Activation Detection and Characterisation in Brain fMRI Sequences. Application to the study of monkey vision.

Bertrand Thirion, Olivier Faugeras

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Activation detection and characterisation in brain fMRI sequences. Application to the study of monkey vision.

Bertrand Thirion — Olivier Faugeras

N° 4213

Juin 2001

THÈME 3

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4213--FR+ENG

Activation detection and characterisation in brain fMRI sequences. Application to the study of monkey vision.

Bertrand Thirion , Olivier Faugeras

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Robotvis

Rapport de recherche n° 4213 — Juin 2001 — 56 pages

Abstract: In this report, we propose a number of new ways of detecting activations in fMRI sequences that require a minimum of hypotheses and avoid any pre-modelling of the expected signal. In particular, we try to avoid as much as possible linear models. The sensitivity of the methods with respect to signal autocorrelation is investigated, in order to reduce or control it. Considering a experimental block design, a key point is the ability of taking into account transitions between different signal levels, but still without the use of predefined impulse response. The methods that we propose are based on well-known Anova and information theoretical models. The problem of statistical test validation is also studied and partly solved. The power of these methods seems high enough to avoid any smoothing, spatial or temporal, of the data.

Once an activation map is obtained, we attempt to characterise activations of the block experiment by studying the pre- and post- activation transitions. This more descriptive part of our work can be continued by searching for brain areas with homogenous characteristics, for example similar impulse responses. Quite naturally, this problem can be formulated as a clustering problem, which we solve through the use of a *fuzzy C-means algorithm*. This part of the analysis is performed without spatial or anatomical constraints, in order to allow for the observation of unexpected phenomena.

A first application is presented on a sequence of visual tasks obtained at Leuven University in order to characterise monkey motion perception. We propose activation maps, and, as a first step towards a spatio-temporal model of the brain, a map of impulse response patterns.

Key-words: fMRI, Anova models, Mutual Information, Markov Chain, Statistical thresholds, fuzzy C-means, robust filtering, activation maps.

Work partially supported by European Grant Mapawamo, No QLRT-1999-30161

Détection et caractérisation d'activations sur des séquences d'IRM fonctionnelles. Application à l'étude de la vision chez les singes.

Résumé : Dans ce rapport, nous proposons un certain nombre de nouvelles manières de détecter des activations dans des séquences d'IRMf, qui nécessitent un minimum d'hypothèses et évitent de pré-modéliser les signaux attendus. En particulier, on évite autant que possible les modèles linéaires. La sensibilité des méthodes par rapport à l'autocorrélation du signal est étudiée et contrôlée, voire réduite. Dans le cas d'un dessin expérimental en bloc, une question très importante est la faculté de prendre en compte les transitions entre les différents niveaux du signal, mais sans introduire de réponse hémodynamique prédéfinie. Les méthodes que nous proposons reposent sur des modèles d'analyse de variance et de théorie de l'information bien connus. Le problème de la validation statistique des tests est également posé et partiellement résolu. Ces méthodes sont assez puissantes pour permettre d'éviter tout lissage, aussi bien spatial que temporel, des images.

Une fois que la carte d'activation est obtenue, nous essayons de caractériser les activations observées en étudiant les transitions entre les blocs du paradigme expérimental. Cette partie plus descriptive de notre travail peut être poursuivie par une recherche des zones du cortex ayant des caractéristiques homogènes, par exemple des réponses impulsionnelles similaires. Assez naturellement, cette question peut être formulée comme un problème de coalescence, que nous résolvons à l'aide d'un algorithme de *fuzzy C-means*. Cette partie de l'analyse se fait en l'absence de contraintes spatiales ou anatomiques, afin de permettre l'observation de phénomènes inattendus.

Une première application est présentée à partir d'une séquence de tâche visuelles obtenues à l'université de Louvain pour caractériser la perception du mouvement chez les singe. Nous montrons des cartes d'activations, et, comme premier pas vers un modèle spatio-temporel de l'activité corticale, une carte de modèles de réponses impulsionnelles.

Mots-clés : IRMf, modèles Anova, Information mutuelle, entropie, chaine de Markov, méthode statistique, fuzzy C-means, filtrage robuste, carte d'activation.

Ce travail est financé en partie par le Fonds Européen Mapawamo, No QLRT-1999-30161

Contents

1 Introduction

Functional Magnetic Resonance Imaging (fMRI) is a powerful tool for investigating brain activity. It relies on the measurements of derived effects of brain activity (BOLD effect or blood volume for example); the physiological relationship between neural events and observed changes in blood volume are still investigated, but many experiments confirm the quantitative dependence of blood volume and blood flow, as well as BOLD signal, on activity (see [GGWF01], [MLL$^+$01], [VN98], [KvC99]).

Information is available as 3-D functional MR images sequences measured with a specific paradigm (sequence of cognitive tasks). Analysing such sequences involves in our view two tasks :
- Detecting activations in order to detect spatial loci of activity.
- Characterising quantitatively this activity in order to rely it with neuro-physiological models.

More specifically, detecting activations means associating cognitive tasks with processing areas in the cortex. The result is called an *activation map.* Activation maps can :

- Take into account the whole experiment, to define areas that were mostly involved by the paradigm tasks altogether (study of an *overall effect*).
- Take into account each condition separately in order to study differences between the neuro-physiological events induced by different tasks; this leads to cognitive subtractions/conjunction studies.

Characterising quantitatively brain activity means computing local parameters that describe more precisely the activations : e.g. impulse response ([KvCD99]), noise level, or delay in the activation peak ([GHLR99]).

The results of activation detection and characterisation should then be summed up in a synthetic representation of the brain as evidenced through the experiment, and mapped onto anatomical data.

fMRI data processing is not a new topic: A great deal of research has been performed in the past ten years to detect activations reliably. A very rich overview is available in [PNPH99a], [PNPH99b]. Let us give a quick account of recent developments.

- The most commonly used methods are based on **linear models** of the data. This is of course the case of statistic parametrical mapping (SPM) (see [FA$^+$97]). Based on a simple linear equation of the signal, this method fits the signal to known models, and gives a score of fit. Further inferences are based on the theory of gaussian fields. This now wide-spread and well-interfaced method is efficient , since it relies on well-known statistical tests, but it also raises some questions:

  - The hypothesis of the linearity of BOLD signal with respect to neuronal activity is questionable (see [MLL+01], for example).
  - The hypothesis of gaussianity of the noise involves the use of spatial and/or temporal smoothing, which biases the data.
  - The use of pre-defined impulse responses is practical, but may also bias the statistical inference, and especially hide transient task-related activations.
  - The whiteness of the noise is an issue with these models ([ZAD97]).
  - Many technical difficulties arise with linear models, due to the possible non-orthogonality of the columns of the design matrix (see [AKKK99]).

  Some solutions have been proposed in a linear framework that are robust to noise autocorrelation ([BD00]), and avoid the use of design matrices and predefined impulse response. This is certainly a good way of allowing for more flexibility in fMRI analysis.
  We also present ANOVA models in section 3.1.1, which are usually thought of as linear models, but that enable nonlinear developments. A comparison and cross-validation with linear models still remains to be done.

- Other authors have embedded the usual linear model in the **maximum likelihood** framework. The methods consist usually in estimating *i)* the response signal of interest and *ii)* the nuisance signal present in the data. Different hypotheses have to be made in this modelling ; the final decision is expressed as a maximum likelihood problem, even if the linear modelling of the data in fact yields classical models. However, the fundamental problem of noise correlation is not completely treated ([AKKK99], [NN99]). Some approaches model a priori activation patterns, [EB99] [HJ00], which is questionable. Some models are more complete, but at the expense of the clarity of the results [KAK99]. Moreover, reliable statistics require Monte-Carlo simulations of the data.

- **Time-frequency methods** have also been adapted to the problem: the temporal fMRI signal can be studied in the Fourier space [LZ97]; the main advantage is that the correlation present in the noise does not prevent frequency components from being independent, which allows the use of statistics that require independence of the random variables [MR00]. New versions of this method use wavelets [bL+01].
  On the other hand, this kind of method is not necessarily well-suited for non-periodic, and non-binary experimental paradigms. The generalisation to event-related experiments remains also problematic [MR00].

- **Mutual information** has also been proposed as a tool for activation detection in temporal signals ([TFW+99] and [KFT+00]). This idea is developed and analysed further (section 3.1.1).

- A method that requires minimal hypotheses has recently been proposed [LU01] : the experiment being repeated, a correlation coefficient between the two time series is computed. A particular voxel is activated if the pattern is similar in the two sequences. Using only the **repeatability** of the experiment output, this algorithm needs no prior hypothesis on the signal to work. But the significance of the scores obtained is questionable, and the decision may be arbitrary.

An important issue is whether to take into account spatial constraints in fMRI. Three solutions are commonly used :

- Spatial smoothing of the data, which is now a standard, and is useful to ensure the gaussianity of the spatial fields used in some methods. An evident drawback is the blurring introduced in the data, particularly when information from different physiological areas is mixed. Some newer approach perform anisotropic smoothing under anatomical constraints ([AKM$^+$01]).
- Treating spatial constraint as information in the framework of decision theory (Maximum a posteriori) : see [HJ00] [KFT$^+$00] [DKvC98] .
- Clustering of the data, and assessment of the clusters [BWM98], [GHLR99].

The difficulty of the problem resides in the fact that it is general difficult to assess the significance of the activation in a signal both in the temporal and spatial domains. This problem has been partly solved with the introduction of multi-variate techniques (PCA, ICA [MJ$^+$98] [MM$^+$98]). We think that these promising methods still have to be compared with univariate methods in order to evaluate their optimality in practice.

In particular, we raise the following two points :

- Activation tests need to be associated with statistical assessment of the results by fitting them to realistic probability distributions (concept of P value).
- The characteristic dimensions of the problem (3×3×3 mm) plead against any form of spatial smoothing of the data, since the structures of interest are likely to be sub-voxelic, and the overall size of the brain is small -in our case, the subject is a 3 years old rhesus monkey. The mapping between anatomical and functional images being not completely satisfactory, we prefer to separate the temporal and spatial dimensions of the problem.

These considerations persuaded us to treat and assess activations on a voxel basis -after coregistration of the images- and to perform clustering afterwards on local and meaningful features (impulse response) in order to obtain a more precise spatio-temporal description of the data.

The next sections are organised as follows : in section 2 we describe the main steps that we propose for fMRI analysis; in section 3 we develop the different aspects, in particular activation detection, for which we propose 6 solutions. A comparison of these 6 methods is proposed in section 4 on both simulated and real data. Section 5 is devoted to an experiment on real data acquired for the study of monkey vision.

The appendix develops technical points : Appendix A presents the derivation of a statistical distribution for Mutual information under the null hypothesis ; so does appendix B for ANOVA models, with the more specific question of noise autocorrelation. The effect of the preprocessing methods presented in section 3 is described in section C. Finally, a quick recall on ARMA regression models is made in section D, with a specific focus on AR(1)MA(1) models.

2 Description of our system for analysing fMRI data

Our system, shown in figure 1 consists of five different modules.

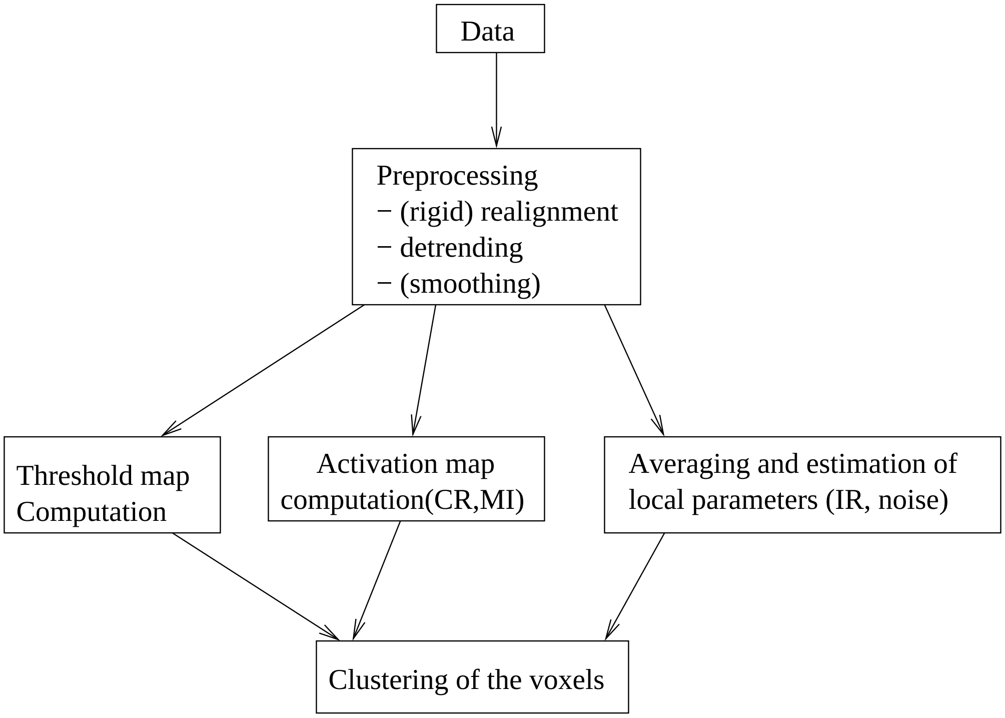

Figure 1: Processing of fMRI data. Note that the computation of activation map and the estimation of local parameters are performed independently, since these steps rely on different hypotheses.

Preprocessing of the data includes first a rigid spatial realignment and interpolation of the data; this is sufficient since the main effect in our data is a translation of up to 6 mm along one of the coordinate axes. Note that realignment reduces in some way fMRI analysis problem from a 4-D to a 1-D problem. The data is detrended : for each voxel separately, constant and linear components in the temporal domain are subtracted (more correction is not necessary) so that grey level distribution is broadly stationary over time. Smoothing should be minimal -if not avoided- in the temporal and spatial dimensions, as soon as SNR is high enough. Henceforth, the data will be considered voxel-wise.

The activation map is produced by computing a statistical test of the data with respect to the paradigm. We call our tests *Correlation Ratio (CR)* or *Mutual Information (MI)*. See section 3.1 for details. At this step, the a priori hypotheses have to be minimised : neither a linear model nor a specific response is expected from the data. The experiment can be treated as a whole, or different conditions of interest can be distinguished and compared.

A threshold map adapted to the activation map has also to be derived in order to assess the results of the activation map. This has sometimes been performed through Monte-Carlo methods [NN99] ; although efficient and reliable, this does not facilitate the interpretation - the results having only a relative value. We rather propose to compute thresholds from modelling considerations. According to the properties involved in the activation test, the threshold can be either uniform or not in each area of interest. A realistic approach is often to view the properties of the brain as spatially non-stationary, thus adapting the thresholds if necessary.

Before estimating **local parameters**, an averaging of the data is performed since the paradigm is periodic. The purpose is to get more precise information from the data, like an estimate of the impulse response (IR) - called hemodynamic response (HR) in the case of BOLD signal. The model is here more constrained, i.e. we use a linear model of the temporal response at a voxel. But the IR is estimated for each voxel independently, and no a priori shape is used. The optimal linear estimate has to be smoothed without introducing a bias (see more details in 3.3).

Clustering the voxels means gathering the voxels with similar characteristics into regions, not necessarily connected, (this is the "qualitative" part of their behaviour), taking into account their score in the activation map (which sounds more like a "quantitative" part). See section 3.4 for more details.

3 The different parts of the system

We now describe four of the modules of our system.

3.1 Activation map

Let us see how to compute an activation map: given the realigned and normalised time sequence of each voxel, and the experimental paradigm, this means evaluating how well the voxels time-series "correspond" to the paradigm -what we mean by "correspond" is made more precise in the sequel. Here we consider a block paradigm; there is one baseline condition and different experimental conditions. Formally, the paradigm will be treated as a discrete variable taking its values in $\mathbb{N}$(the set of positive integers). All the scans under different experimental conditions of interest are separated by baseline conditions, which enables us to consider them separately. But we do not assume any kind of functional model between the paradigm and the voxel time-series.

This section is organised as follows : First, the well-known Analysis of variance (ANOVA) and Mutual Information (MI) activation detectors are recalled. Their main shortcoming is that they model the data in terms of stationary distributions, which is questionable when studying the brain. We therefore present some solutions to improve these basic modelling techniques. This is achieved by introducing the time dimension into the models: The ANOVA method is thus enhanced with a concept of *memory*, while a temporal layer is introduced in the MI method. We furthermore consider temporal time-courses as dynamical systems, and conjecture that most of the information is contained in the asymptotic distribution of the time values: thus we propose a method to compute this distribution before running the ANOVA and MI activation detectors.

3.1.1 Two classical methods : ANOVA and Mutual Information

The ANOVA Model.
Analysis of variance (ANOVA) is a well-known tool of linear statistics. See [Seb77] for a theoretical overview. It is also popular in fMRI analysis ; among others it has an important role in SPM inference -see for example [FA+97] : according to the *extra sum of squares principle*, which states that ANOVA statistical tests are assessed through F-functions, these models are known as F tests.

In a different framework, namely coregistration between medical images (see [RMA99] [RMPA98]), a similar model has been used to reveal the functional dependence between random variables. This technique has been called *Correlation Ratio*, and is explicitely intended to work under nonlinear hypotheses.

Our aim here is to revisit the ANOVA model with similar preoccupations as *Roche et al.* We will also compare it to information theoretic models. First, we briefly recall the model, in our activation detection framework.

We study a discretised temporal signal $y(t)$, and try to check whether it is functionally dependant on an explicative variable, namely the paradigm $p(t)$. Writing

$$y(t) = f(p(t)) + n(t), \tag{1}$$

where $n(t)$ is some noise, we can check the dependence by the Correlation Ratio test :

$$CR(y|p) = \frac{var[E_p(y|p)]}{var(y)} = 1 - \frac{E_p[var(y|p)]}{var(y)}$$

This number varies between 0 and 1 and it can be shown that

$$CR(y|p) = 1 \leftrightarrow \exists \Phi : y = \Phi(p)$$

This test is optimal in the framework of the L^2 space of random variables. Choosing this space amounts to the hypothesis that $n(t)$ is an independently identically distributed (i.i.d.) variable with gaussian law, which we denote $n(t)\, i.i.d. \rightsquigarrow \mathcal{N}(0, \sigma^2)$, where σ is the noise level. The model can be viewed as *linear*, in the sense that the noise is additive, but also as *non-linear*, since f can be any function. This last distinction becomes meaningless when $p(t)$ can take only two values (0/1) -that usually stand for baseline/activation. This case will be referred to as the *binary paradigm*.

In the ANOVA framework, the quantity of interest is in fact

$$\eta(y|p) = \frac{var[E_p(y|p)]}{E_p[var(y|p)]} = \frac{CR(y|p)}{CR(y|p) - 1} \ , \tag{2}$$

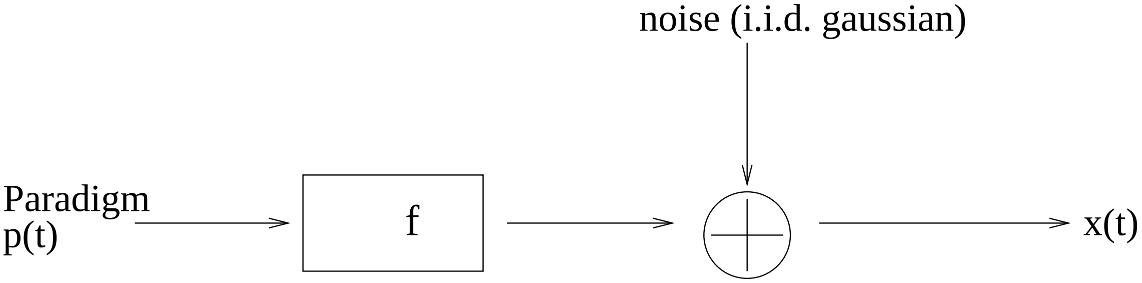

Figure 2: Graphical interpretation of the model (1)

and can be thought of as the quotient of the deterministic part of the signal (given $p(t)$) with its random part ; the numerator and the denominator are independent whenever the noise is not temporally correlated. The hypothesis $f \equiv c$ (no functional dependence) is invalidated with probability P thanks to the test $\int_0^\eta F_{n_1,n_2}(x)dx < 1 - P$, where (n_1, n_2) are proper degrees of freedom, e.g. n_1 is the number of scans minus n_2, which is the number of possible values for the explicative variable minus 1.

The model (1) is described graphically in figure 2.

Let us discuss its characteristics :

- The only hypothesis on the activation model is that the mean value of the signal is stationary conditionally to each value of the paradigm $p(t)$. For example, in the case of the binary paradigm, the activation signal level can be lower or greater than the baseline signal level. In some sense, **minimal a priori knowledge is required**.

- The gaussianity of the residual noise is an issue. We have checked it on our current data with a Kolmogorov-Smirnov test, and it appears that, when the noise level is significantly above 0, it can be considered as gaussian in 75% of the cases, with a confidence of 95 %. This questions the optimality of Anova method. From a different point of view, one can view the variance as a consistent way of measuring a noise level, even though it is non-gaussian. For this reason, we present further developments based on Anova principles. Finally, let us notice that the mutual information test presented beyond does not rely on a gaussian hypothesis.

- The hypothesis that the noise is temporally independent is highly questionable. Noise may be temporally correlated due to multiple experimental artefacts. In such a case, the statistical tests are invalid and usually lead to an over-segmentation of the activated areas.

- Another shortcoming of the model is that transitory aspects of the signal are ignored. This is more or less a problem depending on the temporal resolution of the signal, but it is generally considered that a transition lasts from 5 to 15 seconds, which is not negligible in most experiments.

Mutual Information.

Mutual Information has already been proposed as a way of detecting activations in fMRI data (see [TFW$^+$99] and [KFT$^+$00]). The principle is the following : the measured signal $y(t)$ can be described by a statistical distribution $D(h)$ (h will stand henceforth for the grey level). $p(t)(\in \{0,..,m\})$ being known, we can also derive the distribution $D_i(h), i \in \{0,..,m\}$ which are the distributions of $y(t)$ conditionally to the hypothesis $p(t) = i$. Let us denote $H, H_0,.., H_m$ the entropies of $D, D_0,.., D_m$. Then mutual information (MI) between p and y is defined as the sum of the entropies of the two marginal distributions of y and p minus the entropy of the two-dimensional distribution of (y,p) ; but, since $p(t)$ is deterministic and belongs to a finite set, the formula of MI reduces to :

$$MI(y|p) = H - \sum_{i=0}^{m} P(p=i).H_i \ , \tag{3}$$

where $P(p = i)$ stands for the probability of the event $p(t) = i$.

If there is no activation, then $D \simeq D_0 \simeq ... \simeq D_m$ and $MI(y|p) \simeq 0$. In any case, $MI(y|p) \geq 0$.

In practice, D_i are estimated through a Parzen window method, that involves a convolution with a kernel -usually gaussian, whose parameter β has to be properly tuned. An example corresponding to an activated signal is shown in figure 3.

Let us formulate some remarks on this method :

1. The model relies on very weak hypotheses, which makes it very general and usable in many cases.
2. There exists no straightforward way of thresholding to detect activations, unless using Monte-Carlo simulation.
3. As in the basic ANOVA model, transitions are not modelled, inducing a loss in detection power.
4. The MI model has little sensitivity to noise autocorrelation.
5. On the other hand, MI depends crucially on the number of samples used, the signal amplitude and the method of entropy computation.

We propose a solution to deal with problem 2 : given a distribution D, we show how to determine a threshold T so that the hypothesis $D_0 \simeq D_1 .. \simeq D_m (\simeq D)$ can be rejected with probability P : see section A. The derived formulas, given in the paragraph 3.2 are well confirmed by numerical simulations. Basically, MI behaves as a χ^2 distribution with appropriate parameters.

In the next two paragraphs we describe a number of ways to model transitions when MI is used as a statistical test.

3.1.2 Taking into account temporal aspects

ANOVA method: Adding memory

We propose here some solutions to two of the critics formulated above : the noise should be considered a priori as temporally correlated and the signal at time t_0 is a function of $p(t)$ for $t \leq t_0$.

The second point can be solved by generalising the notion of *paradigm* to a notion of *state* of the system : a state $s(t)$ is defined as the paradigm value at time t, $p(t)$, but also at previous times :

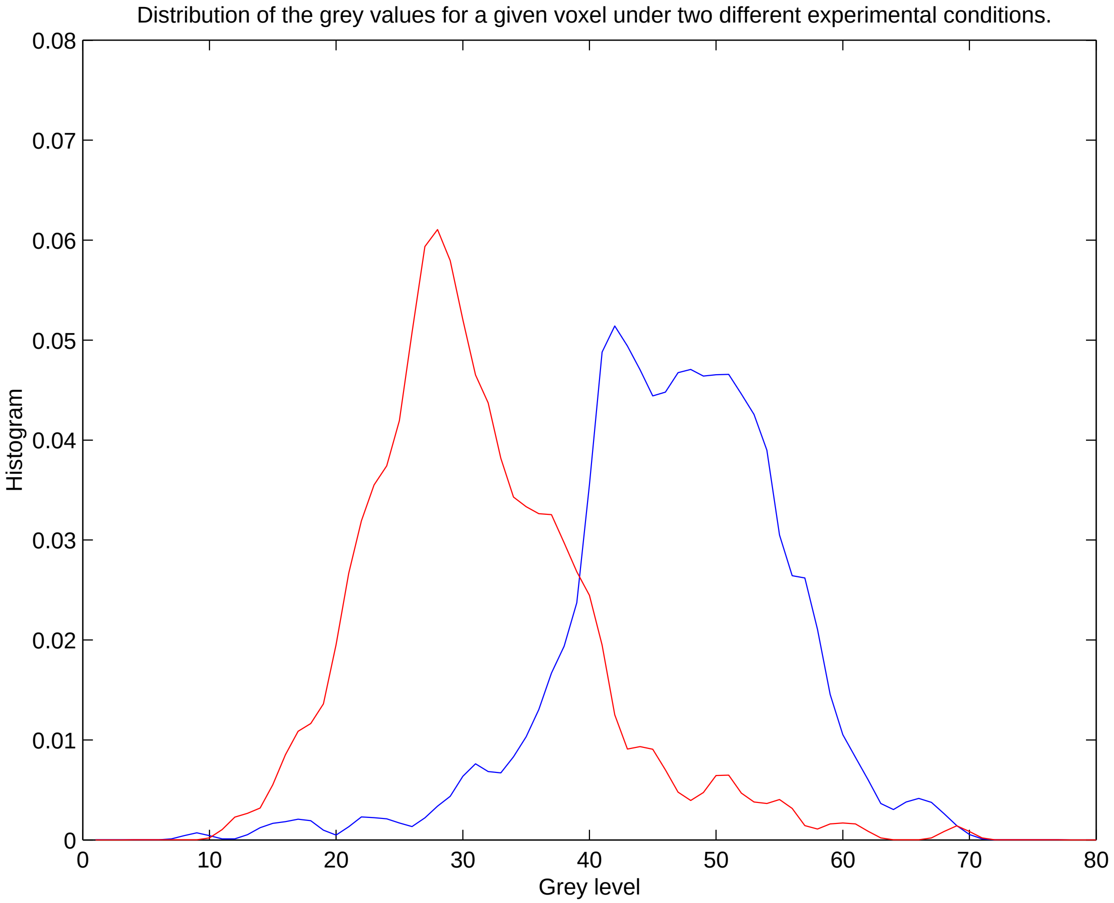

Figure 3: Observation of an activated voxel signal under a binary paradigm. The blue values stands for the activated state, the red value for the baseline state. The two distributions are clearly distinct ; the MI score is 0.3148 -the maximal reachable value being here $\log(2) = 0.69$.. The overlap between the activation and baseline distributions is mainly due to measurements corresponding to transitions.

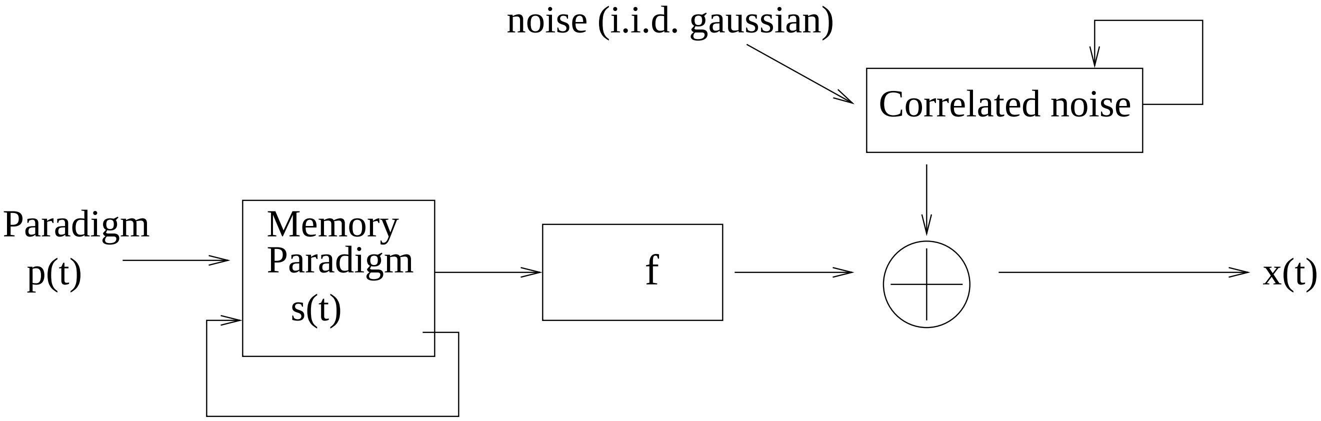

Figure 4: Graphical interpretation of model (4).

$s(t) = (p(t), p(t-1), .., p(t-k))$. Let $S = \{s(t)\}$ be the set of possible values for s. Keeping an additive noise model yields :

$$y(t) = f(s(t)) + n(t) \tag{4}$$

A graphical interpretation of this model (4) is displayed in figure 4.

We have introduced the parameter k a priori; It should be tuned to be greater than 10 seconds in order to model effective transitions. On the other hand, the statistical power of the tests decreases with k ; thus, it should not be too large. Taking 20 seconds is satisfactory and, in practice, the exact value of the parameter has little influence on the results.

The corresponding Anova test for functional dependence becomes :

$$\eta(y|s) = \frac{var[E(y|s)]}{E_s[var(y|s)]} \tag{5}$$

Assuming a temporally uncorrelated noise yields a similar statistical test as previously for η, the only difference being that the number of degrees of freedom has to be modified. Let us study the following example :

The paradigm is a sequence of epochs defined in figure 10. The length of each epoch is 20 scans (as an illustration, we have TR = 3.321s thus we choose k = 7). The experiment is repeated 15 times and $p(t) \in \{0,1,2\}$; then using the notation
$1^l 0^n = \underbrace{1..1}_{l}\underbrace{0..0}_{n}$ we have

$$s(t) = (10^{k-1})(1^2 0^{k-2})..(1^{k-1}0)\underbrace{1^k..1^k}_{20-k}(01^{k-1})..(0^{k-1}1)\underbrace{0^k..0^k}_{20-k}(20^{k-1})..(2^{k-1}0)\underbrace{2^k..2^k}_{20-k}(02^{k-1})..(0^{k-1}2)\underbrace{0^k..0^k}_{20-k}...$$

In that case, $cardinal(S) = 3 + 4(k-1) = 4k - 1$. The F-test is valid, letting $n = cardinal(S)$ and $p = N - cardinal(S) - 1$, where N is the number of scans in the experiment.

The remaining issue is noise temporal correlation. Noise correlation can be described by autoregressive (AR) or moving average (MA) model. We found that an AR(1)MA(1) model, described in section D fits the data correctly. It is defined as :

$$n(t) - \rho.n(t-1) = m(t) + \lambda.m(t-1) \; with \; m(t) \leadsto \mathcal{N}(0,\sigma^2), \; 0 < \rho < 1, \; -1 < \lambda < 1.$$

Unfortunately, the F-test is no-longer valid. However, some approximations are plausible, using the concept of *effective degrees of freedom*. They are presented and discussed in section B.2. Qualitatively, the degrees of freedom are reduced when noise correlation increases.

We found that this model gave a better interpretation of our experimental results (see figure 21), but with some drawbacks:

- It would have been easier to pre-whiten the signal $y(t)$, which is possible since its correlation structure can be computed (see section D). But $s(t)$ is also highly autocorrelated, As a consequence, removing signal correlation involves removing activation $(f(s(t)))$ from the signal.
- The algorithm that we obtain is made of the following steps:
  - Computation of η
  - Estimation of f and of $n(t) = y(t) - f(p(t))$
  - Computation of AR(1) MA(1) coefficients
  - Computation of effective degrees of freedom.
  - F-Test on the η value.

  It relies on successive, more or less accurate, estimations. This raises the question of the final validity of the test.
- Another difficulty intrinsic to our method is the hypothesis that the asymptotical signal level is fixed according to the experimental condition. This may be questionable in practice due to uncorrected drifts in the data.

MI : introducing an implicit Transition model.
Usually, it can be assumed that temporal sequences of activated voxels tend to an asymptotical value after a transition (that lasts broadly between 5s and 15s). The conditional distributions can be well separated from one another when asymptotical values are reached, whereas transitions blur the distributions.

Transitions are usually monotonic in the sense that the grey level behaves monotonically across time, they thus can be distinguished on the $(y(t), y(t+1))$ graph, as shown by the following example in figure 5 where a real signal under a binary paradigm has been taken.

The distributions being well-separated (i.e. non-overlapping) in a 2-D diagram, their mutual information can be computed on the 2-D domain exactly in the same way as in the 1-D case ; given that mutual information

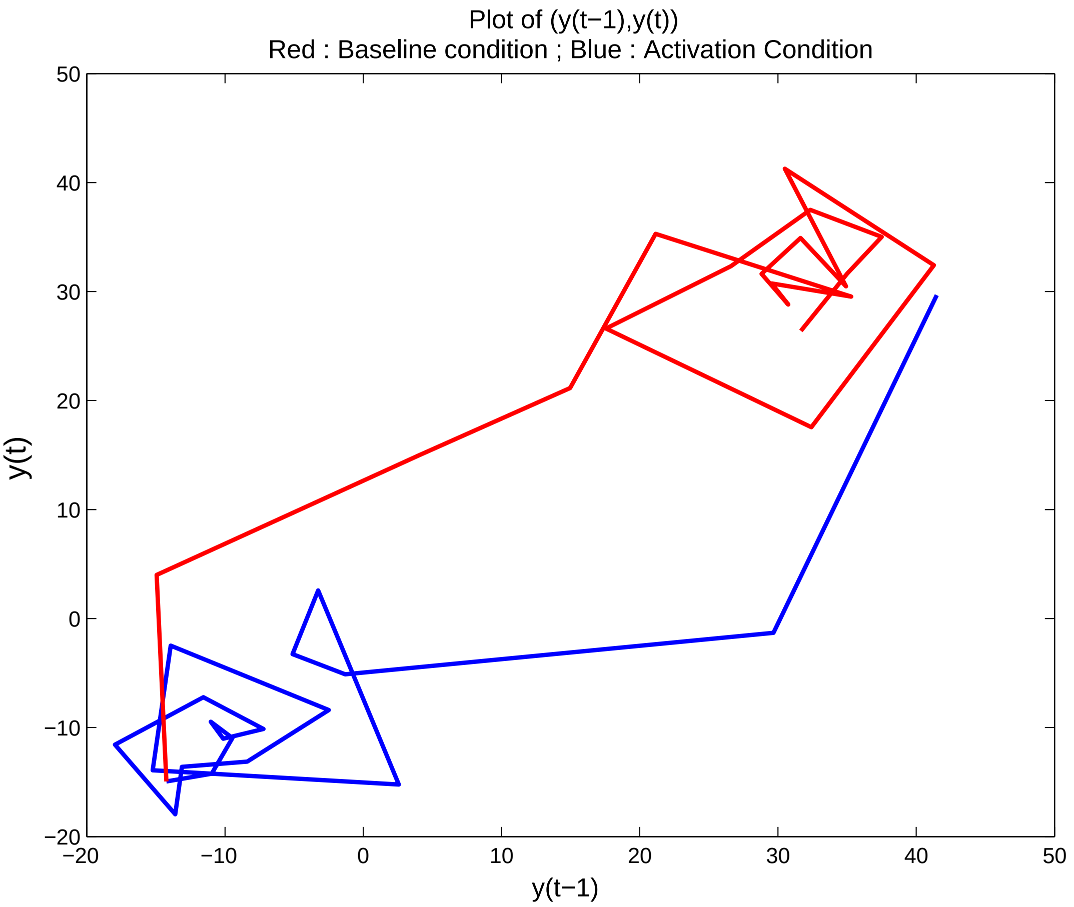

Figure 5: Observation of an activated voxel signal under a binary paradigm. The blue values stands for the activated state, the red values for the baseline state. The two transitions are clearly visible, but they do not induce an overlap in the joint distributions due to their monotonicity : $y(t+1) > y(t)$ under baseline condition, $y(t+1) < y(t)$ under activation condition. Notice that activation is negative in that case. The signal has been averaged over 15 trials.

between two random variables deceases with the overlap of the observed distributions, some signal is gained here to characterise activations. However, the computation of some significant threshold becomes somehow problematic. This model is referred to as *implicit* since the transitions are not defined as such (no temporal windowing has been performed). A nice feature of this method is that, once the paradigm $p(t)$ is known, it requires no additional hypothesis. A difficulty raised by the previous method is that it involves the computation of entropy with 2-D distributions, a process known to be more noise sensitive than in the case of 1-D distributions.

3.1.3 Alternative : Using preprocessing to simulate asymptotical signals

Explicit transition model.
Linear models model transitions through the use of predefined impulse responses. Their disadvantage is the bias introduced in the statistical inference. We propose to model explicitely the transitions while keeping a more descriptive model, so that the post-transitory signal distribution is computed without bias.

Let us define a *system* as the association of a voxel (v) and a given value (i) of the paradigm. The values (grey level) taken by (v,i) across time can be described as a succession of distributions $D(0),..,D(t),D(t+1)..$ where t runs along the successive instants of observation of (v,i) (it is assumed that the experiment is repeated several times ; if not, $D(t)$ reduces to one point). $D(t)$ can be viewed as the intrinsic noise distribution of the system (v,i) at time t.

We propose to model the dynamics of the system as a **Markov Chain**. In other words, we assume that $D(t)$ only depends on $D(t-1)$. Such a system is described by a transition matrix Q such that :

$$D(t) = Q.D(t-1)$$

Q -which depends only on (v,i)- contains all the information about the dynamics of the system. The element $Q_{a,b}$ of Q is $Q_{a,b} = P(x(t)=a|x(t-1)=b)$. This implies the properties : *i)* $Q_{a,b} \geq 0$ *and ii)* $\sum_a Q_{a,b} = 1$. Such a matrix is known to have one and only one eigenvector with eigenvalue 1. Let us denote it by D^∞. It can be viewed as the asymptotic noise distribution of the system ($lim_{t\to\infty} D(t)$); in fact, it is simply the distribution of the signal once the transition is achieved.

Algorithm for Markovian pre-processing and activation detection

- for all the values i of p
  - compute a global distribution D_i of the values taken by (v,i).
  - from all the pairs $(x(t),x(t+1))$, estimate the probabilities $Q_{a,b} = P(x(t)=a|x(t-1)=b)$.
  - let $Q_i = (Q_{a,b})$ be the empirical transition matrix.
  - approximate D_i^∞ by $Q_i^n.D_i$ for n big enough.
- compute the probability-weighted average $D^\infty = \sum_i P(i) D_i^\infty$.
- compute your favourite statistical test , e.g.:

$$MI(y|p) = H(D^\infty) - \sum_i P(i) H(D_i^\infty)$$

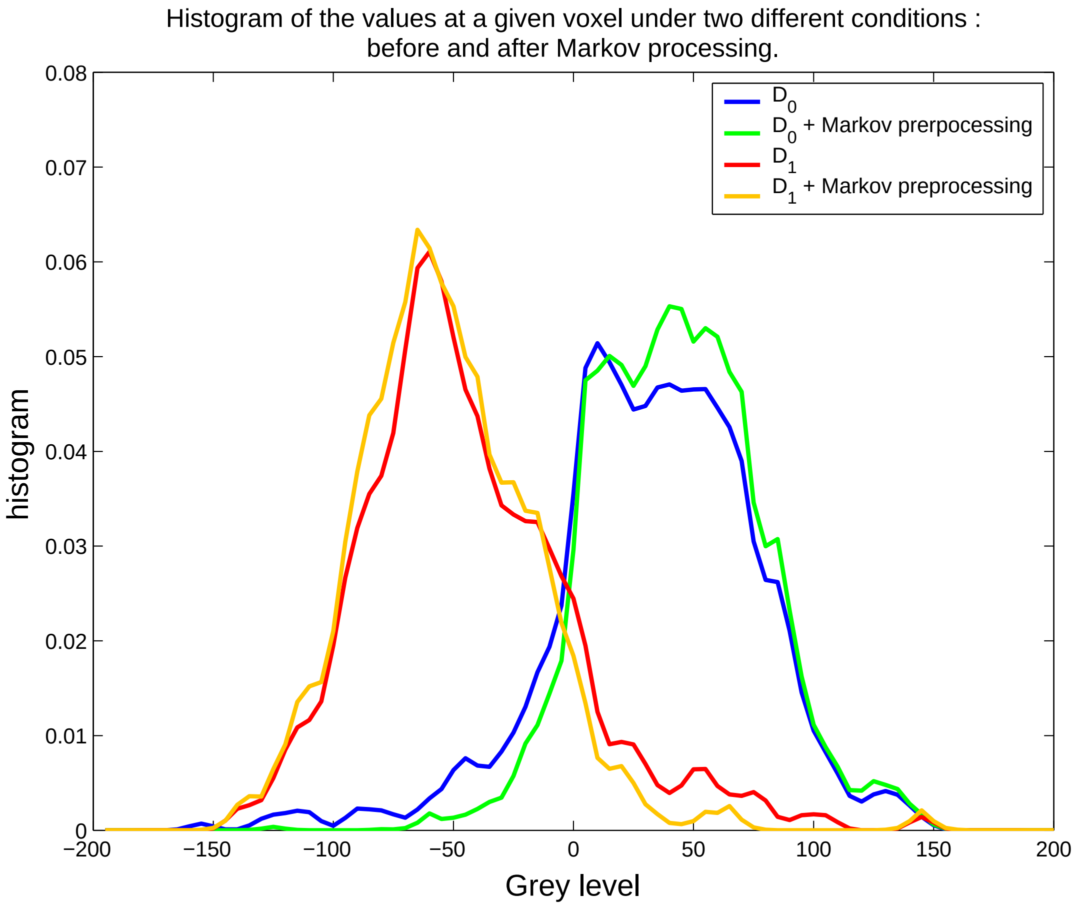

Figure 6: Filtering the grey value distribution to lessen the transitory parts of the signal. The red and blue curves are the global distributions under two different conditions in the same experiment, whereas the magenta and green curves are the Markov-filtered versions. The overlapping between the distributions is significantly reduced, and the MI score increases from 0.3148 to 0.49 (the maximal reachable value being $\log(2) = 0.69..$).

$$CR(y|p) = \frac{var_i(E(D_i^\infty))}{E_i(var(D_i^\infty))}$$

An example is displayed in figure 6 and shows the separating effect of our algorithm.

Convergence of the algorithm.
The transition matrix Q having been estimated, does it have an eigenvector D^∞ as stated in the previous

paragraph ? The following general result [Nev96, chap 2.5 and 2.6] provides existence conditions :

Let Q be a recurrent irreducible transition matrix on its state space E -E is simply the natural space of the observed data, here the grey level; then, given two events x and y different or not, the frequency of reaching the state y, given the initial state x and the transition matrix Q, is independent of x when time $t \to \infty$, that is

$$\frac{\sum_{s=1}^{t} 1_{x(s)=y}}{t} \to \mu(y) \ almost\ surely$$

The limits $\mu(y)$, $y \in E$ are the unique invariant probability densities on E invariant by Q, that is such that $\sum_y \mu(y)Q(y,z) = \mu(z)$, $z \in E$, if this probability density exists (if not, $\mu(y) = 0 \ \forall y \in E$). The Markov chain is said to be ergodic.

Does it apply in our case ? One has to check whether Q is recurrent and irreducible. E being finite, Q is recurrent as soon as it is irreducible. Irreducibility means ; $\forall (x,y) \in E \times E$, $\exists n / Q^n(x,y) > 0$. A slight gaussian smoothing on the coefficients of Q induces $Q(x,y) > 0 \ \forall (x,y) \in E \times E$, making Q irreducible, so that the theorem applies.

3.1.4 Summary on activation detection methods

In the previous sections, six activation detection methods have been proposed. ANOVA and MI methods are *classical*, whereas ANOVA+Memory, MI-2D, Markov+ANOVA and Markov+MI are new. We want to sum up the characteristics of this methods from the modelling viewpoint (a comparison of their performances is given later). Three features have been retained :

- Modelling precision : the ability of the method to analyse and discriminate between different behaviours of the time series.
- Sensitivity to noise correlation : robustness (assessed in terms of statistics) of the test with respect to the correlation of input noise under the null hypothesis. This is obtained through simple numerical simulation.
- Technical difficulty: It involves the complexity of the algorithm implementation, but also the need for tuning hidden parameters in order to obtain meaningful results.

methods	ANOVA	MI	ANOVA+Memory	MI-2D	Markov+ANOVA	Markov+MI
modelling precision	low	low	high	high	high	high
Sensitivity to noise correlation	high	null	high	null	high	low
Technical difficulty	none	medium	low	medium	high	high

3.2 Threshold Map

Our purpose is to derive significant thresholds for the statistical scores (CR,MI) in order to segment the activated parts of the brain. The idea is the following : consider the signal obtained at each voxel as a random variable. Making an activation test is equivalent to filtering the signal through some (highly non-linear) filter. Assessing the values of the test involves answering the following questions : In the case of the null hypothesis, the input signal is random, what is the distribution of the output signal ? Given a probability P_0 (say 10^{-4}) can we find a threshold t_0 so that $P(score > t_0) < P_0$?

It has become usual to perform such computations through Monte-Carlo simulations [NN99], but it is probably more interesting to infer thresholds from theoretical considerations. Such considerations and the computations they imply are presented in appendices A, B and C for both MI and CR. The formulas obtained involve parameters that characterise the time series, e.g. their standard deviation and temporal correlation. this means in practice that the thresholds will be non-uniform across the image. For example, the signals obtained on the skull have different characteristics than those obtained in the white and grey matter, necessitating different threshold values.

The general methodology is :

- Compute local noise parameters (intensity, autocorrelation).
- Infer a statistical distribution $\mathcal{D}$ that corresponds to the activation test under the null hypothesis ($\mathcal{D}$ may depend on the previously estimated parameters).
- Given P_0 compute t such that $\int_{-\infty}^{t} \mathcal{D} = 1 - P_0$
- Label the voxel as *activated with probability* P_0 whenever its test value is greater than t.

Model distributions ($\mathcal{D}$) have been determined for the different activation tests presented above.

The parameters used are the following :

- k : Number of studied experimental conditions .
- N : Number of scans.
- l : When a memory-method is employed, memory size.
- σ : Standard deviation of the signal.
- β : Smoothing parameter used to estimate the grey level distribution.
- R : number of repetitions (the paradigm is supposed to be periodic).

The estimation of $\mathcal{D}$ cannot be obtained exactly in some cases ; thus we propose an upper-bound, whose adequacy is expressed in terms of precision on the next table.

Test	$\mathcal{D}$	Precision
ANOVA	$F_{k-1,N-k}$	exact
MI	$\frac{1}{2\sqrt{2}N}\chi_d^2$, with $d = (k-1)\frac{2.\sigma.\sqrt{2}.erf^{-1}(1-\frac{\log 2}{N})}{\sqrt{2\pi\beta^2}}$	exact
ANOVA+Memory	$F_{4(kl-1),N-4kl}$	exact
MI-2D	$\frac{1}{4N}\chi_d^2$, with $d = (k-1)(\frac{2.\sigma.\sqrt{2}.erf^{-1}(1-\frac{\log 2}{N})}{\sqrt{2\pi\beta^2}})^2$	upper-bound
Markov+ANOVA	$F_{k-1,R-k}$	upper-bound
Markov+MI	$\frac{1}{2\sqrt{2}R}\chi_d^2$, with $d = (k-1)\frac{2.\sigma.\sqrt{2}.erf^{-1}(1-\frac{\log 2}{R})}{\sqrt{2\pi\beta^2}}$	upper-bound

3.3 Estimating the impulse response

3.3.1 Linear model

An activation map gives only the localisation, the extent and the intensity of brain activity related to the experimental conditions. We think that complementary -temporal- information is present in the activation pattern itself. We propose to characterise it through an impulse response (see [KvCD99] for example) which is the hemodynamic response when studying a BOLD signal. The impulse response is readily obtained when the design is event-related, but it has to be inferred from the data in the case of a block experimental design : Let us denote by $a_x(t)$ the activation pattern at a given voxel x, under paradigm $p(t)$; we define the impulse response as the kernel $h_x(t)$ so that $a_x(t) = (h_x * p)(t)$. To simplify notations, the subscript x will be omitted in this section. There remains to relate the observed signal $y(t)$ with the activation pattern $a(t)$.

- If the experimental design is periodic with period T and R repetitions, $a(t)$ will simply be the average $a(t) = \frac{\sum_{i=1}^{R} y(t+iT)}{R}$,
- Else, we can still define the *state variable* $s(t) = (p(t), p(t-1), .., p(t-k))$ then $\alpha(s)$ as the conditional expectancy $E(y|s)$ and $a(t) = \alpha(s(t))$.

In both cases, we postulate the following relationship, which can be viewed as a specialisation of (4):

$$y(t) = a(t) + n(t) = (p * h)(t) + n(t) \qquad (6)$$

Then, supposing $n(t)$ is i.i.d. gaussian, a least-squares estimate of h can be found. From our experience, unlike [DB97], taking into account temporal autocorrelation of the noise has no effect on the results (since most of the temporal autocorrelation is in fact included in the convolution mask p). The hypothesis of gaussianity of the noise is also correct in practice: A Kolmogorov-Smirnov test on gaussianity shows that for 99.5 % of the voxels, the residual noise obtained from equation (6) has a gaussian statistics with a confidence level of 95%. This ensures the optimality of our computation.

IR is estimated separately for the different stimuli. It appears that this separation is crucial in order to interpret correctly the experimental results (see section 5.3).

Let us notice that only finite impulse responses are considered here (infinite responses could be considered, but their estimation and their interpretation would be difficult). A parameter of importance is the length of the impulse response h. With a block design, only few terms in the IR are really significant. In our experiments, 10 terms where computed. It seems that only the first 7 ones are significant.

3.3.2 Relaxing the linearity hypothesis

The impulse response as estimated above is in fact very noisy. The reason is that it is obtained from a deconvolution process known to be very unstable numerically. Furthermore, if the sampling frequency is low, the signal contains quite a lot of aliased artefacts. Thus there is a need for regularising the result. Several methods are available :

- Parametric fitting: h is fitted to a family of known functions and we select the best one : This is efficient, and helps translating the IR parameters into more intuitive ones. But it biases the observations towards a priori models.
- Smoothing y or h with a kernel (a gaussian kernel for example). It is easy and efficient, but the maxima are blurred, and important distortions are added to the signal : the peaks of h can even be moved.
- Adding an adapted penalty to equation (6) : $\phi(|h'(t)|)$, where ϕ -usually a bounded function- is known to preserve discontinuities (see [AV97]). This may be the best choice, but determining ϕ is not easy.

We choose the last method with $\phi(x) = \frac{x^2}{x_0^2+x^2}$, where x_0 is a suitably tuned parameter. The effect of such a function is quite strong since it is bounded, thus very far from the parabolic model. Our experiments show the ability of this method to de-noise the data without smoothing out the main peaks (see figure 7).

Let us end up this paragraph with two remarks:

- It is questionable whether the penalty should be a function of $h'(t)$ or $h''(t)$. The disadvantage of h' is that it may smooth out the main discontinuities. On the other hand, the estimate h'' is extremely noisy, especially when the time lag between two scans is long with respect to the physiological phenomena ($3.3s$ in our reference case). Since the use of a ϕ function enables to overcome the first difficulty, we have chosen a $\phi(h')$ penalty.
- Our estimation method adds two parameters to the problem : One to tune the shape of the ϕ function, and another one to weight the additional penalty. We tuned these constants empirically, since there are no a priori constraints on the expected solution.

3.4 Clustering of voxels

The question we address is to characterise the different regions of the brain involved in a given task. More precisely, we would like to know the level and the characteristics of the different activations in the different parts of the brain.

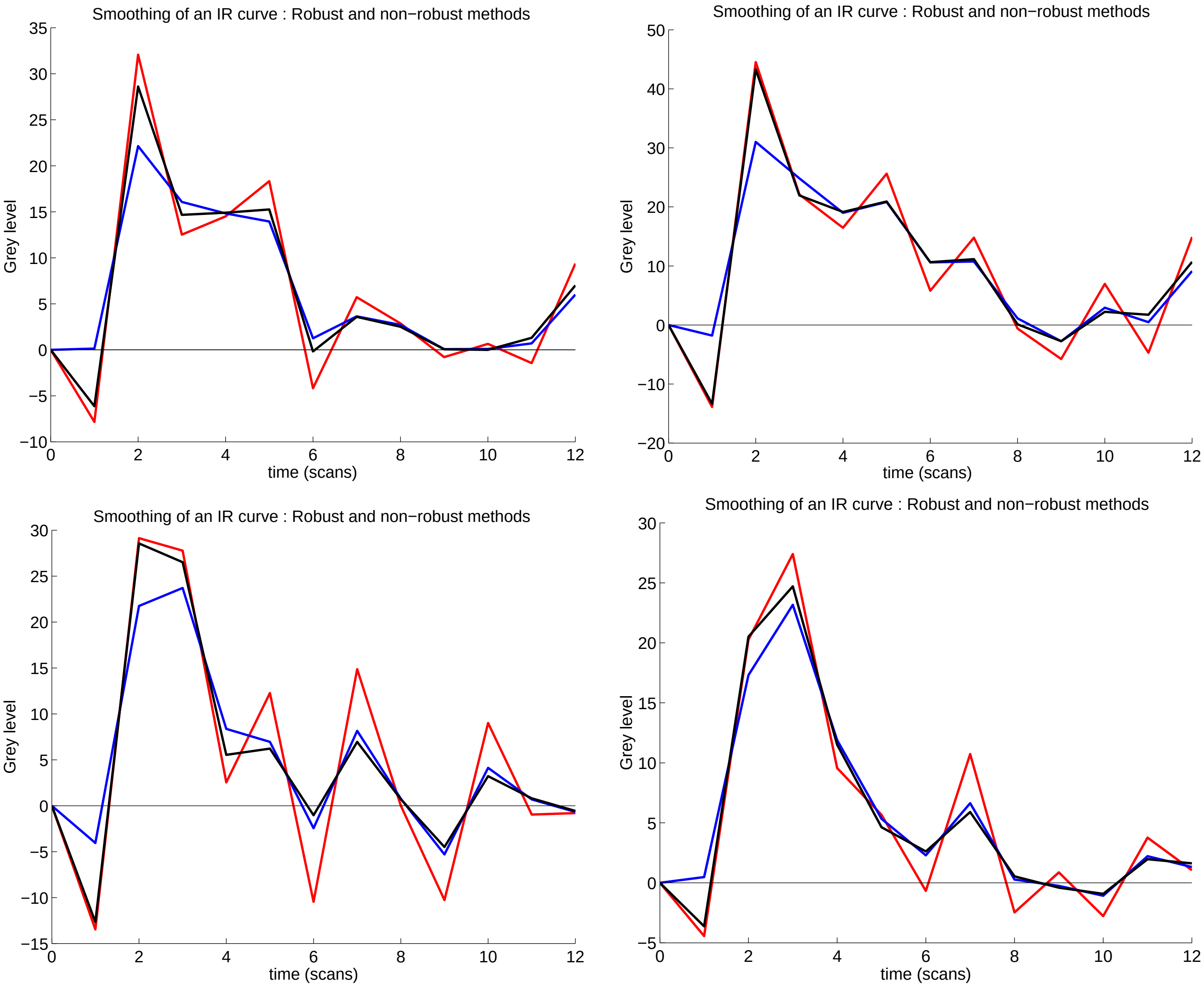

Figure 7: 4 examples of robust and non-robust methods for impulse response filtering. The *raw* and *noisy* I.R. curve is plotted in red, a gaussian-filtered version is plotted in blue and a robust-filtered version in black. Robust filtering enables the reduction of noise of the initial curve while keeping its discontinuities, whereas these are blurred during gaussian filtering.

A question is whether we need to take into account the neighbouring voxels. In fact, spatial proximity should be taken into account only if it is anatomically validity plausible, according to [AKM$^+$01]. For example, though often used ([SKvC00]), an isotropic distance is certainly not suitable. In particular [ZAD97] have noticed that the spatial coherence could not be explained by a smooth autocorrelation function. In the work reported here, we did not take spatial distance into consideration.

Our approach is very close to [GHLR99]. We use a Fuzzy C-means algorithm : the feature considered here is the IR, as computed in the estimation step, and expressed as a vector $H = (h_0, h_1, .., h_{k-1})$ in R^k, with k=10 for example. The computational burden is decreased through a threshold in the voxel activity to keep only those with significant activation score -the threshold can be determined from the statistics obtained under the hypothesis of null-distribution (see section 3.2). Moreover, the IR obtained for voxels with low activation score are essentially noise, and perturb the results of the algorithm. In fact, there are two main difference with existing solutions :

- The input data of the algorithm is pre-selected thanks to the threshold step.
- The input data is significant, since we cluster the IR instead of the raw time serie (in contrast with [BWM98], for example).

The Fuzzy C-means (FCM) algorithm [DK97] let $X = (x_1, .., x_N)$ be a set of points of R^d. A FCM algorithm partitions X into c fuzzy subsets $u_1, .., u_c$. Let $u_i(x_n)$ be the membership in class i and $v_1, .., v_c$ be the centers of the clusters. Then the FCM minimises the functional

$$J(U, V, X) = \sum_{i=1}^{c} \sum_{n=1}^{N} (u_i(x_n))^m \|x_n - v_i\|^2 \tag{7}$$

Where $m > 1$. After random initialisation of the v_i, the weights $u_i(x_n)$ and the centers v_i are updated alternatively until convergence is reached. The centers are an average of the points weighted by u_i, and $u_i(x_n)$ is updated by

$$\forall i, n, \; u_i(x_n) = [\sum_{j=1}^{c} (\frac{\|x_n - v_i\|}{\|x_n - v_j\|})^{\frac{2}{m-1}}]^{-1}$$

In our experiments, we found $m = 1.5$ to be optimal for the speed of convergence.

Important issues The algorithm raises some questions that we are currently investigating :

- The number of classes is unknown a priori, but can be determined through an information theoretical criterion ([GHLR99]). From our own experience, the simple BIC criterion seems the most accurate.
- Use of Euclidian vs Mahalanobis distance : empirically, we found that a Mahalanobis distance may lead to poorer results than the usual Euclidian distance. In our view, the reason is that the Mahalanobis approach makes the variations in the data isotropic in the observation space (say R^d) ; this is certainly not suitable here, since some coordinates of the impulse responses (e.g. at times t=2,3,4 scans) are much more significant than other coordinates (e.g. at times t=1,5,6 scans) ; this implies that some dimensions of the data $d = 2, 3, 4$ contain much more informations than the others.

- Finding a global optimum of (7) can be guaranteed by repeating the process with different random initialisations, since it converges towards a local minimum.

In our view, clustering is a way to take into account quantitative(activation score) as well as more qualitative information (IR). It is also a way to average the observations -the coordinates of the centroids being the average of the coordinates of the points of the clusters- the advantage being that this way of averaging is data-driven. Some results are presented below (see section 5.4).

4 Comparison of the activation detectors

The comparison of activation detectors has two goals : first, to study the ability of the tests to discriminate between activated and non-activated signals, second to evaluate the results of each method (test and statistical assessment) in realistic conditions. The first part has been performed on synthetic data since there is a priori no ground truth on real data. The second part has been performed on real data to provide a more realistic assessment.

4.1 Simulated data

We have generated *i)* $K = 10^4$ synthetic sequences $\{y_t\}$ of length N that are purely correlated gaussian noise, and *ii)* K sequences with an added activation pattern. The signal has an AR(1) structure :

$$y_t = \rho y_{t-1} + a_t + z_t \, , \; t = \{1, .., N\}$$

where $\rho \in [0,1]$ is the regression parameter, $a_t = 0$ when no activation is simulated ($i)$) or $a_t = 0|a$ alternatively according to the simulated paradigm for activated signal ($ii)$), and $z_t \rightsquigarrow \mathcal{N}(0, \sigma^2)$ and N is the signal length . In our case, $\frac{a}{\sigma} = 0.5$, $N = 400$, the paradigm is binary, made of 20 blocks of 20 scans, ρ was set to 0.5 and 0.8 to study the influence of noise autocorrelation.

The results for the six proposed tests (Anova, Anova+memory, Markov+Anova, MI, MI-2D, Markov+MI) are presented as ROC curve : the detection rate in *ii)* is plotted as a function of the false-positive rate in *i)*. It is curve in the $[0,1] \times [0,1]$ square. Ideally it should be reduced to : $\{0\} \times [0, 1[\cup]0, 1] \times 1$. In practice, it is more or less close to that limit. The results are presented in figure 8. Only the upper left corner of the curves is displayed, since it contains the relevant information.

We observe that the six proposed methods fall into two classes :

- Anova, MI work better for low autocorrelation levels $\rho = 0.5$.
- Anova+memory, Markov+Anova and Markov+MI work better for higher correlation rates.

MI-2D has relatively poor results in both cases. It may be because the method needs a fine tuning of its parameters and more data to be efficient.

These results are not surprising, since the methods with Markovian pre-processing have been designed to model transitional and, more generally, correlated aspects in the signal. The same holds for the *Anova+memory* model. The lower results obtained with these methods under low correlation can be explained by the loss of

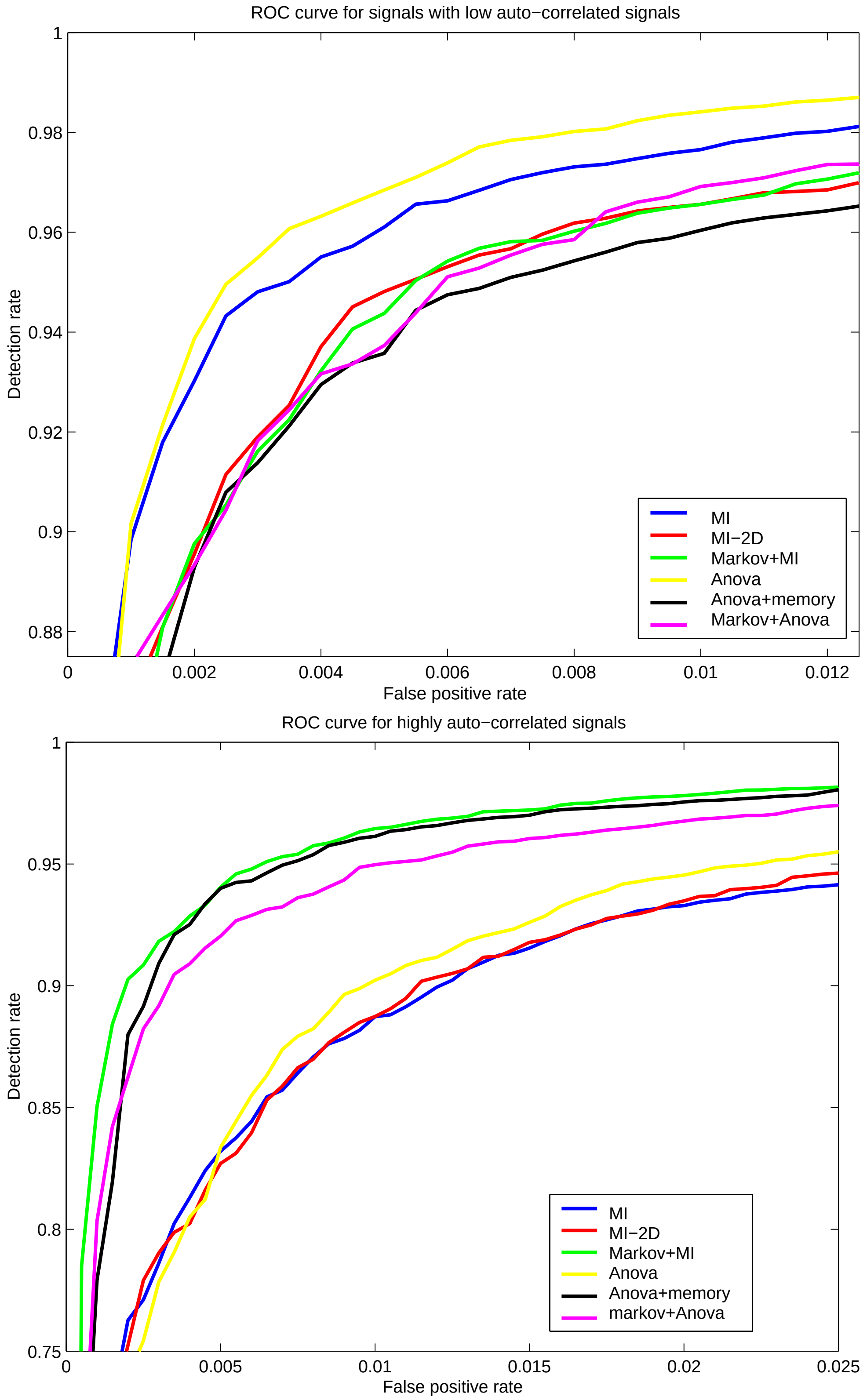

Figure 8: ROC curves of the statistical tests : the rate of activation detection is plotted against the rate of false-positive. The six curves correspond to yellow : Anova; black : Anova+memory ; magenta :Markov+Anova ; blue : MI; red: MI-2D; green : Markov+MI. Left $\rho = 0.5$ (low autocorrelation) ; right $\rho = 0.8$ (high autocorrelation).

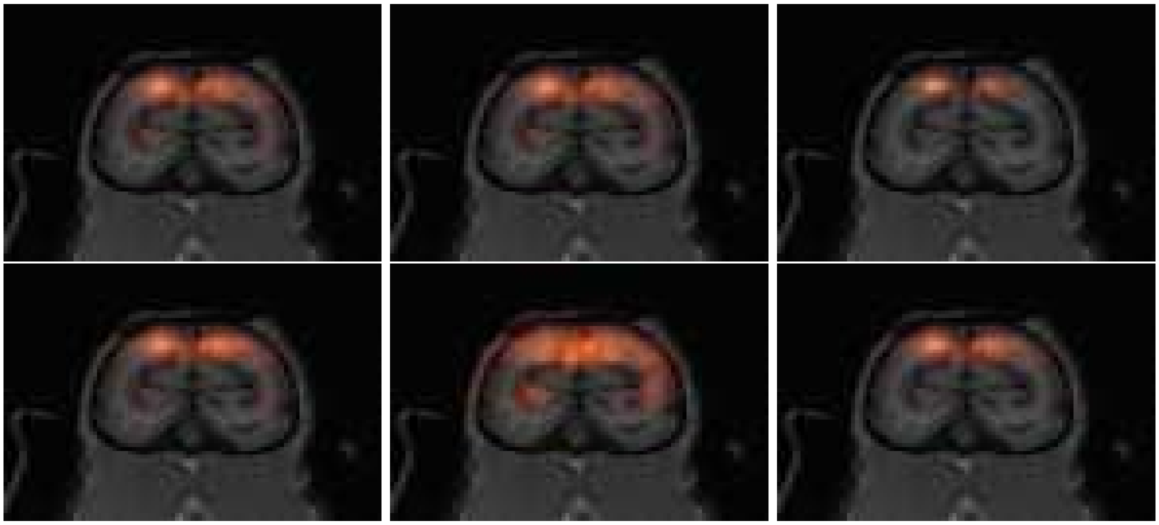

Figure 9: Comparison of the activation tests at a coronal slice. Colour intensity and brightness represent the result of the test scaled by the statistical assessment function. From up left to bottom right the statistical tests are : Anova, Anova+memory, Markov+Anova, MI, MI-2D, Markov+MI. The width of activation loci is related to the ability of the test to distinguish between large and low activation. For example, the result with MI-2D is blurred, which indicates a poor efficiency (this may be related to the weakness of the method and corresponding statistical assessment)

statistic power of these methods (indeed, their ground model is a non-stationary signal, while basic MI and Anova methods have all the statistical power of stationary models).

A possible conclusion could be that the performances of the Anova+memory, Markov+Anova and Markov+MI models do not depend much on noise characteristics, making these methods more suitable in general.

4.2 Real data

Here the purpose is to evaluate the results obtained with the tests on real data in relation with the proposed statistical assessment. Since the results cannot easily be checked we consider that it is difficult to quantify the efficiency/specificity of the tests. Thus we give a rather qualitative evaluation. The experiment has been made on the sequence described in the next paragraph. The scans corresponding to *baseline* and *stimulus 2 : moving texture* have been retained. For each voxel of a coronal slice of the occipital lobe, the six tests t have been performed, and the statistical significance threshold s computed. The ratios $r = \frac{t}{s}$ are displayed for the six methods on the anatomical slice. Only the values of r that are greater than 1 have been considered. See figure 9.

The resulting maps are to be viewed as follows : tight activation peaks indicate an ability to distinguish between strong and weak activations. With respect to this property, one has the following ranking :

$$\text{Markov+Anova} > \text{Markov+MI} > \text{MI} > \text{Anova+Memory} \simeq \text{Anova} > \text{MI-2D},$$

which is understandable : Markov pre-processing tends to improve the strength of great activation. On the other hand, MI-2D seems to be sensitive to measures artefact, since it relies on a 2-D plot of the data ; thus the performed estimations are fragile, and due to the boundedness of mutual information, there is little margin to characterise activation peaks.

Caveat: It should not be inferred from the previous remarks that activations detected with MI-2D are not less reliable than with other techniques : activations encountered with MI-2D -for example near the calcarine sulcus on figure 9- are systematically also encountered with other detectors, which in turn give a better idea of peak heights.

4.3 Conclusion on activation detectors

In our view, two main conclusions can be drawn from the previous experiments :

- The techniques that model temporal aspects in the data (memory model, Markov pre-processing), may be less efficient when the data is weakly correlated. On the other hand, their efficiency does not depend on the correlation in the data, which makes their use safer.
- Markovian pre-processing is a powerful tool to distinguish between low, medium and high activation strength.

MI-2D seems to be weaker than other techniques to assess activation strength. Some work may still be necessary to make it a better tool.

5 Experiments with data from Leuven

5.1 Description of the data

Our data is a set of 15 runs of 80 scans of a monkey performing a vision task. Image resolution is $3 \times 3 \times 3\ mm$, and temporal resolution $TR = 3.321s$; $TE = 32msec$. Coronal slices were acquired sequentially, from front to rear. The size of the images is $128 \times 128 \times 32$ voxels.

The monkey is a 4 years old male rhesus, whose skull is maintained fixed during the experiments with a MR-compatible headset and plastic screws. He was trained during several months to fix a screen while staying still, and used to noisy environments.

Data was acquired with a 1.5 T Siemens Vision scanner equipped with echo-planar imaging. During the experiment, visual stimuli were projected from a LCD projector using customised optics onto a screen which was positioned 15cm in front of of the monkey's eyes.

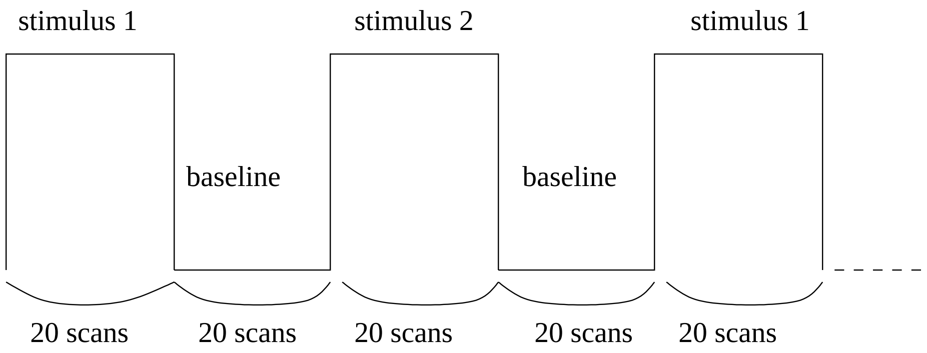

Figure 10: Experimental paradigm of our data. Epochs of two stimuli alternate, separated by baseline periods. Each epoch is 20 scans ($\simeq 66 sec.$) long. Stimulus 1 means staring at a static set of dots, Stimulus 2,staring at a moving set of dots.

The experimental paradigm was the same for the 15 runs : during the first 20 scans, the monkey observed a static texture (a texture is a set of dots with random aspect), followed by 20 scans rest (the monkey looks at a fixation cross) , then a uniformly moving texture is projected during 20 scans followed again by 20 scans rest. The uniform motion direction changed randomly every 427 msec in 45 degree increments, and the uniform speed was adjusted to 6 degrees/second. The temporal layout of the experiment is illustrated in figure 10. During the experiment, eccentricity and luminance were carefully equated across conditions.

The monkey was administrated a super-paramagnetic contrast agent called mion, so that the measured signal does not correspond to bold effect but to blood volume. In particular, activations are in fact negative (decrease of the signal). Let us notice also that mion and BOLD effect may also compete in the final signal. This is of course an additional reason not to use usual predefined models for BOLD response.

5.2 Activation maps

We have chosen to present 3 activation maps produced from the data ; the test being *Anova+memory* in each case : the first activation map is based upon the whole experiment ; the following tests are restricted to the study of one of the stimulus conditions + baseline condition. In each case, a significance threshold corresponding to a probability of activation under the null condition of 10^{-4} has been computed, and only supra-threshold scores are displayed.

Figure 11 represents the activation map obtained taking into account all the scans (baseline and stimuli 1 and 2). 20 coronal planes separated by 2mm are displayed.

This gives a first idea of the regions of interest in the experiment : middle temporal gyrus, occipital gyrus (middle anterior and posterior parts), annectant gyrus, and possibly the upper part of the lingual gyrus.

In the sequel, the activation conditions are studied separately, in order to determine how they contribute to a global effect. This could also be achieved through a kind of “subtractive” analysis, but our analysis of

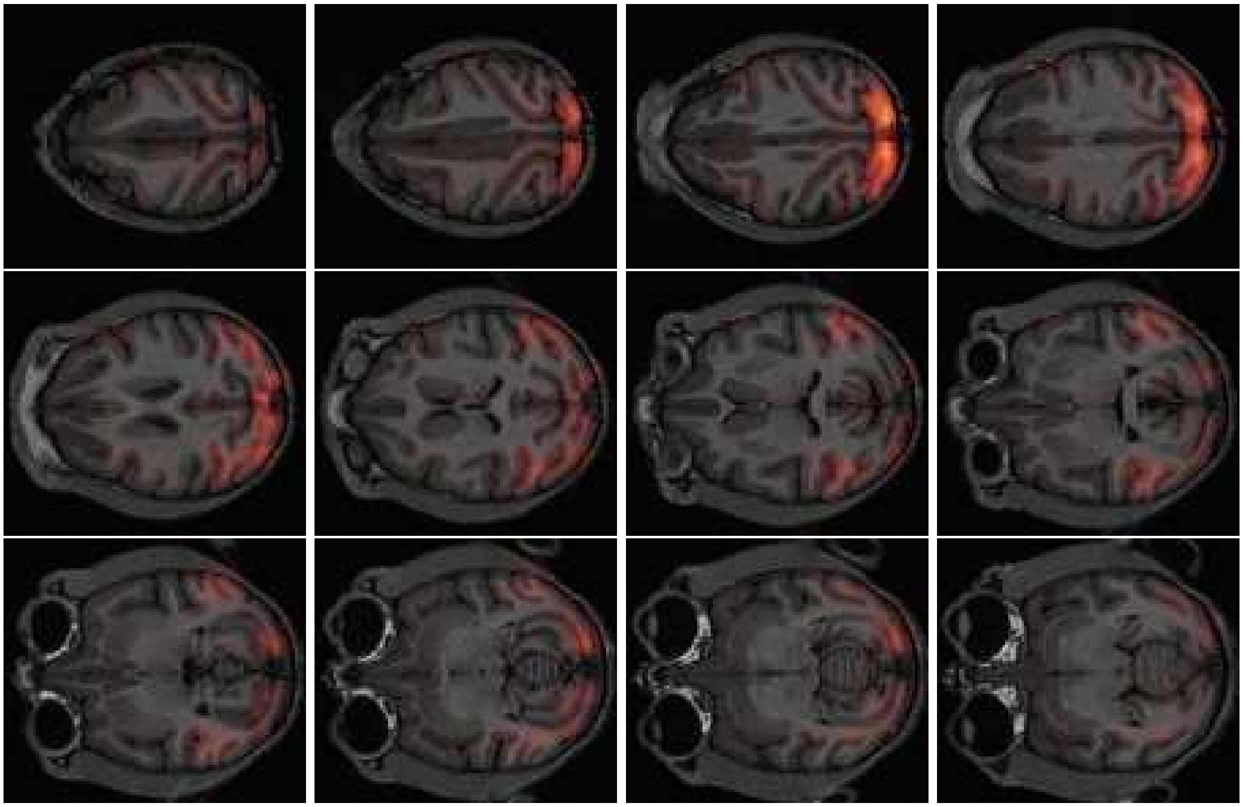

Figure 11: Activation map obtained from considering the whole experiment, and displayed in horizontal slices: Activity seems to be quite diffuse, but some foci are visible. The slices are 2mm apart from one another.

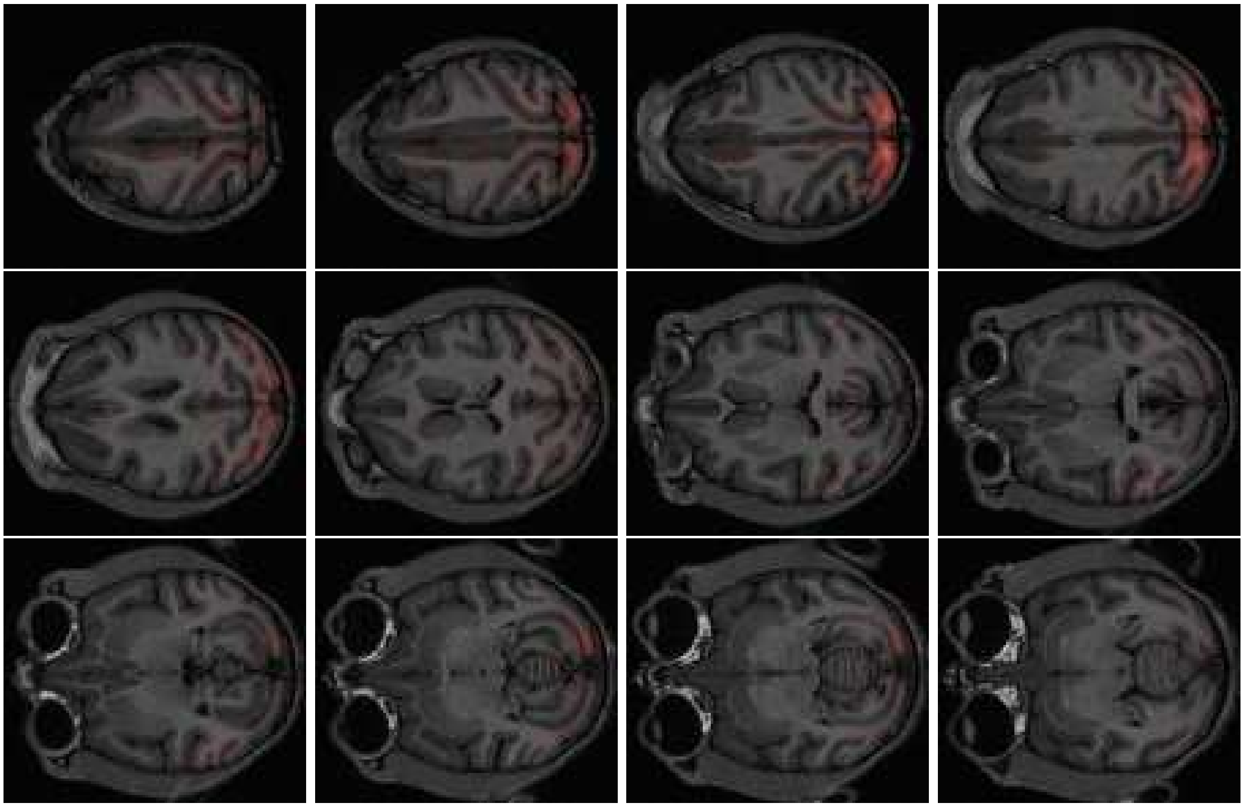

Figure 12: Activation map obtained with stimulus number 1 : static texture. The -horizontal- slices are 2mm apart from one another.

variance is not really suitable for such purposes. Moreover, in a first approach, we prefer to perform a more descriptive analysis on the data.

Figure 12 represents the activation map obtained for the first stimulus (static texture). 20 horizontal slices separated by 2mm are displayed. A few loci of the brain are above the threshold, but the maxima are clearly located in the occipital gyrus. However, they are far below the maxima obtained for the other stimulus.

Figure 13 represents the activation map obtained for the second stimulus (uniformly moving texture). 20 horizontal slices separated by 2mm are displayed. Activations are localised mainly in the upper posterior part of the occipital gyrus, but also in the posterior part of the annectant gyrus, and in the middle temporal gyrus and very weakly in the upper part of the lingual gyrus.

To sum up our activation detection results on this data, we give an additional map presented on coronal slices of the monkey's brain in figure 14. The colour map is related to which type of activation occurs at each voxel : blue for stimulus 2 only activation, green for activations under both conditions (there is no occurrence of activation under stimulus 1 only).

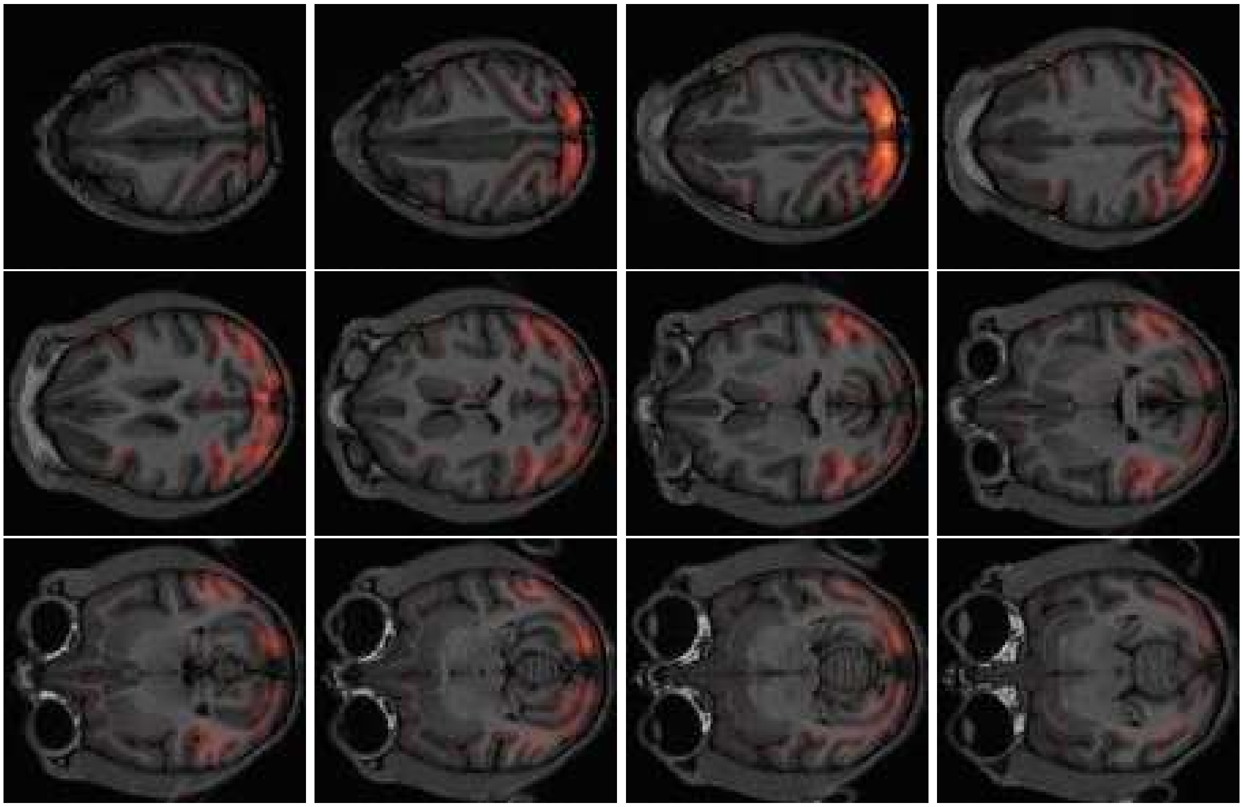

Figure 13: Activation map obtained with stimulus number 2 : uniformly moving texture. The -horizontal-slices are 2mm distant from one another.

The results show clearly activations under both conditions in V1 cortex, whereas activations under stimulus 2 only occur in the MT/V5 area.

5.3 Local parameters

We present the impulse responses computed for the two stimuli. After being computed on each voxel separately, the impulse responses are clustered. Each cluster can be given an activation score, which is the average activation scores of the voxels of the cluster. This enables us to distinguish between the most significant clusters and the others. In our case, 9 clusters -this number being justified through a BIC criterion on the clustering functional - have been derived for each condition, and each time the 5 best ones have been kept and displayed. Each of the clusters represents a population of a few dozens of voxels at least : The curve that represents a cluster is the weighted average of the members of the class. Hence it is not too noisy.

Before describing the results, let us remind the reader that :

- In our case, the impulse response is not a classical hemodynamic response but a mion response, since the signal mainly comes from mion volume present in the voxel. The mion reduces the signal, thus we have plotted in fact the opposite of the impulse response to give it a more usual aspect.
- The experimental design is explicitely a block design, with long blocks (20 scans $\simeq 66s$). The impulse responses are recovered from the first and last scans of each epoch with a linear deconvolution technique. Moreover, the sampling grid is large with respect to the typical impulse response delay. For all those reasons, the resulting curves are quite noisy. We concentrate our description on basic features : delay and amplitude of the maximum, presence of dips.

Stimulus 1 "static" See figure 15 : Though 10 scans of impulse response are plotted, probably only the 5 or six first ones are significant. In order to interpret these results properly, it is useful to remember the hierarchy *red ¿ black ¿ blue ¿ green ¿ yellow* which is provided by the average activation score of each cluster. Each curve presents the usual features of impulse responses : initial dip, peak, decay, but the maxima do not occur at the same time for the different curves : scan 2 for the blue curve, scan 3 for the red curve, scan 4 for the black curve ; this clearly suggests that activations of the corresponding areas occur successively. Note that slice acquisition is sequential, so that the delay between the curves cannot be explained by an artefact related to the temporal acquisition. The green and yellow curves have a maximum at scan 3, but the main increase in yellow area occurs clearly earlier than in the green one.

Stimulus 2 "moving" See figure 16 : The impulse responses are now quite similar with the usual models for four of the curve models (the red, black, blue and yellow ones) : initial dip, maximum, slower decay. Differences are found in the amplitude and delay of the response, maximum at scan 2 for the red and blue models, at scan 3 for the yellow and black models. The yellow curve reaches surprisingly a high maximum, but the width of the peak is inferior to the other models. The green curve behaves differently from the others : no initial dip, and the maxima is reached after four scans. It seems to indicate a slow, smooth response.

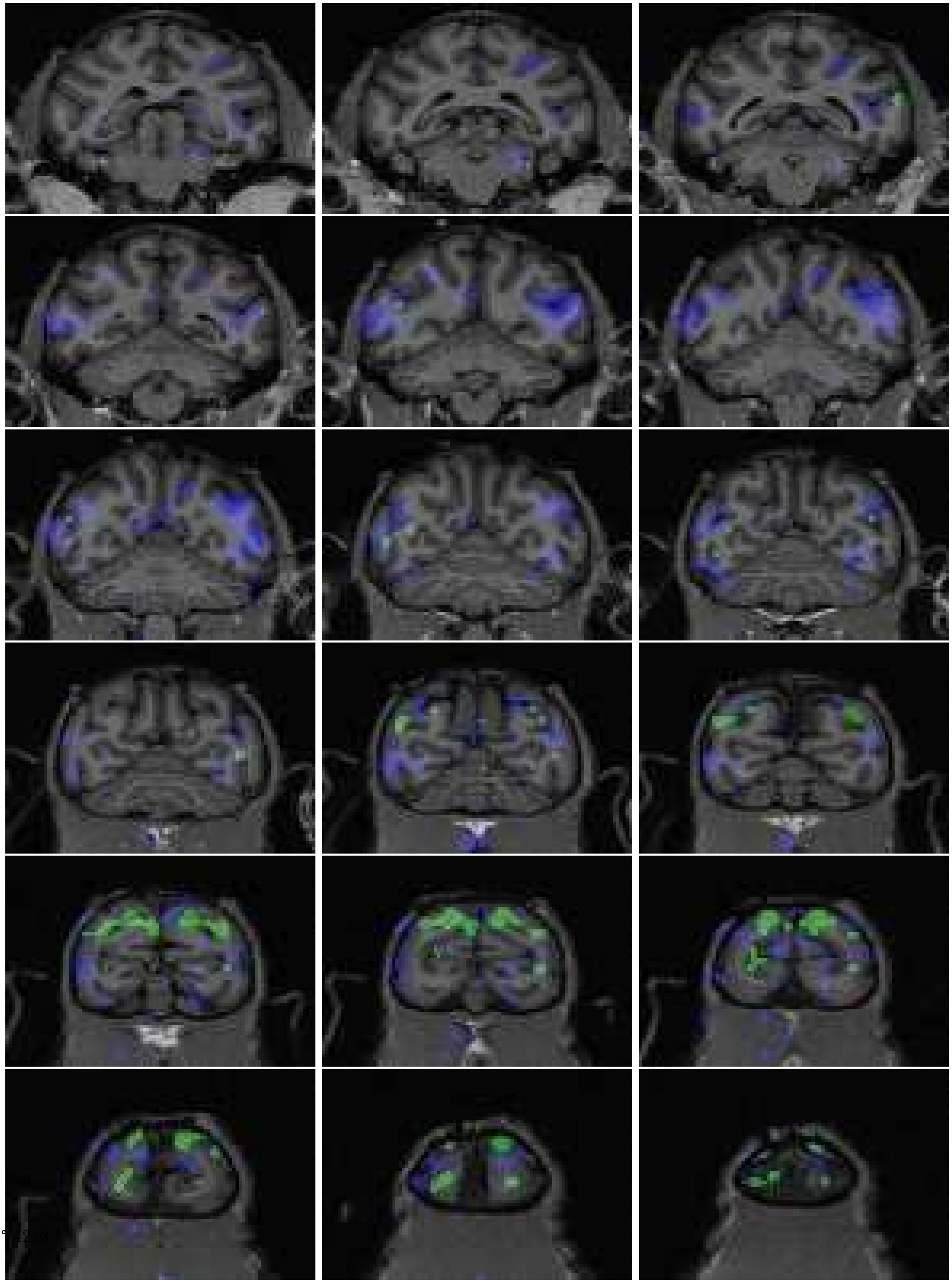

Figure 14: Coronal views of the rear part of a monkey's brain. The green areas correspond to activation areas under both activation conditions, whereas the blue areas correspond to activations under stimulus 2 only. No voxel had activation under stimulus 1 only. The statistical test used here is mutual information; an activation score is computed for each experimental condition separately. A voxel is said active under a given condition when its activation score overcomes the adapted threshold.

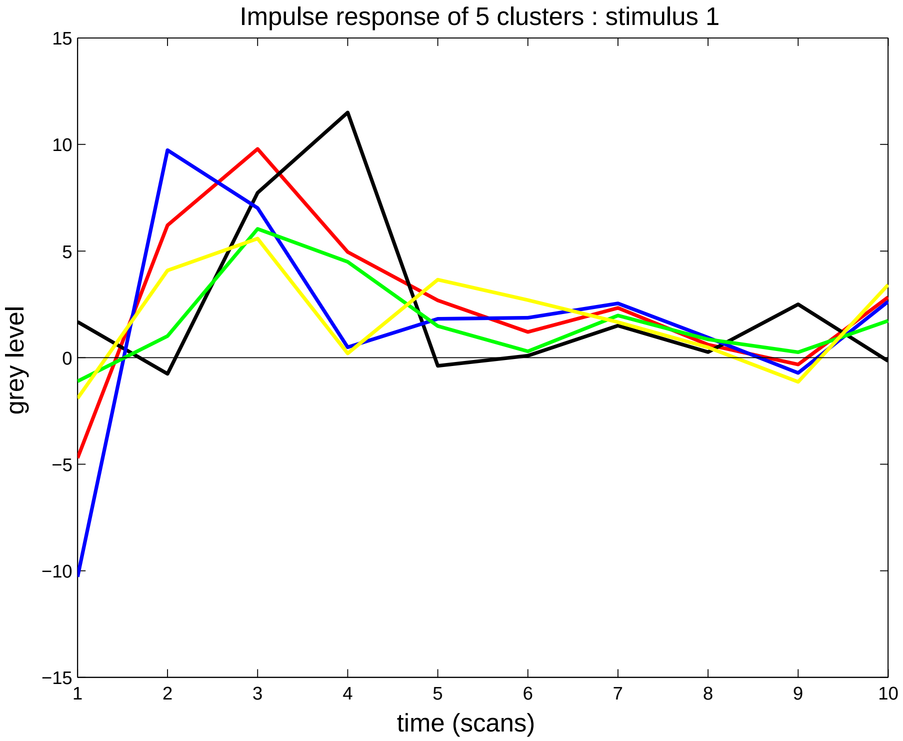

Figure 15: Plot of the impulse response for the first stimulus. The responses obtained for the first 5 clusters are plotted here. Average activation scores have been computed for each cluster: the highest scores are for the red cluster, then the black one, then the blue one, the green one and the yellow one.

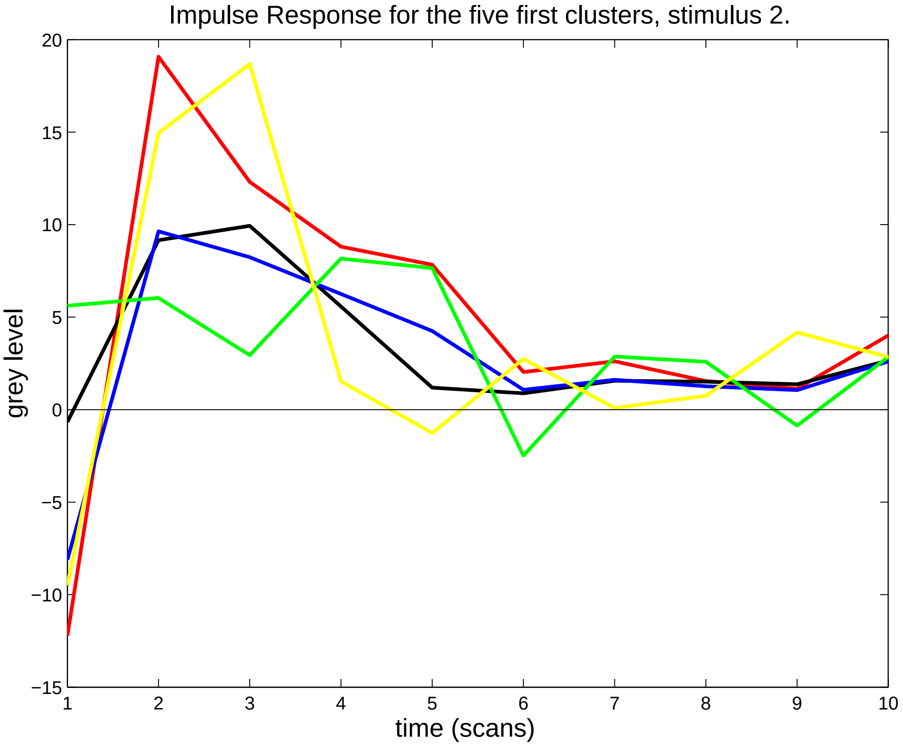

Figure 16: Plot of the hemodynamic response for the second stimulus. The responses obtained for the first 5 clusters are plotted here. Average activation scores have been computed for each cluster: the highest scores are for the red cluster, then the black one, then the blue one, the green one and the yellow one.

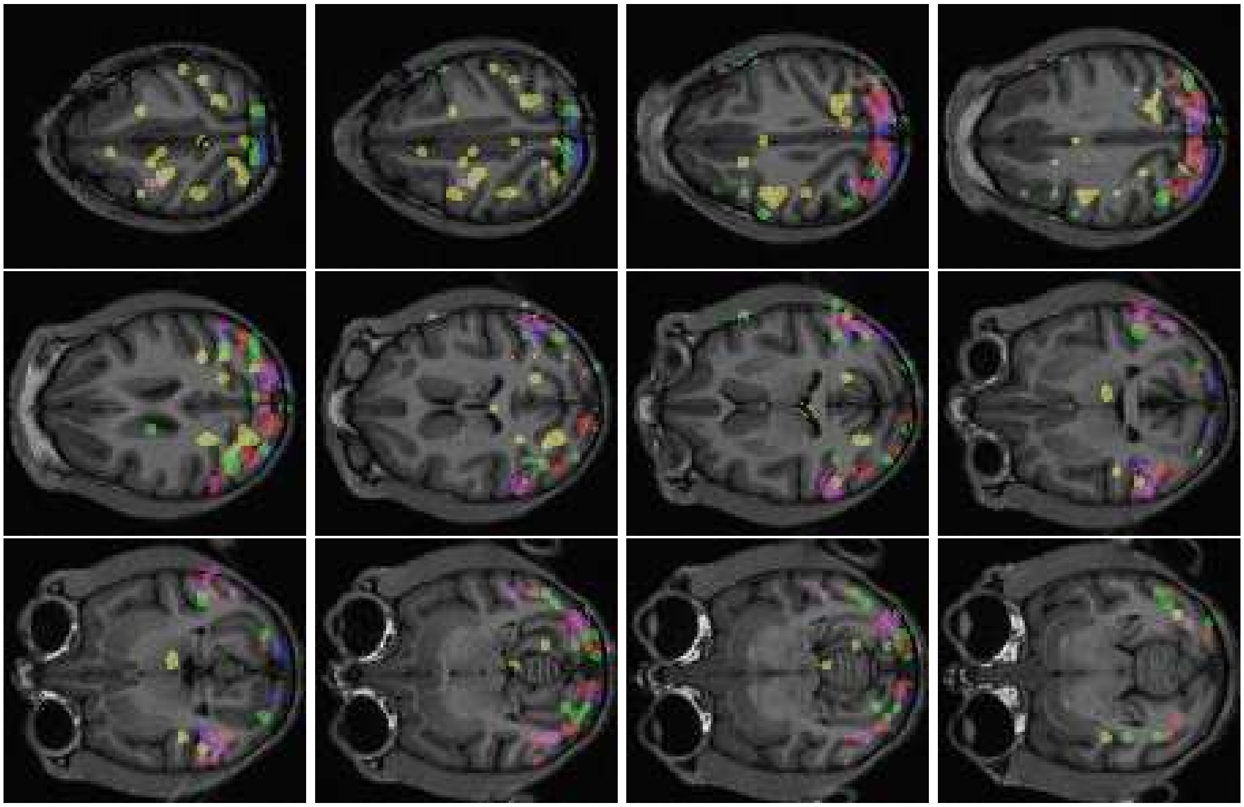

Figure 17: Clustering on activated areas with stimulus number 1 : static texture. The slices are 2mm apart from one another.

5.4 Clustering

We present, separately for each condition, the regions corresponding to the different kinds of detected activations. The colour of the region is the same as the corresponding curve for the impulse response (but the black curve is displayed in purple). Only the supra-threshold areas are represented. Even if spatial correlation has not been taken into account, one can notice that the activated areas are quite homogenous spatially. Not astonishingly, the red and purple areas correspond to the maxima of activation scores. See figures 17 and 18.

stimulus 1 : The red, blue and purple areas are quite intricate; they make up the posterior part of the occipital gyrus, and probably indicate the successive activation of the different parts of visual cortex V1. While the yellow and green ones are situated in more lateral parts of the gyrus, where activations are lower.

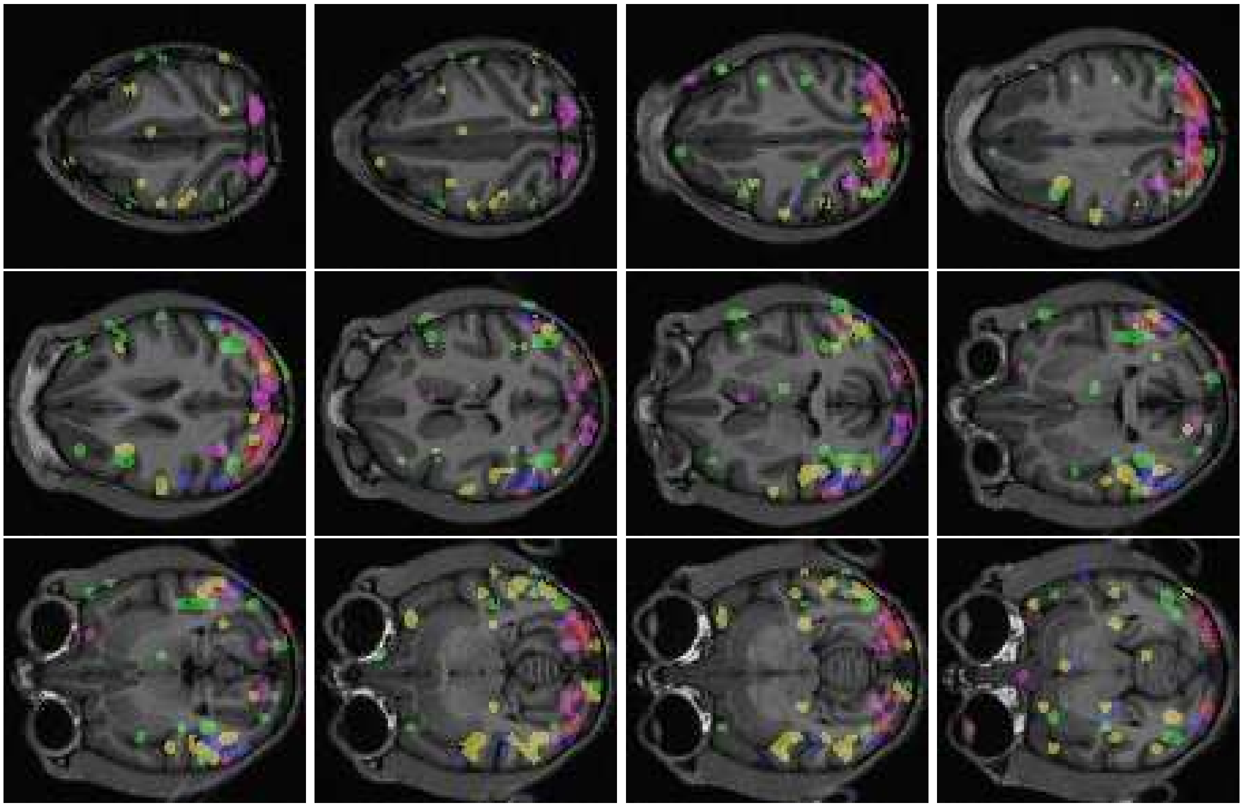

Figure 18: Clustering on activated areas with stimulus number 2 : uniformly moving texture. The slices are 2mm apart from one another.

stimulus 2 : The *red*, *blue* and *purple* areas make up the visual cortex V1 that is in the posterior part of the occipital gyrus. At this place, activations are maximal ; these two areas are quite interleaved, thus not distinguishable; this calls for more precise study (and data), since the *red* activation seems to occur slightly earlier than the other one. The *yellow* area also belongs to the occipital gyrus, but it is more peripheral, in places where activations are lower. The *green* part seemingly belongs to the MT area.

5.5 Interpretation

The Effect of moving stimulus on the activation is two-fold :

- It enhances the observed activation in the occipital lobe (upper posterior part)
- It lets activations appear in the middle temporal and angular gyri.

A more careful observation of the impulse response shows that activations inside of the visual cortex V1 occur at different time lags, which calls for a more precise spatio-temporal analysis in that region. The behaviour in MT parts is also of interest : it is less standard than in other areas (see the green curve in figure 16), which indicates a complex relationship with visual areas.

6 Conclusion

We have proposed a way of detecting activation reliably, with a minimum of hypotheses, avoiding thus any bias in the statistical inference : in particular, no model of impulse response is used, and the hypothesis on the noise are kept as free as possible. Models for values of the tests under the null hypothesis have also been derived, allowing for reliable activation detection. In our view, a important improvement in activation detection can be obtained by a careful modelling of the transitions between activation levels.

We consider that robust computation of the impulse response and spatial clustering of these responses is also a first step towards a complete spatio-temporal model of the brain during activation. Even if block designs are not suited for connectivity and causality analysis in the brain, we think that these experiments provide statistically reliable information on the activation loci, from which inferences can be drawn with much profit.

Some improvements are still to carry on to our work :

- Studying the influence of initial coregistration of the scans on the activation tests. Activation detection should rely on regions rather than individual voxels. This is however linked to an increase in data resolution.
- Adapt our methods to more complicated experiments : event-related paradigms, shorter blocks, variable stimuli length.
- Enabling different kinds of inference from the activation maps (subtraction, conjunction)
- Taking into account anatomy constraints in the data processing.
- Embed our description in a physiologically more convenient framework. Simple ones have already been proposed ([VN98], [KvC99]) and could be improved.

7 Acknowledgements

We wish to thank Professor G. Orban, coordinator of the laboratory of Neuro- and Psychophysiology, R.U. Neurophysiology, Dept. of Neurosciences and Psychiatry, K.U.Leuven, Medical School, Leuven, Belgium (LEUNEURO), who provided us with the functional MR images we used to test our algorithms, and gave us insight in the study of monkey vision, as well as Wim Wanduffel and Denis Fize, whose advice helped us in the design of our methods.

We also wish to thank some members of Robotvis project : P. Kornprobst, who helped us to develop robust deconvolution methods, G. Hermosillo for his insight in information theoretic techniques, and T. Viéville for his advice on many fundamental and formal aspects of this work.

A Mutual Information from a statistical point of view

A.1 Artefactual Mutual Information obtained in the case of a binary paradigm

The goal of the section is to provide a meaningful threshold on the mutual information values obtained empirically between the time series and the experimental paradigm. The question is raised but unsolved in [TFW$^+$99] and [KFT$^+$00]. As in most approaches, our goal is to avoid false-positive, that is to reduce the probability of detection in the case of null hypothesis. This involves studying the distribution of the results coming from input noise.

Let us consider the following case : the paradigm p is "binary" i.e. it results from two conditions (0/1) temporally interleaved. Let us suppose that the observed data is a random variable $y(t)$ with a given statistical distribution h on a domain X. Let h_0 and h_1 be the distribution conditionally to the events $(p = 0)$ and $(p = 1)$, and let π_0 and π_1 be the probabilities of $(p = 0)$ and $(p = 1)$ events $(\pi_0 + \pi_1 = 1)$. Then

$$h(x) = \pi_0 h_0(x) + \pi_1 h_1(x). \tag{8}$$

The mutual information between $p(t)$ and $y(t)$ is defined as

$$MI(y,p) = -\int_X (h(x)\log(h(x)) - \pi_0 h_0(x)\log(h_0(x)) - \pi_1 h_1(x)\log(h_1(x)))dx.$$

Theoretically, under a null hypothesis, one should have $h_0 = h_1 = h$ thus $MI(y,p) = 0$. In fact, the estimation of h_0, h_1 and h is not perfect, and there is a residual noise in the estimate of MI (which is always positive). Our problem is to evaluate this noise :

Let us note for $i \in \{0,1\}$ $z_i(x) = \frac{h_i(x)-h(x)}{h(x)}$. Then (8) yields

$$\pi_0 z_0(x) + \pi_1 z_1(x) = 0. \tag{9}$$

One gets :

$$MI(y,p) = -\int_X h(x)(\pi_0(1+z_0(x))\log(1+z_0(x)) + \pi_1(1+z_1(x))\log(1+z_1(x)))dx.$$

Then, in the hypothesis of no activation, one can assume that $h \simeq h_0 \simeq h_1$ which implies for $i \in \{0,1\}$, $x \in X$ $z_i(x) \ll 1$. Keeping the second order terms yields:

$$MI(y,p) = \int_X h(x)(\pi_0(1+z_0(x))(z_0(x) - \frac{1}{2}z_0^2(x)) + \pi_1(1+z_1(x))(z_1(x) - \frac{1}{2}z_1^2(x)))dx + o(z_0^2, z_1^2)$$

And finally

$$MI(y,p) \simeq \int_X \frac{1}{2}h(x)(\pi_0 z_0^2(x) + \pi_1 z_1^2(x))dx$$

Replacing z_i by its definition, this can be rewritten as :

$$MI(y,p) \simeq \int_X \frac{1}{2} h(x) \frac{(\pi_0 \pi_1^2 + \pi_1 \pi_0^2)(h_0 - h_1)^2(x)}{h^2(x)} dx = \frac{1}{2} \pi_0 \pi_1 \int_X \frac{(h_0 - h_1)^2(x)}{h(x)} dx \qquad (10)$$

A.2 Estimating the value of Mutual Information

Now, one needs to evaluate the quantities in the above equation. Let us consider that N samples of data $(x_1, .., x_N)$ were collected, N_0 of which under the condition$(p = 0)$ and N_1 under the condition $(p = 1)$, so that $\pi_0 = N_0/N$ and $\pi_1 = 1 - \pi_0 = N_1/N$. Then h, h_0 and h_1 are evaluated through Parzen estimates. That is

$$h(x) = (\frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}) \star \mathcal{N}(0, \beta^2)(x)$$

where $\mathcal{N}(0, \beta^2)$ is the normal distribution with 0 mean and variance β^2 ; β is a positive parameter that defines the Parzen window ; each x_i is distributed according to the "real" distribution $h(x)$. To avoid confusions, let us denote by $H(x)$ the "ground truth" distribution, and $h(x)$ its Parzen estimate. Of course, similar formulas hold for h_0 and h_1 (but the "ground truth" distribution is the same, that is H).

Thus from a probabilistic point of view, one has

$$E(h(x)) = H \star \mathcal{N}(0, \beta^2)(x)$$

and

$$var(h(x)) = \frac{1}{N} \left(H \star (\mathcal{N}(0, \beta^2))^2 - E(h(x))^2 \right)$$

Now, in equation (10), we may replace $h(x)$ by $E(h(x))$ and $(h_0 - h_1)^2(x)$ by

$$var(h_0 - h_1)(x) = var(h_0) + var(h_1) = \left(\frac{1}{N_0} + \frac{1}{N_1} \right) \left(H \star (\mathcal{N}(0, \beta^2))^2(x) - E(h(x))^2 \right)$$

To have further estimates, let us suppose that $H = \mathcal{N}(\mu, \sigma^2)$. The problem is unchanged if we take $\mu = 0$, so $H = \mathcal{N}(0, \sigma^2)$.

$$\frac{(h_0 - h_1)^2(x)}{h(x)} \simeq \frac{var(h)(x)}{E(h)(x)} \left(\frac{1}{N_0} + \frac{1}{N_1} \right) = (\frac{\frac{1}{\sqrt{4\pi\beta^2}} \mathcal{N}(0, \sigma^2 + \frac{\beta^2}{2})}{\mathcal{N}(0, \beta^2 + \sigma^2)(x)} - \mathcal{N}(0, \beta^2 + \sigma^2)(x)) \left(\frac{1}{N_0} + \frac{1}{N_1} \right)$$

let us simplify one more step, supposing that $\beta \ll \sigma$:

$$\frac{(h_0 - h_1)^2(x)}{h(x)} \simeq \left(\frac{1}{\sqrt{4\pi\beta^2}} - \mathcal{N}(0, \sigma^2)(x) \right) \left(\frac{1}{N_0} + \frac{1}{N_1} \right)$$

This yields

$$MI(y,p) \simeq \frac{1}{2} \pi_0 \pi_1 \left(\frac{1}{N_0} + \frac{1}{N_1} \right) \int_X (\frac{1}{\sqrt{4\pi\beta^2}} - h(x)) \, dx$$

$$= \frac{1}{2}\pi_0\pi_1 \left(\frac{1}{N_0} + \frac{1}{N_1})\right) \left(\frac{|X|}{\sqrt{4\pi\beta^2}} - 1\right)$$

since $\int_X h(x)dx = 1$

$$= \frac{1}{2N} \left(\frac{|X|}{\sqrt{4\pi\beta^2}} - 1\right)$$

since $\pi_0\pi_1(\frac{1}{N_0} + \frac{1}{N_1}) = \frac{1}{N}$.

Note that H plays no role in the above formula apart from the size of its support $|X|$. It is thus possible to generalise it to more general distributions (but this still has to be checked).

The question is to estimate X. Let us write $X = [-x_0, x_0]$ We propose the following, in the gaussian case :

$$X = [-x_0, x_0] / \ if \ (X_1, .., X_N) \rightsquigarrow \mathcal{N}(0, \sigma^2), \ P(\exists i, |X_i| \geq x_0) \leq \frac{1}{2}$$

Let π be the probability that $|X_i| \geq x_0$. In the gaussian case, $x_0 = \sigma.\sqrt{2}.erf^{-1}(1 - \pi)$. Now,

$$P(\forall i \in 1..N, \ |X_i| \leq x_0) = (1 - \pi)^N (= \frac{1}{2})$$

Thus $\pi \simeq \frac{\log 2}{N}$, and $x_0 = \sigma.\sqrt{2}.erf^{-1}(1 - \frac{\log 2}{N})$

This yields

$$MI(y, p) \simeq \frac{1}{2N} \left(\frac{2.\sigma.\sqrt{2}.erf^{-1}(1 - \frac{\log 2}{N})}{\sqrt{4\pi\beta^2}} - 1\right) \tag{11}$$

The formula (11) obtained in the gaussian case has been checked and validated experimentally; in particular, the dependence on the parameters is correct in a first approximation (see figure 19) : MI is affine with $\frac{\sigma}{\beta}$; unexpectedly, it does not depend on π_0 or π_1.

A.3 Estimating the distribution of MI

Now not only the expected value for MI but also its distribution has to be estimated.
First, let us suppose that **one** sample of the distribution H is observed, and compute h through the Parzen technique. $h(x)$ is a random variable with two first moments $e = E(h)(x) = H \star \mathcal{N}(0, \beta^2)$, and

$$v = var(h)(x) = \left(H \star \mathcal{N}(0, \beta^2)^2 - E(h)^2\right)(x)$$

v and e both depend on x, but to simplify matters, one will consider in the following that $\frac{v}{e}$ $(= \frac{1}{\sqrt{4\pi\beta^2}} - e \sim \frac{1}{\sqrt{4\pi\beta^2}}$ in the gaussian case) only depends on β.

Let us suppose now that **N independently drawn** observations are used to estimate h, N being large. Applying the great numbers law yields $h(x) \rightsquigarrow \mathcal{N}(e, \frac{v}{N})$
In our case, N_0 samples are drawn under hypothesis $(p = 0)$ and N_1 are drawn under hypothesis $(p = 1)$. The same algorithm yields : $h_0(x) \rightsquigarrow \mathcal{N}(e, \frac{v}{N_0})$ and $h_1(x) \rightsquigarrow \mathcal{N}(e, \frac{v}{N_1})$.
We deduce thus that $(h_0 - h_1)(x) \rightsquigarrow \mathcal{N}\left(0, v(\frac{1}{N_1} + \frac{1}{N_0})\right)$ and thus $(h_0 - h_1)^2(x) \rightsquigarrow (v(\frac{1}{N_1} + \frac{1}{N_0}))\chi_1^2$.

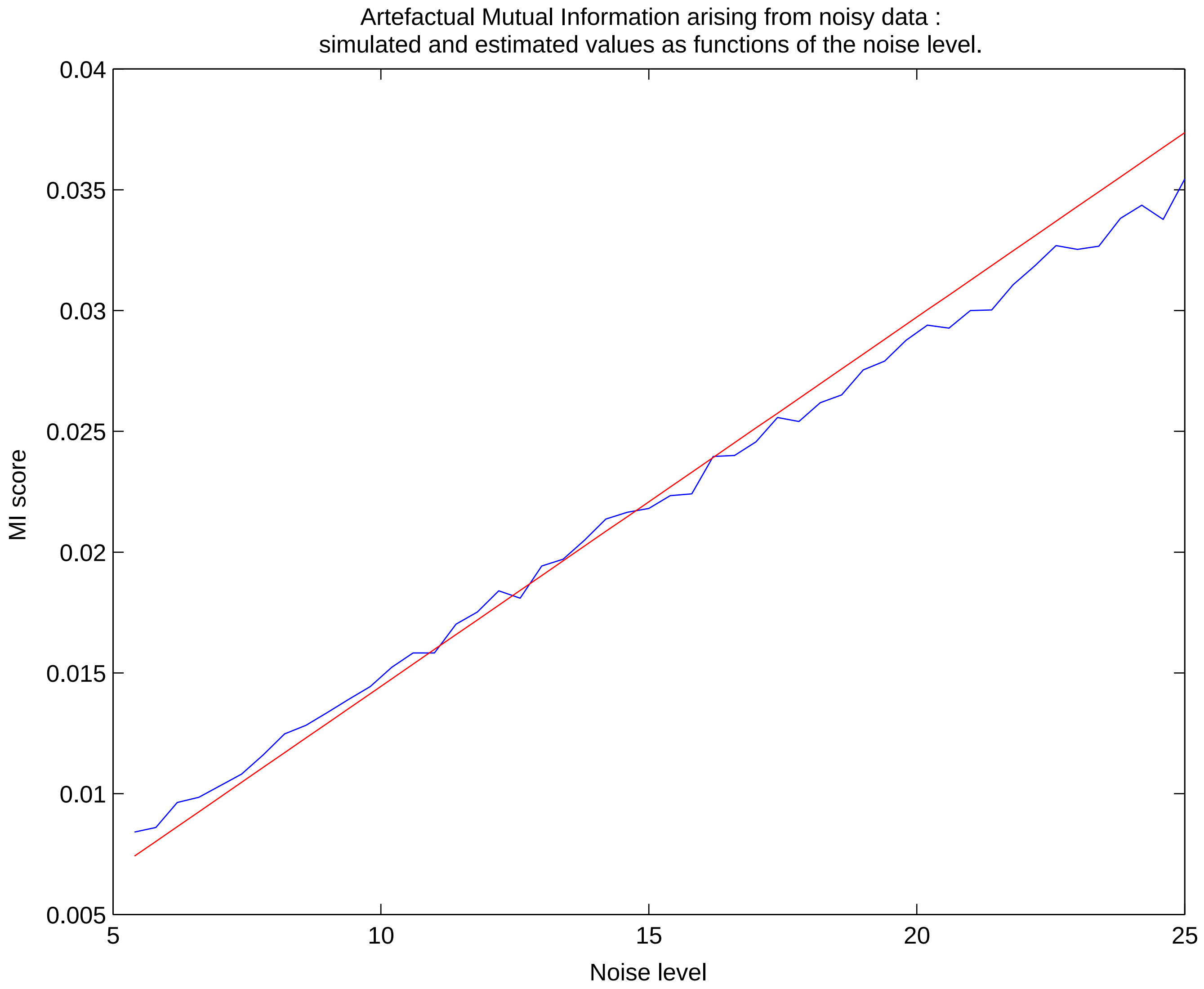

Figure 19: Artefactual Mutual information arising when the input data are random variables with gaussian distribution. The standard deviation of these distributions (σ) is on the x axis. The blue curve represents the MI score averaged from many simulations, the red curve represents the estimate computed through equation (11). The discretisation step is equal to 1, $\beta = 2$, $N = 600$, $\pi_0 = \pi_1 = 0.5$.

Last, $h(x) = (\pi_0 h_0 + \pi_1 h_1) \rightsquigarrow \mathcal{N}(e, \frac{v}{N_0+N_1})$.

Thus, as soon as $N_0 + N_1$ is big enough $h(x)$ can be viewed as a constant in the expression $\frac{(h_0-h_1)^2(x)}{h(x)}$, and, in fact,

$$\frac{(h_0 - h_1)^2(x)}{h(x)} \rightsquigarrow \left(\frac{v}{e}(\frac{1}{N_1} + \frac{1}{N_0})\right) \chi_1^2 \Leftrightarrow \frac{(h_0 - h_1)(x)}{\sqrt{h(x)}} \rightsquigarrow \alpha\mathcal{N}(0,1) \tag{12}$$

where $\alpha = \sqrt{\frac{v}{e}(\frac{1}{N_1} + \frac{1}{N_0})}$

Let $\tilde{h}(x) = \frac{1}{\alpha}\frac{(h_0-h_1)(x)}{\sqrt{h(x)}} \rightsquigarrow \mathcal{N}(0,1)$. $\tilde{h} = (\tilde{h}(x))_{x\in X}$ is a vector whose coordinates are a family of correlated normal variables.

Let us rewrite equation (10) :

$$MI(y,p) \simeq \frac{1}{2}\pi_0\pi_1 \int_X \frac{(h_0 - h_1)^2(x)}{h(x)} dx$$

in the discretised form

$$MI(y,p) \simeq \frac{1}{2}\pi_0\pi_1 \sum_{x\in X} \left(\frac{(h_0 - h_1)(x)}{\sqrt{h(x)}}\right)^2$$

that is

$$MI(y,p) \simeq \frac{1}{2}\pi_0\pi_1\alpha^2\tilde{h}^T\tilde{h} \tag{13}$$

Let us derive the statistical distribution of $\tilde{h}^T\tilde{h}$. If there was no convolution with $\mathcal{N}(0,\beta^2)$ during the estimation of the distribution, then the coordinates of $\tilde{h}$ would be independent, and $\tilde{h}^T\tilde{h}$ would be a χ^2 with $|X|$ degrees of freedom. In the presence of the correlating kernel K, $\tilde{h}^T\tilde{h}$ is no longer a χ^2 variable, but can be approximated by a χ^2 with a reduced number of degrees of freedom, which is $d = \frac{trace(K^TK)^2}{trace((K^TK)^2))}$ When K is the gaussian convolution kernel of size $|X|$ and parameter β, $d = \frac{|X|}{\sqrt{2\pi\beta^2}}$; that is

$$\frac{\tilde{h}^T\tilde{h}}{\sqrt{2\pi\beta^2}} \rightsquigarrow \chi_d^2 \; with \; d = \frac{|X|}{\sqrt{2\pi\beta^2}}$$

Equation (13) becomes

$$MI(y,p) \simeq \frac{1}{2}\pi_0\pi_1\alpha^2\sqrt{2\pi\beta^2}\chi_d^2$$

where

$$\frac{1}{2}\pi_0\pi_1\alpha^2\sqrt{2\pi\beta^2} = \frac{1}{2}\pi_0\pi_1\frac{v}{e}(\frac{1}{N_1} + \frac{1}{N_0})\sqrt{2\pi\beta^2} = \frac{1}{2\sqrt{2}N}$$

The simplifications come from the facts that $N_0 = N\pi_0$, $N_1 = N\pi_1$ and $\frac{v}{e} = \frac{1}{\sqrt{4\pi\beta^2}}$

Finally, we obtain the following distribution for artefactual Mutual information under a binary experimental paradigm :

$$MI(y,p) \rightsquigarrow \frac{1}{2\sqrt{2}N}\chi^2_d \tag{14}$$

with

$$|X| = 2.\sigma.\sqrt{2}.erf^{-1}(1-\frac{\log 2}{N})$$

and

$$d = \frac{|X|}{\sqrt{2\pi\beta^2}}$$

Experiments confirm this as soon as $1 \ll \beta \ll \sigma$ (see figure 20).

A.4 Case of a non-binary paradigm

Here we suppose that the paradigm does not reduce to the succession of two conditions. Let $i \in [0,..,I], I \geq 2$ be these conditions. They can be associated to numbers of occurrences N_i and probabilities π_i.

What happens to the estimations of MI and its distribution ?

One just has to remember that these estimations are based on the integration of $\frac{var(h)}{E(h)}$, h being the empirical estimate of the underlying signal distribution. Thus, if $var(h)$ is a χ^2_1 in the binary case, its is distributed as χ^2_{I-1} in the general case.

Straightforwardly equations (11) and (14) become

$$MI(y,p) \simeq \frac{I-1}{2N}\left(\frac{2.\sigma.\sqrt{2}.erf^{-1}(1-\frac{\log 2}{N})}{\sqrt{4\pi\beta^2}} - 1\right) \tag{15}$$

$$MI(y,p) \rightsquigarrow \frac{1}{2\sqrt{2}N}\chi^2_{(I-1)d} \tag{16}$$

where

$$|X| = 2.\sigma.\sqrt{2}.erf^{-1}(1-\frac{\log 2}{N})$$

and

$$d = \frac{|X|}{\sqrt{2\pi\beta^2}}$$

The correctness of these formulas has also been checked through numerical simulations. They hold as well as the formulas obtained in the binary case.

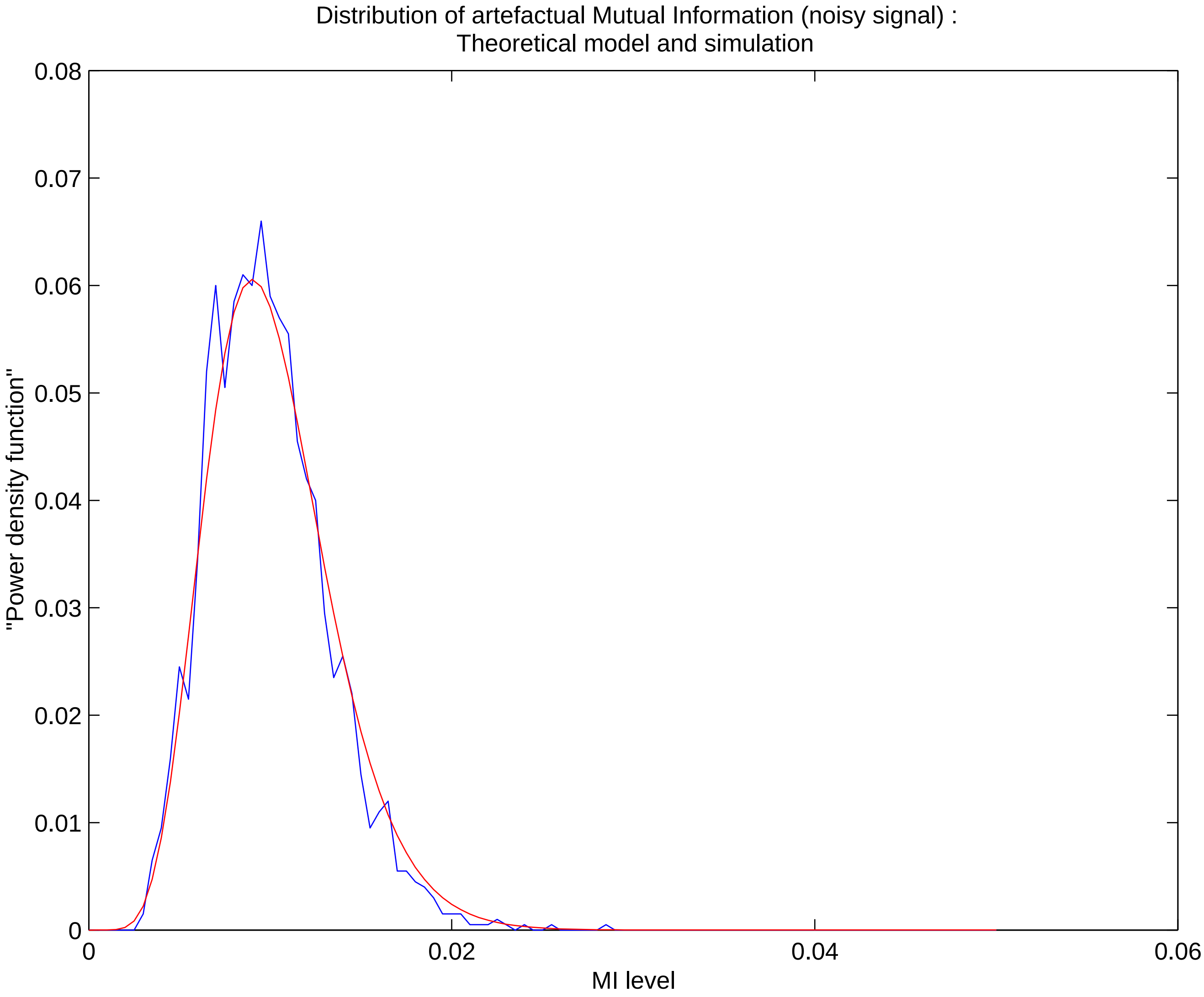

Figure 20: Distribution of artefactual MI values when the input signal is a set of random variables drawn from a gaussian distribution. The red curve is the estimate from equation (14), the blue one is the the result of a simulation with 2000 repetitions. The parameters have been set as : $\beta = 3$, $N = 600$, $\pi_0 = \pi_1 = 0.5$, $\sigma = 20$.

A.5 Mutual Information in the 2-D case

In section (3.1.2), we proposed to estimated Mutual Information scores on a 2-D histogram : The signal is no longer x_t but (x_{t-1}, x_t). In the hypothesis of a null distribution, this has little influence on the results displayed in equations (11) and (14). Let us suppose that (x_{t-1}, x_t) are uncorrelated ; then the only difference is that the support $|X|$ becomes a square $|X|^2$, as well as the smoothing term $\sqrt{2\pi\beta^2} \to 2\pi\beta^2$.
Therefore, equations (11) and (14) become :

$$MI(y,p) \simeq \frac{1}{2N}\left((\frac{2.\sigma.\sqrt{2}.erf^{-1}(1-\frac{\log 2}{N})}{\sqrt{4\pi\beta^2}})^2 - 1 \right) \tag{17}$$

$$MI(y,p) \rightsquigarrow \frac{1}{2N}\frac{v}{e}(2\pi\beta^2)\chi^2_d \tag{18}$$

where

$$|X| = 2.\sigma.\sqrt{2}.erf^{-1}(1-\frac{\log 2}{N})$$

and

$$d = (\frac{|X|}{\sqrt{2\pi\beta^2}})^2$$

Our simulations show that MI is overestimated with this method, and the discrepancy increases with σ ; this shows that the hypothesis of small fluctuations in the estimated density is violated, and the concavity of entropy functions lowers the value. We propose to view our estimated as an upper bound on artefactual mutual information.

Practically, the presence of correlation in the signal does not significantly change the estimate nor the distribution of MI.

B ANOVA models from a statistical point of view

B.1 Model *ANOVA with memory* : no temporal correlation

The goal of the section is to provide a meaningful threshold on ANOVA tests values obtained empirically between the time series and the experimental paradigm. As in most approaches, our goal is to avoid false-positive, that is to reduce the probability of detection in the case of null hypothesis.

Let $y(t)$ be a temporal signal, and $p(t)$ be the corresponding paradigm, with $t = 1..N$. Let us denote p_k the paradigm with a memory of length $k-1$ (in scans) : $p_k(t) = (p(t), p(t-1), .., p(t-k))$. From [RMPA98] or [RMA99] one has :

$$var(y) = var[E(y|p_k)] + E_{p_k}[var(y|p_k)] \tag{19}$$

Under the null hypothesis, y follows a given distribution. For simplicity, one chooses $y(t) \rightsquigarrow N(0,1)$, whose main disadvantage is to be temporally decorrelated (whereas normality is, from our experiments, a correct model). The question of the temporal autocorrelation of the signal is raised later (section B.2).

Let us consider a block paradigm with two states 0 and 1:

$$p(t) = \underbrace{1..1}_{m_1}\underbrace{0..0}_{m_0}\underbrace{1..1}_{m_1}\cdots$$

then with the notation $1^l0^n = \underbrace{1..1}_{l}\underbrace{0..0}_{n}$

$$p_k(t) = (10^{k-1})(1^20^{k-2})..(1^{k-1}0)\underbrace{1^k..1^k}_{m_1-k}(01^{k-1})..(0^{k-1}1)\underbrace{0^k..0^k}_{m_0-k}(10^{k-1})..(1^{k-1}0)..$$

The cardinal of p_k is $2k$. One denotes by N_{p_k} the number of samples associated to the condition p_k of the paradigm. To simplify the notations, the sum symbols on the following lines are taken implicitly on the different states of the paradigm.

Caveat : In the next paragraphs, if $(x_i)_{i=1..N}$ is a random variable, $var(x)$ will be used for : $var(x) = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N}$, where $\bar{x} = \frac{\sum_{i=1}^{N}x_i}{N}$. Usually, var is normalised by $N-1$.

The next three relations follow from definition :

$$N.var(y) \rightsquigarrow \chi^2_{N-1},$$
$$E(y|p_k) \rightsquigarrow \mathcal{N}(0, 1/N_{pk}),$$
$$N_{p_k}.var(y|p_k) \rightsquigarrow \chi^2_{N_{p_k}-1}.$$

On one hand,

$$E_{p_k}[var(y|p_k)] = \sum \frac{N_{pk}}{N} var(y|p_k),$$

which yields

$$N.E_{p_k}[var(y|p_k)] \rightsquigarrow \chi^2_{\sum(N_{p_k}-1)} = \chi^2_{N-2k}.$$

On the other hand, supposing to make matters easier that $E_{p_k}(E(y|p_k)) = 0$, one has

$$var[E(y|p_k)] = \sum \frac{N_{p_k}}{N} E(y|p_k)^2.$$

Thus, $\sqrt{N_{p_k}}E(y|p_k) \rightsquigarrow N(0,1)$ implies

$$N.var[E(y|p_k)] \rightsquigarrow \chi^2_{2k-1}$$

Now, in the gaussian framework, the two terms of the right member of equation (19) are independent. According to the preceding computation, (19) can be rewritten symbolically -and trivially- as

$$\chi^2_{N-1} = \chi^2_{2k-1} + \chi^2_{N-2k} \qquad (20)$$

Finally, one concludes that

$$\eta = \frac{var[E(y|p_k)]}{E_{p_k}[var(y|p_k)]} = \frac{CR}{1-CR} \rightsquigarrow F_{2k-1,N-2k}\frac{2k-1}{N-2k}$$

Which is the usual assessment for ANOVA tests.

B.2 Model Anova *with memory* : Adding temporal correlation

The gaussianity of the signal is not a problem. One intuitive reason is that departure from gaussianity are the same on the numerator and denominator of η. In fact, the real problem is the temporal autocorrelation of fMRI data : It is noticeable that the data has such correlation, even if there is no signal. Besides, it is well-known [ZAD97] that such correlation biases the observation under the null hypothesis; this is clearly shown in an experiment on real data on figure (21). Experimentally, the autocorrelation coefficients of fMRI noise seem to be well-described as:

$$R(0) = \sigma^2,\ R(1) = \nu R(0),\ R(n > 1) = \rho^{n-1} R(1) \tag{21}$$

with $1 > \rho > \nu > 0$. If there is some *activation signal* mistakenly present in the so-called noise, the autocorrelation is even stronger, since the paradigm $p(t)$ is highly autocorrelated. This is a reason why the relation $p(t) \to y(t)$ should be kept as unconstrained as possible.

The way of treating the problem is to evaluate the effective degrees of freedom of the F-function devised above for statistical description of noise. But let us first notice that it is unlikely that η will be distributed as a F-function, because the numerator and denominator are correlated, thus not independent -in particular, equation (20) no longer holds. In a first approach, this problem will be neglected. Instead, one will try to correct it by changing the degrees of freedom value, using the concept of *effective degrees of freedom* (*edf*).

Derivation of the *edf* for $var(y)$. First, let us assume that (y) is given by an AR-1 process $(0 < \rho < 1)$

$$y_n - \rho y_{n-1} = z_n,\ with\, z_n \rightsquigarrow N(0,1)$$

This corresponds to $R(n) = \frac{\rho^n}{(1-\rho)^2}$, and the solution is $y_n = \sum_{i=0}^{n-1} \rho^i z_{n-i}$; thus $var(y) = \frac{1}{1-\rho^2}$.

Let us study $V = \frac{1}{N}\sum_{n=1}^{N} y_n^2$. Let $Y = (y_1, ..., y_n)$, $Z = (z_1, ..., z_n)$ and P be the matrix

$$P = \begin{pmatrix} 1 & 0 & .. & 0 \\ \rho & 1 & 0 & .. \\ .. & \rho & 1 & 0 \\ \rho^{n-1} & .. & \rho & 1 \end{pmatrix}$$

Then $Y = PZ$, and $V = \frac{1}{N}Y^T Y = \frac{1}{N}Z^T P^T P Z$. Let $S = P^T P$. S is symmetric, hence $S = U^T \Lambda U$, where $\Lambda = diag(\lambda_1, .., \lambda_n)$ is diagonal and U is orthogonal. $V = \frac{1}{N}\sum_{i=1}^{N} \lambda_i (UZ)_i^2$. Can V be approximated a χ^2 distribution ? Rigorously, V is a χ^2 if and only if $\lambda_i \in \{0,1\}\, \forall i$. We propose the following approximation, which is correct up to the first two moments of the distribution:

The moment generating function of a χ^2 with q degrees of freedom is $\phi_r(t) = (1-2t)^{-\frac{q}{2}}$ (see[Seb77] for details) ; the moment generating function of a sum of weighted squared gaussians with weights $(\nu_1..\nu_n)$ is $\psi_{\nu_1..\nu_n}(t) = \Pi_{i=1}^{n}(1-2t\nu_i)^{-\frac{1}{2}}$. Let us choose q in order to make the first two moments equal. Then *edf* will be taken equal to q. The equality of the first moments yields

$$q = \sum_{i=1}^{n} \nu_i,$$

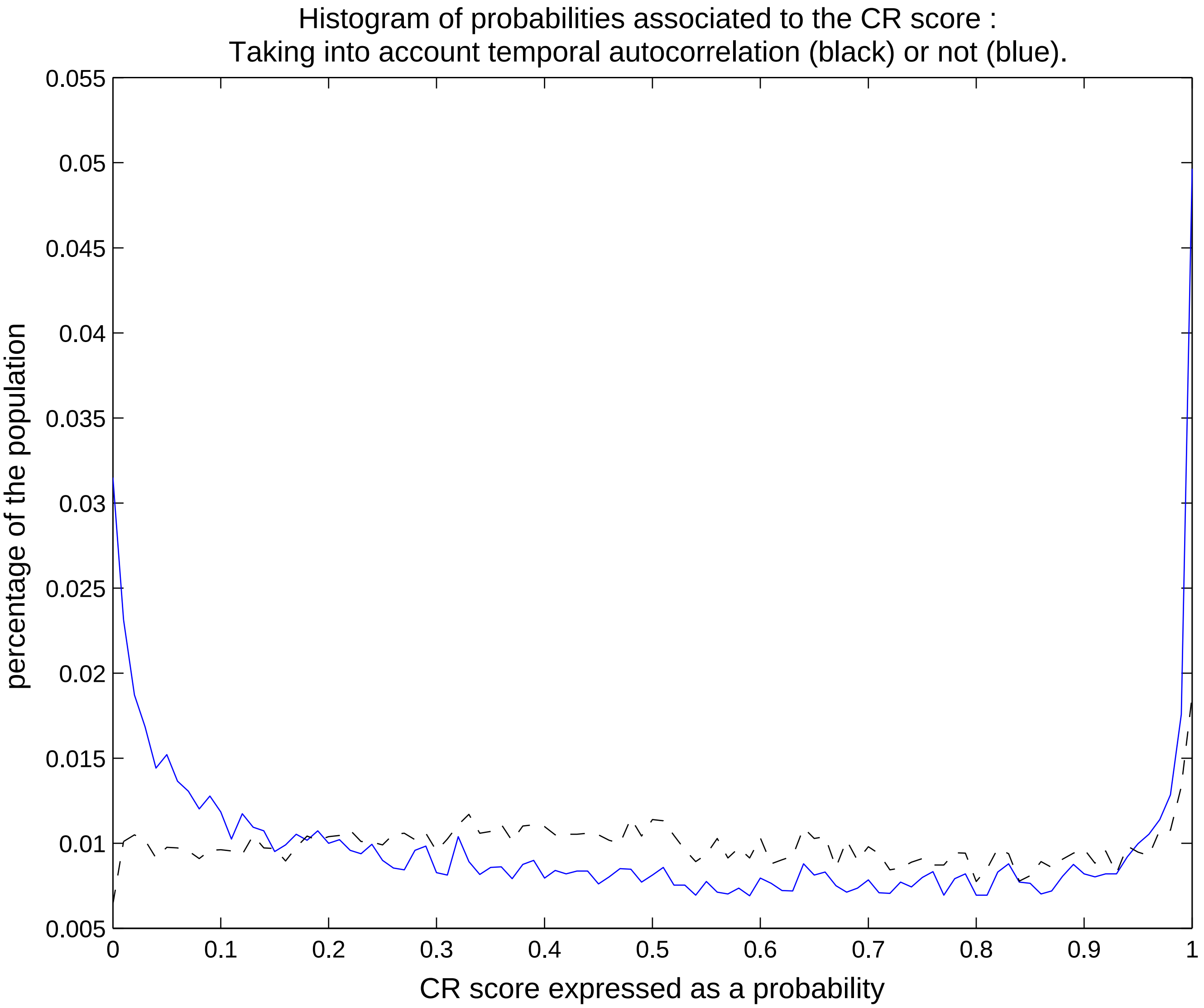

Figure 21: We illustrate the importance of taking into account temporal autocorrelation. A histogram of the CR scores obtained on real data is plotted above. The scores have been warped into probabilities by using a F-function. The function in blue does not take temporal autocorrelation into consideration, whereas the black curve does. Apart from a few percents of activated voxels, the remainder can be considered as non-activated ; thus the distribution of probabilities should be flat, since the CR scores are modelled as the realization of a random process. We notice that the black distribution is indeed a lot flatter than the other one : we interpret it as a much more realistic modelling of the data.

and that of the second moments yields

$$q(q-1) = (\sum_{i=1}^{n} \nu_i)^2 - \sum_{i=1}^{n} \nu_i^2.$$

Taking ν_i proportional to λ_i yields $\nu_i = \frac{\lambda_i}{\nu}$ with $\nu = \frac{\sum_{i=1}^n \lambda_i^2}{\sum_{i=1}^n \lambda_i}$. Finally, the solution is :

$$q = \frac{(\sum_{i=1}^n \lambda_i)^2}{\sum_{i=1}^n \lambda_i^2}$$

Our next task is to evaluate $\sum_{i=1}^n \lambda_i$ and $\sum_{i=1}^n \lambda_i^2$.

First, we can approximate $S = P^T P$ with

$$S \simeq \frac{1}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & .. & \rho^{n-1} \\ \rho & 1 & \rho & .. & \rho^{n-2} \\ .. & \rho & 1 & \rho & .. \\ & & .. & & \\ \rho^{n-1} & \rho^{n-2} & .. & \rho & 1 \end{pmatrix}$$

This yields :

$$\sum_{i=1}^n \lambda_i = Tr(S) \simeq \frac{N}{(1-\rho^2)},$$

and

$$\sum_{i=1}^n \lambda_i^2 = Tr(S^2) = \sum_{i,j=1}^n S_{i,j}^2 \simeq \left(N - \frac{1}{(1-\rho^2)}\right) \frac{1+\rho^2}{(1-\rho^2)^3}$$

.

The approximations above are correct if ρ is far from 1. For N=100 and $\rho = 0.9$ the error is 7%.

We obtain an estimate for *edf* :

$$edf = q \simeq \frac{\left(\frac{N}{(1-\rho^2)^2}\right)^2}{\left(N - \frac{1}{1-\rho^2}\right)\frac{1+\rho^2}{(1-\rho^2)^3}} \simeq N\frac{1-\rho^2}{1+\rho^2}$$

Before generalising the computation, let us notice that the above approximations are equivalent to the following :

$$S \simeq \begin{pmatrix} R(0) & R(1) & R(2) & .. & R(n-1) \\ R(1) & R(0) & R(1) & .. & R(n-2) \\ .. & R(1) & R(0) & R(1) & .. \\ & & .. & & \\ R(n-1) & R(n-2) & .. & R(1) & R(0) \end{pmatrix}$$

$$E(V) \simeq R(0)$$

$$var(V) \simeq \frac{2}{N^2}\left(N.R(0)^2 + \sum_{n=1}^{N-2} R(n)^2(2N-2n)\right)$$

which gives

$$edf = \frac{N}{1+\sum_{n=1}^{N-2}\frac{R(n)^2}{R(0)^2}\frac{2N-2n}{N}} \tag{22}$$

Now, a more precise model is AR-1 MA-1 ($|\rho|, |\lambda|, |\mu| < 1$) :

$$y_n - \rho y_{n-1} = \lambda z_n + \mu z_{n-1},\ with\, z_n \rightsquigarrow N(0,1)$$

The corresponding autocorrelation coefficients are :

$$R(0) = \frac{\lambda^2+\mu^2+2\rho\lambda\mu}{1-\rho^2},\ R(1) = \rho R(0) + \mu\lambda,\ R(k>1) = R(1)\rho^{k-1}$$

Which is, up to a scale factor, the model (21) that we experimentally verified for fMRI data.

The formula for y_n is : $y_n = \sum_{i=0}^{n-2}\rho^i(\lambda z_{n-i} + \mu z_{n-i-1})$; applying (22) yields

$$edf = \frac{N}{1+(1+\frac{\mu\lambda}{R(0)\rho})^2\frac{2\rho^2}{1-\rho^2}} = \tau N$$

The straightforward interpretation of this computation is that the effective degrees of freedom of $var(y)$ are reduced by this τ factor.

Now, let us compute the effective degrees of freedom for $var[E(y|p_k)]$. One can argue that the variables $E(y|p_k)$ are also consecutive in times, and correlated as the samples y_k so that the effective degrees of freedom are obtained through multiplication by the same factor τ. More precisely, among the *states* p_k, the states 1^k and 0^k are repeated $m_1 - k$ and $m_0 - k$ times respectively, and thus the quantities $E(y|1^k)$ and $E(y|0^k)$ are more or less independent from the others. The effective degrees of freedom of $var[E(y|p_k)]$ are thus $2 + (2k-2)\tau$ (a numerical simulation confirmed the result).

The effective degrees of freedom for $E_{p_k}[var(y|p_k)]$ are $\tau(N-2k)$, since this quantity behaves essentially as $var(y)$ (using a more exact number would have almost no impact on the final distribution).

The statistics of interest for $\eta = \frac{CR}{1-CR}$ are now $F_{c',N'}$ with $c' = 2 + (2k-2)\tau$ and $N' = \tau(N-2k)$. A simulation is presented in figure 22.

B.3 Summary : How to perform a threshold on ANOVA test values ?

The method for thresholding the CR values is presented in figure 23.

C Statistical tests after pre-processing through the markovian model

In section (3.1.3) we proposed to keep the usual tests -ANOVA and MI- but to pre-process the data through a Markov Chain model in order to simulate the asymptotical behaviour of the signal under fixed conditions.

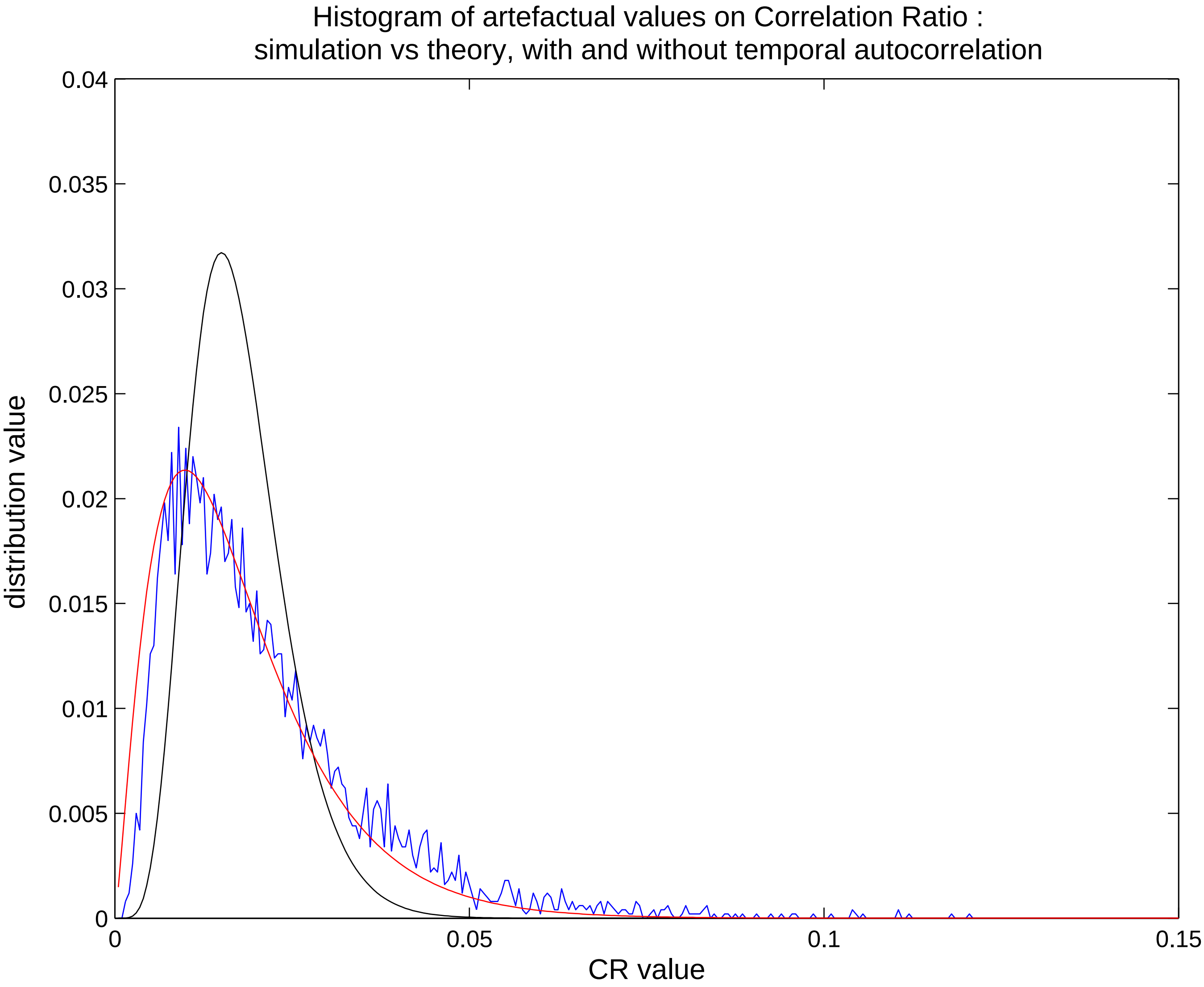

Figure 22: Experiment on residual or artefactual Correlation Ratio in the computed from temporally autocorrelated noise. The approximated theoretical model is plotted in red and values of a simulated model is plotted in blue. The noise is AR-1, with ρ=0.8, k=7, N=810. A model that does not take temporal correlation into account is plotted in black. It fits the data less than the "correlated" model.

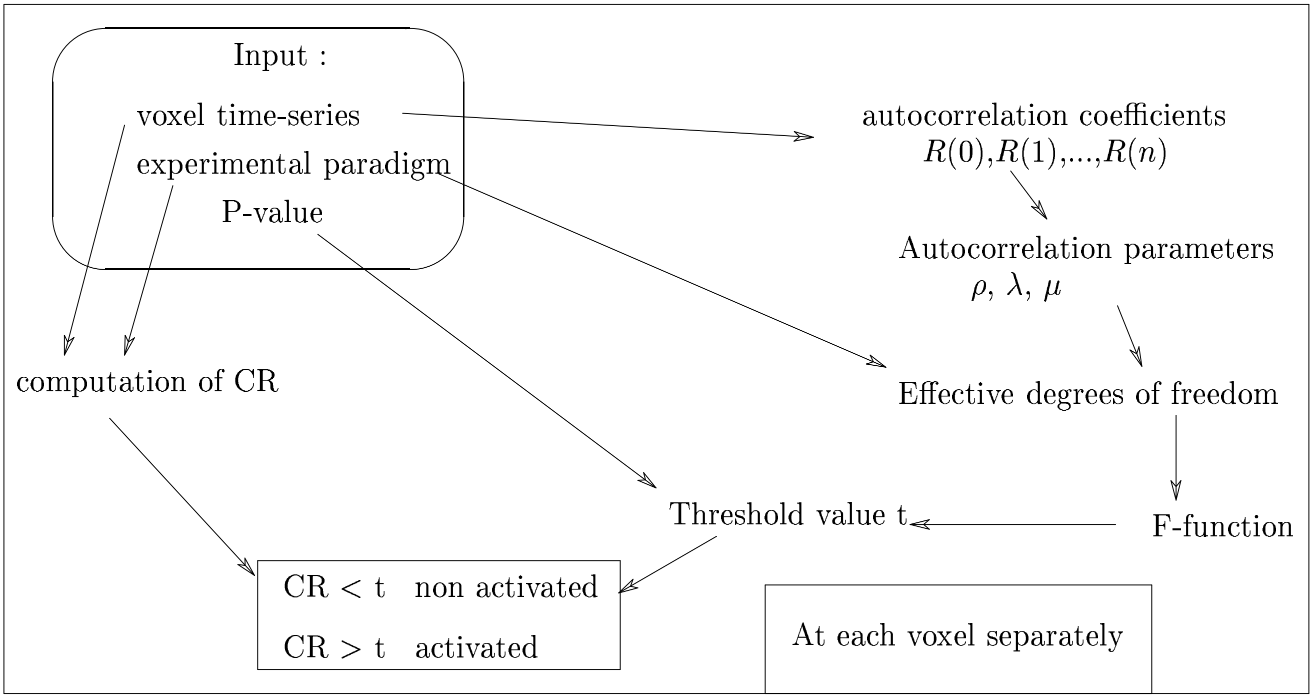

Figure 23: This figure displays the different steps in order to threshold the CR values, taking into account the temporal autocorrelation in the signal.

This kind of processing is highly non-linear, and its result can hardly be viewed as a kind of filtering. Is it still possible to fit a statistical test to assess these methods ?

Qualitatively, there are two possible ways :

- If the input data is some uncorrelated white noise, the *markovian* pre-processing has little or no effect, so that the parameters of the estimate can be kept unchanged.
- On the other hand, if the input data is strongly correlated, then the preprocessing behaves like a mapping : $x(t) \to x(T)$ $\forall t \in \{1,..,T\}$ where T is the epoch length. Thus an epoch reduces to a single degree of freedom in the statistics.

We think that this alternative can be overcome by choosing the *worst case*, hence reducing each epoch to a single degree of freedom or equivalently a single measure in the parameters of statistical tests. The significance threshold is certainly over-estimated, but this is a security against false positive.

Practically, the parameter N in equation (14), and in the test function $F_{k-1,N-k}$ has to be replaced by R, the number of repetitions of the experimental condition.

A consequence of this approach is that, on a statistical point of view, this kind of method is well suited for paradigms with short epochs and many repetitions.

D On ARMA models

The goal of this paragraph is simply to make explicit the relationship between the autocorrelation coefficients and the ARMA coefficients of a regressive model.
Let us consider, with $(|\rho|, |\lambda|, |\mu| < 1)$:

$$y_n - \rho y_{n-1} = \lambda z_n + \mu z_{n-1},\ with\ z_n \rightsquigarrow N(0,1)$$

then it is straightforward that

$$R(0) = \frac{\lambda^2 + \mu^2 + 2\rho\lambda\mu}{1-\rho^2},\ R(1) = \rho R(0) + \mu\lambda,\ R(k > 1) = R(1)\rho^{k-1}$$

In practice, one computes the coefficients $R(n)$ empirically.
Then, Let $\sigma^2 = R(0)$, $R(1) = \nu\rho\sigma^2$, and $R(n > 1) = \nu\rho^n\sigma^2$.
The parameters ρ and ν are easily computed from linear regression on the $(\log R(n))$ sequence as in [DB97].

There remains only to solve the system

$$\frac{\lambda^2 + \mu^2 + 2\rho\lambda\mu}{1-\rho^2} = \sigma^2$$
$$\rho\sigma^2 + \mu\lambda = \nu\rho\sigma^2$$

to compute the parameters λ and μ of the model. These parameters could be used to whiten the data, but this would spoil the signal contained in the data.

References

[AKKK99] B.A. Ardekani, J. Kershaw, K. Kashikura, and I. Kanno. Activation detection in functional mri using subspace modeling and maximum likelihood estimation. *IEEE Transactions on Medical Imaging*, 18(2):101–114, February 1999.

[AKM$^+$01] Alexandre Andrade, Ferath Kherif, Jean-Francois Mangin, Keith Worsley, Anne-Lise Paradis, Olivier Simon, Stanislas Dehaene, Denis Le Bihan, and Jean-Baptiste Poline. Detection of fmri activation using cortical surface mapping. *Human Brain Mapping*, 12(2):79–93, February 2001.

[AV97] G. Aubert and L. Vese. A variational method in image recovery. *SIAM Journal of Numerical Analysis*, 34(5):1948–1979, October 1997.

[BD00] M.A. Burock and A.M. Dale. Estimation and detection of event-related fmri signals with temporally correlated noise : A statistically efficient and unbiased approach. *Human Brain mapping*, 11:249–260, 2000.

[bL$^+$01] Ed. bullmore, Chris Long, et al. Colored noise and computational inference in neurophysiological (fmri) time series analysis : Resampling methods in time and wavelets domains. *Human Brain mapping*, 12:61–78, 2001.

[BWM98] R. Baumgartner, C. Windischberger, and E. Moser. Quantification in functional magnetic resonance imaging : fuzzy clustering vs correlation analysis. *magnetic resonance Imaging*, 16(2):115–125, 1998.

[DB97] A.M. Dale and R.L. Buckner. Selective averaging of rapidly presented individual trials using fmri. *Human brain mapping*, 5:329–340, 1997.

[DK97] R. N. Davé and R. Krishnapuram. Robust clustering methods: A unified view. *IEEE Transactions on Fuzzy Systems*, 5(2):270–293, 1997.

[DKvC98] Xavier Descombes, Frithjof Kruggel, and D.Y. von Cramon. fmri signal restoration using an edge preserving spatio-temporal markov random field. *Neuroimage*, 8:340–349, 1998.

[EB99] Brian S. Everitt and Edward T. Bullmore. Mixture model mapping of brain activation in functional magnetic resonance images. *Human Brain Mapping*, 7:1–14, 1999.

[FA$^+$97] K.J. Friston, J. Ashburner, et al. *spm 97 course notes*. Wellcome Department of Cognitive Neurology, University College london, 1997.

[GGWF01] G. Gratton, M.R. Goodman-Wood, and M. Fabiani. Comparison of neuronal and hemodynamic measures of the brain response to visual stimulation : An optical imaging study. *Human Brain Mapping*, 13:13–25, 2001.

[GHLR99] Cyril Goutte, Lars Kai Hansen, Mattew G. Liptrot, and Egill Rostrup. Feature space clustering for fmri meta-analysis. Technical Report 13, Technical University of Denmark, 1999. submitted to NeuroImage.

[HJ00] Niels Vaever Hartvig and Jens Ledet Jensen. Spatial mixture modeling of fmri data. *Human Brain mapping*, 11:233–248, 2000.

[KAK99] J. Kershaw, B.A. Ardekani, and I. Kanno. Application of bayesian inference to fmri data analysis. *IEEE transactions on Medical imaging*, 18(12):1138–1153, December 1999.

[KFT$^+$00] J. Kim, J.W. Fisher, A. Tsai, C. Wible, A.S. Willsky, and W.M. Wells. Incorporating spatial priors into an information theoretic approach for fmri data analysis. In G. Goos, J. Hartmanis, and J. van Leeuwen, editors, *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2000*, volume 1935 of *Lectures Notes in Computer Science 1935*, pages 62–72. Springer, October 2000.

[KvC99] Frithjof Kruggel and D.Y. von Cramon. Physiollogically oriented models of the hemodynamic response in functional mri. In A. Kuba et al., editors, *IPMI'99*, volume 1613 of *Lecture Notes in Computer Science*, pages 294–307. Springer-Verlag Berlin Heidelberg, 1999.

[KvCD99] F. Kruggel and von Cramon D.Y. Temporal properties of the hemodynamic response in functional mri. *Human Brain mapping*, 8(4):259–271, 1999.

[LU01] D.N. Levin and S.J. Uftring. Detecting brain activation in fmri data without prior knowledge of mental event timing. *NeuroImage*, 14:153–160, 2001.

[LZ97] Nicholas Lange and Scott L. Zeger. Non-linear fourier time serie analysis for human brain mapping by functional magnetic resonance imaging. *Appl. Statist.*, 46(1):1–29, 1997.

[MJ$^+$98] M.J. McKeown, T.P. Jung, et al. Spatially independant activity patterns in functional mri data during the stroop color-naming task. *Proc. Natl. Acad. Sci. USA*, 95:803–810, February 1998.

[MLL$^+$01] K.L. Miller, W.-M. Luh, T.T. Liu, et al. Nonlinear temporal dynamics of the cerebral blood flow response. *Human Brain Mapping*, 13:1–12, 2001.

[MM$^+$98] Martin J. McKeown, S. Makeig, et al. Analysis of fmri data by blind separation into independant spatial components. *Human Brain Mapping*, 6:160–188, 1998.

[MR00] Jonathan L. Marchini and Brian D. Ripley. A new statistical approach to detecting significant activation in functional mri. *Neuroimage*, 12:366–380, 2000.

[Nev96] Jacques Neveu. *Probabilités*. Ecole Polytechnique, Palaiseau, France, 1996 edition, 96.

[NN99] F. N. Nan and R.D. Nowak. Generalized likelihood ratio detection for fmri using complex data. *IEEE Transactions on Medical Imaging*, 18(4):320–329, April 1999.

[PNPH99a] K.M. Petersson, T.E. Nichols, J.B. Poline, and A.P. Holmes. Statistical limitations in functional neuroimaging. 1. non-inferential methods and statistical models. *Phil. Trans. R. Soc. Lond.*, 354:1239–1260, 1999.

[PNPH99b] K.M. Petersson, T.E. Nichols, J.B. Poline, and A.P. Holmes. Statistical limitations in functional neuroimaging. 2. signal detection and statistical inference. *Phil. Trans. R. Soc. Lond.*, 354:1261–1281, 1999.

[RMA99] A. Roche, G. Malandain, and N. Ayache. Unifying maximum likelihood approaches an medical image registration. Rapport de recherche 3741, INRIA, July 1999.

[RMPA98] Alexis Roche, Grégoire Malandain, Xavier Pennec, and Nicholas Ayache. Multimodal image registration by maximization of the correlation ratio. Technical Report 3378, INRIA, August 1998.

[Seb77] G.A.F. Seber. *Linear regression Analysis*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, 1977.

[SKvC00] M. Svensén, F. Kruggel, and D.Y. von Cramon. Probabilistic modeling of single-trial fmri data. *IEEE Transactions on Medical Imaging*, 19(1):25–35, January 2000.

[TFW$^+$99] A. Tsai, J.W. Fisher, C. Wible, W.M. Wells, J. Kim, and A.S. Willsky. Analysis of functional mri data using mutual information. In *Second International Conference on medical Image Computing and Compuer-Assisted Intervention*, volume 1679 of *Lecture Notes in Computer Science*, pages 473–480. Springer Verlag, September 1999.

[VN98] Alberto L. Vazquez and Douglas C. Noll. Nonlinear aspects of the bold response in functional mri. *Neuroimage*, 7:108–118, 1998.

[ZAD97] E. Zarahn, G.K. Aguirre, and M. D'Esposito. Empirical analyses of bold fmri statistics. 1. spatially unsmoothed data collected under null-hypothesis. *Neuroimage*, 5:179–197, 1997.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399