deconvolved into $s(t)$ and a_τ in detail especially for the sparse prior case. Examples of estimated $s(t)$ by minimizing L_s are shown in Figure 4. The time series shown in Figures 4 (a) and (b) are estimated from the fMRI data shown in Figures 1 (a) and (b), respectively. That is, Figure 4 (a) shows $s(t)$ from an activated voxel and Figure 4 (b) shows $s(t)$ from a deactivated voxel. In $s(t)$ estimated from the data from an activated voxel, there are large peaks mostly every 5 scans and the time series is similar to a true time series of stimuli. The time series $s(t)$ estimated from the data from a deactivated voxel is also peaky, however, we can not find a periodic structure.

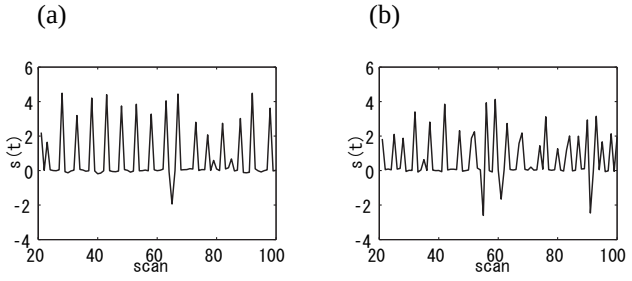

Figure 4: Estimated $s(t)$ by the proposed sparse blind deconvolution. (a) $s(t)$ estimated from $x(t)$ shown in Figure 1 (a). (b) $s(t)$ estimated from $x(t)$ shown in Figure 1 (b)

We can use the obtained time series $s(t)$ to estimate a time series of presentations of stimuli. However, in this data, the accuracy of estimation of stimuli time series from $s(t)$ was not very different from estimation results of $x(t)$.

We examined how much negative values exist in $s(t)$. After normalizing the variance of $s(t)$ to one, we counted times when $s(t) < -0.1$, that is, when it takes a significant negative value. In a sparse prior case, 1.7 % of $s(t)$ from activated voxels and 3.6 % of $s(t)$ from deactivated voxels take values less than -0.1. In a Gaussian prior case, 13 % of $s(t)$ from activated voxels and 11 % of $s(t)$ from deactivated voxels take values less than -0.1. In any case, the rate is small and we anticipate that a nonnegative constraint to $s(t)$ would not affect estimation results strongly.

The estimated HRFs by minimizing L_s are shown in Figures 5 (a) and (b). They are estimated from $x(t)$ shown in Figures 1 (a) and (b) respectively. The HRFs estimated from $x(t)$ obtained from activated regions tend to have wider peak with non-zero a_0 or a_5. Some HRFs estimated from $x(t)$ obtained from deactivated regions showed large decrease after taking the maximum value at $\tau = 3$. However, for both sets of data, the obtained HRFs have certain variety. Detailed analysis is left for a future study. We also learned a_τ with random initial values and confirmed that with data from activated voxels, in many cases, they converged to unimodal shapes similar to Figure 5 (a).

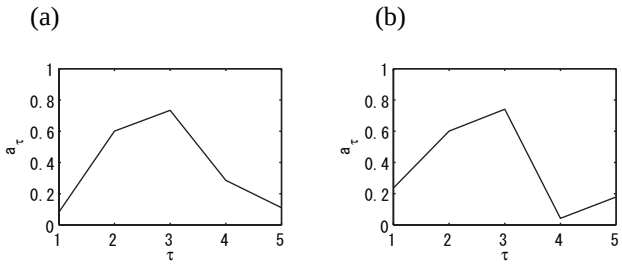

Figure 5: HRFs estimated from fMRI data by using the proposed sparse blind devonvolution. (a) a_τ estimated from data shown in Figure 1 (a). (b) a_τ estimated from data in Figure 1 (b).

4. DISCUSSION

We assumed a generative model of fMRI data which is a convolution of an HRF and a time series of stimuli. In addition, we assume that the values of time series of stimuli $s(t)$ distribute sparsely or normally. We constitute cost functions to estimate a_τ and $s(t)$ from fMRI data based on the model and the assumptions. Values of the cost functions are different between data from activated regions and those from deactivated regions. The cost function L_s with a sparse prior of $s(t)$ showed clearer difference. This means that data from activated regions are more compatible to the generative model defined by eq. (1) and the sparsity of $s(t)$. The value of the cost function can be calculated without a true time series of stimuli. Therefore, it could be used to detect activated brain regions when we could not obtain an exact time series of stimuli, for example, with natural stimuli or complex stimuli.

By using the proposed algorithm, we could estimate a time series of stimuli from results of deconvolution $s(t)$.

However, for these data, the advantage over estimating them from $x(t)$ was not very large. We also estimated voxel-wise HRFs. For sets of data from activated regions, we could obtain unimodal HRF shapes consistent with the conventional HRF shape.

Parameters $\tau_{\max}$, λ_s, λ_n and β were determined in heuristic ways. We shifted the used values to some extent and confirmed that the obtained results basically hold around used values of these parameters. For example, with a little longer a_τ such as $\tau_{\max} = 8$, the qualitatively similar results are obtained.

Some extensions of the proposed algorithm is possible. First, we could impose a nonnegative constraint to $s(t)$ and/or a_τ and see how different the results from the current study. Second, we could construct an algorithm to estimate a time series of stimuli using data obtained from multiple voxels together. This could give a better estimation of the time series.

Acknowledgments

We thank Uri Hasson from Weizmann Institute of Science, his colleagues and the fMRI data center for making the fMRI data available. We thank Karl Friston and his colleagues from the Wellcome Trust Centre for Neuroimaging for the SPM software.

REFERENCES

[1] U. Hasson, Y. Nir, I. Levy, G. Fuhrmann, and R. Malach, "Intersubject synchronization of cortical activity during natural vision," *Science*, vol. 303, pp. 1634–1640, 2004.

[2] M. J. McKeown, S. Makeig, G. G. Brown, Jung T.-P., S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, "Analysis of fmri data by blind separation into independent spatial components," *Human Brain Mapping*, vol. 6, pp. 160–188, 1998.

[3] C. Goutte, P. Toft, E. Rostrup, F. Å. Nielsen, and L. K. Hansen, "On clustering fmri time series," *NeuroImage*, vol. 9, pp. 298–310, 1999.

[4] P. P. Mitra and B. Pesaran, "Analysis of dynamic brain imaging data," *Biophysical Journal*, vol. 76, pp. 691–708, 1999.

[5] J.-R. Duann, T.-P. Jung, W.-J. Kuo, T.-C. Yeh, S. Makeig, J.-C. Hsieh, and T. J. Sejnowski, "Single-trial variability in event-related bold signals," *NeuroImage*, vol. 15, pp. 823–835, 2002.

[6] L. H. Hansen, "Ica of fmri based on a convolutive mixture model," *NeuroImage*, vol. 19, no. 2, suppl. 1, pp. e1637–e1638, 2003.

[7] R. A. Hutchinson, R. S. Niculescu, T. A. Keller, I. Rustandi, and T. M. Mitchell, "Modelling fmri data generated by overlapping cognitive processes with unknown onsets using hidden process models," *NeuroImage*, vol. 46, pp. 87–104, 2009.

[8] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.

[9] U. Hasson, T. Hendler, D. B. Bashat, and R. Malach, "Vace or face? a neural correlate of shape-selective grouping processes in the human brain," *Journal of Cognitive Neuroscienc*, vol. 13, pp. 744–753, 2001.

[10] M. S. Cohen, "Parametric analysis of fmri data using linear systems methods," *NeuroImage*, vol. 6, pp. 93–103, 1997.