A Numerical Study of the Kullback-Leibler Distance
in Functional Magnetic Resonance Imaging

Brenno Caetano Troca Cabella, Márcio Júnior Sturzbecher, Walfred Tedeschi,
Oswaldo Baffa Filho, Dráulio Barros de Araújo, and Ubiraci Pereira da Costa Neves
Departamento de Física e Matemática,
Faculdade de Filosofia Ciências e Letras de Ribeirão Preto,
Universidade de São Paulo,
Av. Bandeirantes, 3900,
14.040-901, Ribeirão Preto, SP, Brazil
brenno@pg.ffclrp.usp.br

The Kullback-Leibler distance (or relative entropy) is applied in the analysis of functional magnetic resonance (fMRI) data series. Our study is designed for event-related (ER) experiments, where a brief stimulus is presented and a long period of rest is followed. In particular, this relative entropy is used as a measure of the "distance" between the probability distributions p_1 and p_2 of the signal levels related to stimulus and non-stimulus. In order to avoid undesirable divergences of the Kullback-Leibler distance, a small positive parameter δ is introduced in the definition of the probability functions in such a way that it does *not* bias the comparison between both distributions. Numerical simulations are performed so as to determine the probability densities of the mean Kullback-Leibler distance $\overline{D}$ (throughout the N epochs of the whole experiment). For small values of N ($N < 30$), such probability densities $f(\overline{D})$ are found to be fitted very well by Gamma distributions ($\chi^2 < 0.0009$). The sensitivity and specificity of the method are evaluated by construction of the receiver operating characteristic (ROC) curves for some values of signal-to-noise ratio (SNR). The functional maps corresponding to real data series from an asymptomatic volunteer submitted to an ER motor stimulus is obtained by using the proposed technique. The maps present activation in primary and secondary motor brain areas. Both simulated and real data analyses indicate that the relative entropy can be useful for fMRI analysis in the information measure scenario.

Keywords: fMRI Analysis; Kullback-Leibler Distance; Hypothesis Testing

I. INTRODUCTION

The functional magnetic resonance imaging (fMRI) is one of the main techniques for mapping human brain activity in a noninvasive way [1]. This technique allows to assess cognitive functions with high spatial and temporal resolution imaging. Neural activity due to stimulation causes an increase of the blood flow and oxygenation in the local vasculature. The blood oxygenation level dependent (BOLD) contrast of the image is induced by changes in relative concentrations of oxy and deoxy-hemoglobin [2–4]. The BOLD signal is the response of the application of an experimental protocol (known as paradigm) which is composed of periodic changes between stimulus and non-stimulus of different brain areas. The result of such an experiment is a time series corresponding to the signal for each voxel. In the processing of fMRI time series, several methods that separate the physiologically induced signals (corresponding to active voxels) from noise (non-active voxels) have been employed. Deterministic methods quantify the similarity between the time series of each voxel with a hemodynamic response function (HRF) model. On the other hand, statistical methods can infer how significative is the difference between the signals corresponding to periods of stimulation and non-stimulation but do not need *a priori* knowledge of the form of the HRF. In the late years, methods based on information measures, such as the Shannon and Tsallis entropies and the generalized mutual information, have been employed as alternatives for the conventional analysis of fMRI data [5–9].

These informational methods also present the advantage of not requiring a HRF model.

In this context, we now propose to use the Kullback-Leibler distance (or relative entropy) as an information measure to analyse fMRI data series. Our study is designed for event-related (ER) paradigms, where stimulation is presented in a short period of time and followed by a long period of rest. Like other informational methods, the proposed technique only considers the general structure of the signal and needs no HRF model. Indeed the relative entropy is able to quantify the difference between the probability distributions p_1 and p_2 of the states of the BOLD signal acquired during periods of stimulation and rest, respectively. The sample average of the Kullback-Leibler distance is elected as the statistic for testing the null hypothesis that a voxel is non-active (noise). The probability distributions of the test statistic are numerically determined by Monte Carlo simulations and found to be fitted very well by Gamma distributions. The sensitivity and specificity of the method are computed for different signal-to-noise ratios (SNRs). Finally, the proposed technique is applied to real data series (from an asymptomatic volunteer submitted to an ER motor stimulus) in order to determine the corresponding activation map. Our results from both simulated and real data indicate that the technique can be in fact a very good option for fMRI analysis.

II. RELATIVE ENTROPY

In information theory, the relative entropy is a measure of the "distance" between the probability functions p_1 and p_2 of two discrete random variables X_1 and X_2, respectively. This information measure is also known as the Kullback-Leibler distance (D) and defined as [10]

$$D = \sum_{j=1}^{L} p_{1j} \log_2 \left(\frac{p_{1j}}{p_{2j}} \right) \ , \tag{1}$$

where p_{ij} is the probability that X_i assumes the j-th value from its set of L elements. The logarithm is taken to base 2 so as the entropy is measured in bits.

Relative entropy is not symmetric and is always non-negative. It is zero if and only if $p_1 = p_2$. It should also be noted that $\lim_{\substack{p_{1j} \to 0 \\ p_{2j} \neq 0}} p_{1j} \log_2 \left(\frac{p_{1j}}{p_{2j}} \right) = 0$ whereas

$\lim_{\substack{p_{2j} \to 0 \\ p_{1j} \neq 0}} p_{1j} \log_2 \left(\frac{p_{1j}}{p_{2j}} \right) = \infty$.

The relative entropy of the time course $s(t)$ of a signal (from real or even simulated data) is proposed to be calculated as follows. Consider one single event-related trial (also named *epoch*). During that epoch the time course $s(t)$ of the (simulated) BOLD signal of an active voxel exhibits a peak and then returns to a baseline level (Fig.1). As for the calculation of Shannon entropy in a previous work [7], an entire epoch (window W) of the experiment is divided into two time intervals: half windows W_1 (related to the signal increase) and W_2 (corresponding mainly to baseline values). The probability distribution of signal levels (p_1) corresponding to the BOLD response in the first time period is clearly expected to differ from that one (p_2) corresponding to the baseline signal values so that the relative entropy within this epoch is calculated as a measure of the "distance" between p_1 and p_2. In this way we consider the discrete sets $S_1 = \{s(t_k), k = 1, 2, ..., n_1\}$ and $S_2 = \{s(t_l), l = n_1 + 1, n_1 + 2, ..., n_1 + n_2\}$ of signal amplitudes measured at n_1 instants ($t_1 < t_2 < ... < t_{n_1}$) within period W_1 and n_2 instants ($t_{n_1+1} < t_{n_1+2} < ... < t_{n_1+n_2}$) within period W_2, respectively. In order to compute the probability functions p_1 and p_2, the accessible states of the system are then defined in terms of subdivisions of the interval of variation of the signal amplitude [5]. Consider the whole set $S = S_1 \cup S_2$ and let $s_0 = \min[S]$ and $s_L = \max[S]$ be the inferior and superior limits of S, respectively. An equipartition of S is defined by the amplitude values $s_0, s_1 = s_0 + \Delta s, s_2 = s_0 + 2\Delta s, ..., s_L = s_0 + L\Delta s$ where L is the number of subdivisions (levels) and $\Delta s = (s_L - s_0)/L$. It is assumed that each subdivision corresponds to a possible state of the system. Each integer j from the label set $\zeta = \{1, 2, ..., L\}$ refers to the the j-th disjoint interval I_j which is defined as $[s_{j-1}, s_j)$ (for $j \leq L - 1$) and $[s_{L-1}, s_L]$ (for $j = L$).

As already noted, the relative entropy might diverge in the limit $p_{2j} \to 0$ (for $p_{1j} \neq 0$). From a practical point of view, this undesirable divergence would be susceptible to occur if the probability function p_2 were allowed to assume null values. In order to avoid such divergence, we introduce a small positive parameter δ ($0 < \delta \leq 1$) in the definition of the probabilities of the signal levels,

$$p_{ij} = \frac{n_{ij} + \delta}{n_i + L\delta} , \tag{2}$$

where $n_{ij} = \left| S_i \cap I_j \right|$ is the number of points within window W_i ($i = 1, 2$) and level I_j ($j = 1, 2, ..., L$) and $n_i = |S_i|$ is the whole number of points in window W_i. In the above definition, the condition $\sum_{j=1}^{L} p_{ij} = 1$ is clearly satisfied whereas if $\delta \to 0^+$ then $p_{ij} \to \frac{n_{ij}}{n_i}$ so that one recovers the usual meaning that p_{ij} is the probability of occupation of level I_j (within W_i), i.e., the relative frequency of points on I_j. Furthermore, eq. 2 might be interpreted as the relative frequency of points on level I_j (within W_i) *given* that a fractionary point (of weight δ) is equally placed on each level I_j of both windows W_1 and W_2. In this sense, the introduction of parameter δ does *not* bias the comparison between distributions p_1 and p_2 since it still remains balanced.
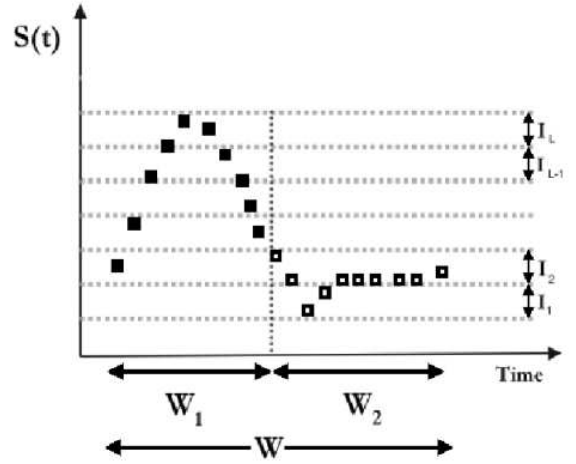

FIG. 1: Time course of the BOLD signal in a single activated voxel. The plot illustrates the set of points within the time interval related to signal increase (W_1) and the set of points within the second time interval (W_2) corresponding mainly to baseline values. The intervals $I_1, I_2, ..., I_L$ define the accessible levels (states) of the signal.

III. SIMULATED DATA ANALYSIS

The simulated time series $s(t)$ is composed of a hemodynamic response function $h(t)$ and a gaussian noise η, i.e., $s(t) = h(t) + \eta$. The exact hemodynamic response function that follows neuronal activity in an ER-fMRI paradigm depends on many variables [11] but in general the BOLD response reaches a maximum soon after a brief stimulation (4 to 5s) and then decays to the baseline level (at about 20s after stimulus). Pre-undershoot and a post-undershoot signals are sometimes observed [12]. In our simulations, we use a fisiologically reasonable model for the HRF which was proposed by Friston [13] and consists of the difference between

two gamma functions,

$$h(t) = \left(\frac{t}{d}\right)^a \exp\left(\frac{d-t}{b}\right) - c\left(\frac{t}{d'}\right)^{a'} \exp\left(\frac{d'-t}{b'}\right), \quad (3)$$

where $d = ab$ is the time corresponding to the peak and $d' = a'b'$ is the time for the undershoot, with $a = 6$, $a' = 12$, $b = b' = 0.9$ and $c = 0.35$.

Noise η is sampled from a gaussian distribution with mean zero and a given variance σ_n^2. Otherwise the variance σ_s^2 of the signal $h(t)$ is evaluated once along time and is the same for all simulations. Time courses $s(t)$ are simulated for different values of SNR (signal-to-noise ratio) which is usually defined as

$$SNR = 10\log\left[\frac{\sigma_s^2}{\sigma_n^2}\right] \quad (4)$$

and measured in decibels (dB).

Following the same parameters used in the real ER-fMRI experiment (see section IV), each simulated time course $s(t)$ consists of $N = 24$ epochs. For each epoch, $n_1 = n_2 = 7$ signal amplitudes are simulated within windows W_1 and W_2. Some time series $s(t)$ simulated for different signal-to-noise ratios (-4dB, 0dB, 4dB and 8dB) are plotted in Fig. 2.
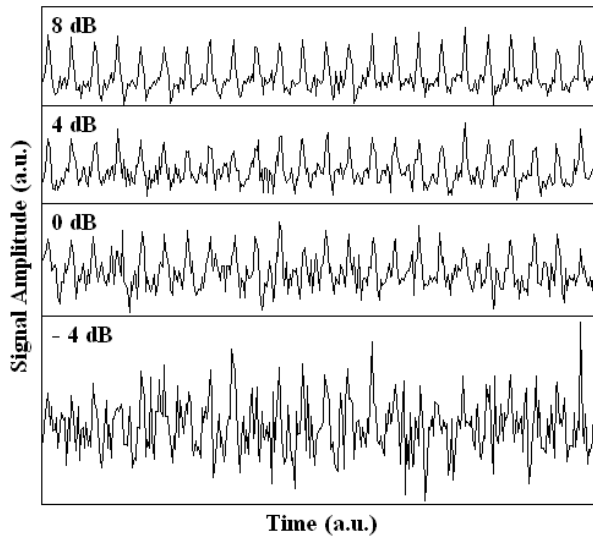

FIG. 2: Plot of some time courses $s(t) = h(t) + \eta$ simulated for different SNR's.

To determine the probability functions p_1 and p_2, we have chosen the parameters $L = 2$ and $\delta = 0.1$ for evaluation of eq. 2. As we shall see, that choice for δ leads to optimum results. The Kullback-Leibler distance D must be computed within each epoch. The mean Kullback-Leibler distance $\overline{D}$ is the sample average of the N distances computed along the whole time course. We perform $N_{exp} = 10^6$ simulations of the whole time series. Results from simulations show that the Kullback-Leibler distance D exhibits a discrete probability distribution with periodic peaks. On the other hand, the probability density of the mean distance, $f(\overline{D})$, is observed to approach a

continuous distribution for large values of N. According to the central limit theorem, $f(\overline{D})$ tends to a Gaussian distribution as $N \to \infty$. However, for small values of N ($N < 30$), we have found that the probability densities $f(\overline{D})$ can be fitted very well by Gamma distributions,

$$f(\overline{D}) \approx \frac{\overline{D}^{(\alpha-1)}}{\beta^\alpha \Gamma(\alpha)} \exp\left(-\frac{\overline{D}}{\beta}\right), \quad (5)$$

where α and β are parameters of the distribution related to its mean $\mu = \alpha\beta$ and variance $\sigma^2 = \alpha\beta^2$. Fig. 3 presents the probability densities $f(\overline{D})$ corresponding to simulations for pure noise ($SNR \to -\infty$) and some finite values of SNR (-8dB, -4dB, 0dB and 4dB); the continuous curves correspond to the Gamma fittings ($\chi^2 < 0.0009$) and overlap the simulated points very well.
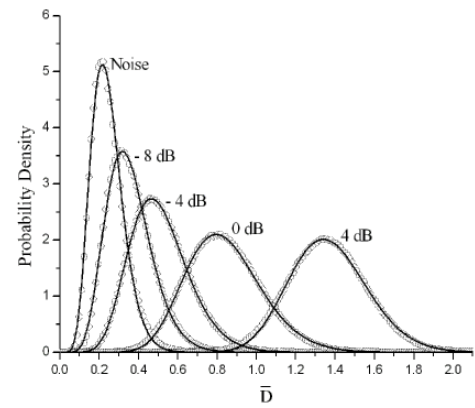

FIG. 3: Probability densities $f(\overline{D})$ for simulations of pure noise and some finite values of SNR with parameters $N = 24$, $n_1 = n_2 = 7$, $L = 2$ and $\delta = 0.1$. The continuous curves corresponding to Gamma fittings ($\chi^2 < 0.0009$) overlap the simulated points.

The mean Kullback-Leibler distance $\overline{D}$ is elected as the statistic for testing the null hypothesis H_0 that the time course $s(t)$ in a particular voxel is pure noise (non-active voxel) against the alternative hypothesis H_1 that $s(t)$ is composed of both BOLD signal and noise with a given signal-to-noise ratio (active voxel).

A signal detection scheme can be usually appraised by the construction of the receiver operating characteristic (ROC) curves [14, 15]. In the present analysis, the outcome of the signal detection scheme is the mean distance $\overline{D}$. Such outcome is classified as signal (+) if the observed $\overline{D}$ belongs to a given critical region (above some threshold $\overline{D}_c$); otherwise it is classified as noise (−). If $f(\overline{D}|-)$ and $f(\overline{D}|+)$ are the probability densities obtained from simulations of pure noise and signal (with a given SNR), respectively, then the *sensitivity* is defined as the probability of a true positive,

$$P(+|+) = \int_{\overline{D}_c}^\infty f(\overline{D}|+)\, d\overline{D}, \quad (6)$$

whereas the *specificity* is defined as the probability of a true

negative,

$$P(-|-) = \int_{-\infty}^{\overline{D}_c} f(\overline{D}|-) \, d\overline{D} \qquad (7)$$

(the probability functions are null for $\overline{D} < 0$). Fig. 4 illustrates the areas under the probability distributions $f(\overline{D}|-)$ (for noise only) and $f(\overline{D}|+)$ (for signal with SNR=-4dB) which correspond to specificity and sensitivity, respectively. In hypothesis testing, a type I error consists of rejecting the null hypothesis H_0 when it is true. The significance level α_{PT} per test is the probability of a type I error, i.e., $\alpha_{PT} = P(+|-) = 1 - P(-|-)$ is the probability of a false positive (or 1-specificity).
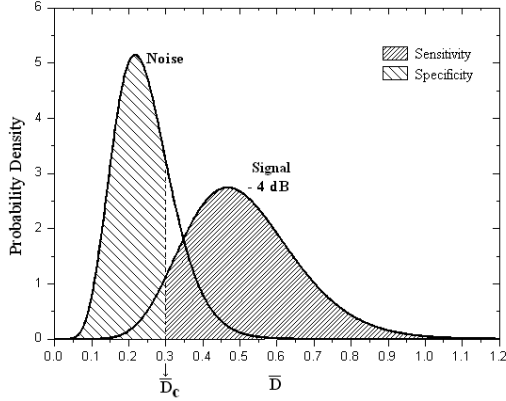

FIG. 4: Areas corresponding to specificity and sensitivity under probability densities $f(\overline{D}|-)$ (for simulated noise only) and $f(\overline{D}|+)$ (for simulated signal with SNR=-4dB).

The ROC curve is the plot of sensitivity (probability of a true positive) versus 1-specificity (probability of a false positive), resulting from the variation of the threshold $\overline{D}_c$. The ROC curves for several SNR's are shown in Fig.5.
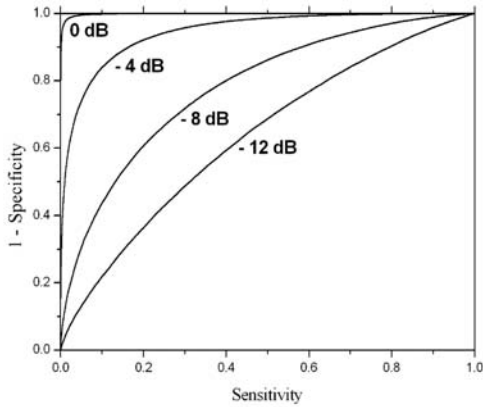

FIG. 5: ROC curves for different values of SNR.

The performance of the method is evaluated by its ability to detect active brain areas with high sensitivity and specificity. Thus the higher is the concentration of points in the left

superior corner of the ROC curve graph, the better is the performance. The *area* below the ROC curve is a way to quantify this performance. The discriminate power of the test increases as the ROC area varies from 0.5 (lowest value) to 1 (optimum performance). Our simulations show that the ROC area (and so the discriminate power) grows as SNR is increases according to Fig. 6. Finally, we may investigate the dependence of the power of the test with the choice of parameter δ. Fig. 7 shows that the ROC area varies smoothly with δ for various SNR'S and that $\delta = 0.1$ seems to be the best choice for all studied cases.
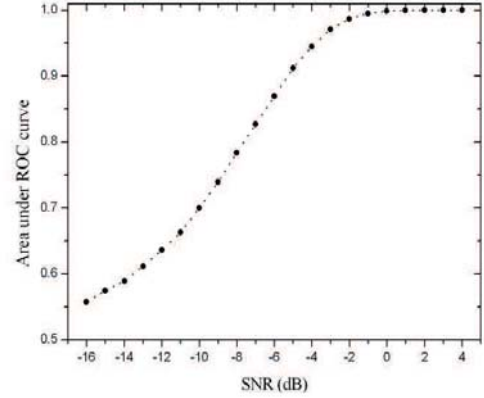

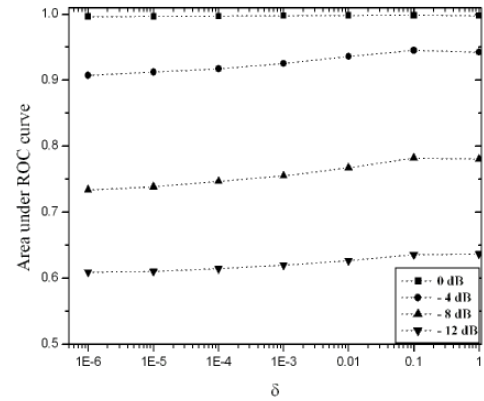FIG. 6: Area under ROC curve for different values of SNR.

FIG. 7: Area under ROC curve for different values of δ.

IV. REAL DATA ANALYSIS

An asymptomatic volunteer has been submitted to an event related motor stimulus. The fMRI data were acquired in a 1.5 T scanner (Siemens, Magneton Vision) with quadrature transmit/receive head RF coil and circular polarization commercially available. The study protocol consisted of 24 re-

peated epochs with two states: left- and right-handed finger-tapping with duration of 3-5 seconds and 17-21 seconds of rest. Three axial slices (Fig.8) with 4 mm of width (located on portions of the primary motor cortex) showed 14 temporal points throughout each epoch, three during the activity states and eleven during the period of rest, totalizing 336 images each slice. All images were acquired with echo-planar image (EPI) protocol with repetition time (TR) of 1680 ms, time to echo of 118 ms, flip angle of 90. Each slice was acquired with field of view (FOV) of 210 mm with a matrix 128 x 128 resulting a resolution of 1,64 x 1,64. Preprocessing was carried out using the Brain Voyager™ QX (Brain Innovation Inc., The Netherlands) software. The data analysis included several steps as motion correction, time series image realignment by the first slice as reference, spatial smoothing with a FWHM of 4 mm and temporal filtering with a high pass filter of 3 cycle/s.
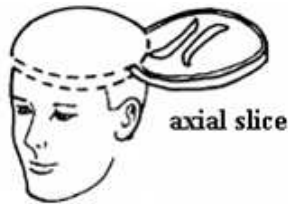

FIG. 8: Axial Slice.

For each voxel of the image, the mean Kullback-Leibler distance ($\overline{D}$) of the corresponding real data series has been computed (with parameters $L = 2$ and $\delta = 0.1$). One should remember that this processing requires *no* model for the HRF. The result of such analysis are the activation maps (overlaid onto corresponding anatomical images) for three axial slices, as shown in Fig. 9. The maps present activation in primary and secondary motor areas due to a motor stimulus. In order to avoid a lot of spurious activated voxels (false positives) in the functional maps, one should consider the *Šidàk − Bonferroni* correction [16],

$$\alpha_{PF} = 1 - (1 - \alpha_{PT})^C , \qquad (8)$$

where α_{PF} is the significance level per family of tests, defined as the probability of making at least one type I error for the whole family of tests (i.e., for all voxels); α_{PT} is the usual significance level (per test) and C is the number of tests (voxels). For a signicance level $\alpha_{PF} = 0.05$ per family of tests, the threshold for voxel activation is $\overline{D}_c = 0.9337$ so that voxels with $\overline{D} < \overline{D}_c$ are decided to be non-active.

V. CONCLUSION

In summary, we have proposed the use of the Kullback-Leibler distance (or relative entropy) as an alternative method

to analyse fMRI data series. The probability distributions of the mean Kullback-Leibler distance (throughout the *N* epochs of the event related experiment) were determined by numerical simulations. The sensitivity and specificity of the method
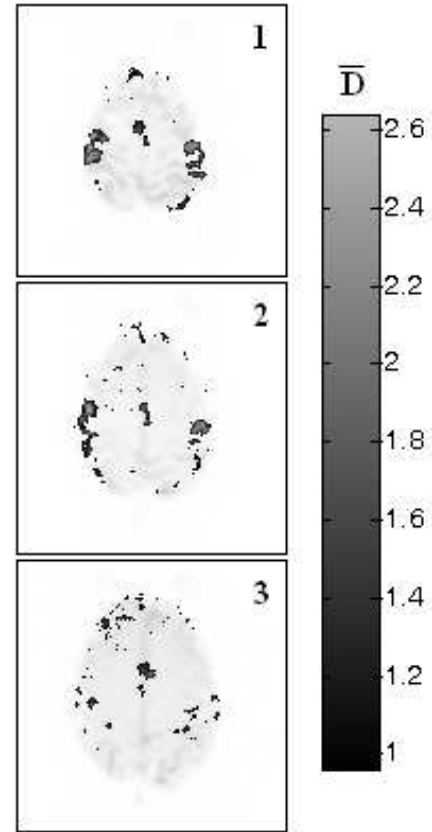

FIG. 9: Functional maps of three axial slices obtained by employment of Kullback-Leibler information measure. For a signicance level per family $\alpha_{PF} = 0.05$, the threshold for voxel activation is $\overline{D}_c = 0.9337$ so that voxels with $\overline{D} < \overline{D}_c$ are cut off (non-active).

were evaluated by construction of ROC curves for different SNR's. The functional map corresponding to real data series from an asymptomatic volunteer submitted to an ER motor stimulus was obtained by using the proposed technique. Our results from both simulated and real data indicate that the technique can be a very good option for fMRI analysis. As a further study, the use of the *generalized* Kullback-Leibler distance in fMRI analysis might be considered.

[1] D. Lebihan, *Diffusion and Perfusion Magnetic Resonance Imaging Applications to Functional MRI*, Raven Press, New York, 1995.

[2] S. Ogawa, T. Lee, A. Nayak, P. Glynn, *Magnetic Resonance in Medicine* **14**, 68 (1990).

[3] S. Ogawa, T. Lee, A. Kay, D. Tank, *Proceedings of the National Academy of Sciences USA*, 9868 **87** (1990).

[4] P. van Zijl, S. Eleff, J. Ulatowski, J. Oja, A. Ulug, R. Traystman, R. Kauppinen, Nature Medicine **4**, 159 (1998).

[5] A. Capurro, L. Diambra, D. Lorenzo, O. Macadar, M. T. Martin, C. Mostaccio, A. Plastino, J. Pérez, E. Rofman, M. E. Torres, J. Velluti, *Physica A* **265**, 235 (1999).

[6] T. Yamano, *Phys. Rev. A* **63**, 046-105 (2001).

[7] D. B. de Araujo, W. Tedeschi, A. C. Santos, J. Elias Jr., U. P. C. Neves, O. Baffa, *Neuroimage* **20**, 311 (2003).

[8] W. Tedeschi, D. B. de Araujo, H.-P. Müller, , A. C. Santos, U. P. C. Neves, S. N. Ernè, O. Baffa, *Physica A* **344**, 705-711 (2004).

[9] W. Tedeschi, H.-P. Müller, D. B. de Araujo, A. C. Santos, U. P. C. Neves, S. N. Ernè, O. Baffa, *Physica A* **352**, 629-644 (2005).

[10] T. M. Cover, J. A. Thomas, *Elements of information theory*, John Wiley & Sons, New York, (1991).

[11] F. M. Miezin, L. Maccotta, J. M. Ollinger, S. E. Petersen and R. L. Buckner, *NeuroImage* **11**, 735 (2000).

[12] D. Malonek, A. Grinvald, *Science* **272**, 551 (1996).

[13] K. J. Friston, P. Fletcher, O. Josephs, A. Holmes, M. D. Rugg, R. Turner, *NeuroImage* **7**, 30 (1998).

[14] J. Hanley, B. McNeil, *Radiology* **148**, 839 (1983).

[15] D. Goodenough, K. Rossmann, L. Lusted, *Radiology* **110**, 89 (1974).

[16] N. Salkind, *Encyclopedia of Measurement and Statistics*, Sage Publications, Inc, 2006.