



# **A new strategy for model order identification and its application to transfer entropy for EEG signals analysis.**

Chunfeng Yang, Regine Le Bouquin Jeannes, Jean-Jacques Bellanger,  
Huazhong Shu

## **► To cite this version:**

Chunfeng Yang, Regine Le Bouquin Jeannes, Jean-Jacques Bellanger, Huazhong Shu. A new strategy for model order identification and its application to transfer entropy for EEG signals analysis.. IEEE Transactions on Biomedical Engineering, 2013, 60 (5), pp.1318-27. 10.1109/TBME.2012.2234125 . hal-00880641

**HAL Id: hal-00880641**

**<https://hal.science/hal-00880641>**

Submitted on 15 Nov 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A New Strategy for Model Order Identification and Its Application to Transfer Entropy for EEG Signals Analysis

Chunfeng Yang, Régine Le Bouquin Jeannès, *Member, IEEE*, Jean-Jacques Bellanger, and Huazhong Shu, *Senior Member, IEEE*

**Abstract**—The background objective of this work is to analyze electroencephalographic (EEG) signals recorded with depth electrodes during seizures in patients with drug-resistant epilepsy. Usually, different phases are observed during the seizure evolution, including a fast onset activity. We aim to ascertain how cerebral structures get involved during this phase, in particular whether some structures “drive” other ones. Regarding a recent theoretical information measure, namely the Transfer Entropy (TE), we propose two criteria, the first one based on Akaike's information criterion, the second on the Bayesian information criterion, to derive models' orders that constitute crucial parameters in the TE estimation. A normalized index, named PTE, allows for quantifying the contribution or the influence of a signal to the global information flow between a pair of signals. Experiments are first conducted on linear autoregressive models, then on a physiology-based model, and finally on real intracerebral EEG epileptic signals to detect and identify directions of causal interdependence. Results support the relevance of the new measures for characterizing the information flow propagation whatever unidirectional or bidirectional interactions.

**Index Terms**—Bayesian information criterion, causality, EEG signal, physiology-based model, transfer entropy.

Manuscript received June 12, 2012; revised October 1, 2012; accepted November 3, 2012. This work was supported by the National Basic Research Program of China under Grant 2011CB707904, by the National Natural Science Foundation of China under Grants 61073138, 61271312, and 60911130370, by the Ministry of Education of China under Grant 20110092110023, the Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, and by Natural Science Foundation of Jiangsu Province under Grant SBK200910055.

C. Yang is with Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, 210096, Nanjing, China; the INSERM, U1099, Rennes; Université de Rennes 1, LTSI, F-35000; Campus de Beaulieu, Bât 22, 35042, Rennes Cedex, France and with Centre de Recherche en Information Biomédicale Sino-français (CRIBs) (e-mail: chunfeng.yang@seu.edu.cn).

R. Le Bouquin Jeannès and J. J. Bellanger are with the INSERM, U1099, Rennes; Université de Rennes 1, LTSI, F-35000; Campus de Beaulieu, Bât 22, 35042, Rennes Cedex, France and with Centre de Recherche en Information Biomédicale Sino-français (CRIBs) (corresponding author: R. Le Bouquin Jeannès; tel: 33223236220; fax: 33223236917; e-mail: {regine.le-bouquin-jeannes, jean-jacques.bellanger}@univ-rennes1.fr).

H. Shu is with Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, 210096, Nanjing, China and with Centre de Recherche en Information Biomédicale Sino-français (CRIBs) (e-mail: shu.list@seu.edu.cn).

## I. INTRODUCTION

EPILEPSY is a neurological disorder characterized by repetitive seizures. In 30% of cases, seizures remain drug-resistant and considerably affect all aspects of the patient's life [1]. Drug-resistant epilepsies are often partial, with an epileptogenic zone (EZ) located in a relatively circumscribed brain area. For these partial epilepsies, surgical treatment is considered. The difficulty that arises is to determine the organization of the EZ and, thus, the part of the brain that should be excised to remove seizures without affecting normal brain function. In some patients, the pre-surgical evaluation may include recording of intracerebral Electroencephalographic (iEEG) signals using depth electrodes. The analysis of such signals which remains a difficult task is aimed at determining which sites of the brain belong to the EZ, prior to surgery. In this context, signal processing techniques provide quantitative information that cannot be easily obtained by visual inspection. This is typically the case of correlation (in wide-sense) measures that prove useful for assessment of functional couplings between remote brain sites. Beside functional connectivity, effective connectivity provides complementary information on the directionality of couplings, i.e. the influence one neural population exerts on another one. Recently, a theoretical information measure, namely Transfer Entropy (TE), was proposed to identify the direction of the information flow and to quantify the strength of coupling between complex systems [2]. This “model-free” technique is based on the transition probabilities between states of the considered systems from output signals. This method has been applied and tested in some nonlinear benchmark models, real EEG signals and magnetoencephalographic data in cortical and cerebellar networks with no ground reference [3]–[6]. Moreover, from these foregoing literatures, certain calibration parameters involved in the TE estimation play a significant role in acquiring the correct information flow direction between the two systems under investigation, i.e. the driven system and the driving system, for instance, the orders of the underlying Markov processes. In [2], only the Markov process order of the driving system is mentioned: it is fixed to that of the driven one or set to 1 with the reason that the latter is preferable for computational reasons. In [3], [4], Sabesan *et al.* further

investigated this problem. Two measures, the delayed mutual information and the autocorrelation function [7], [8], were proposed to estimate the Markov process order of the driven system. However, the Markov process order of the driving system was still set to 1 since the authors presumed that the current state of the driving system was sufficient to arouse a considerable change in the dynamics of the driven system with one time step in the future. Furthermore, Vicente *et al.* [5] and Wibral *et al.* [6] used Cao's criterion [9] based on the false neighbor computation and autocorrelation function to estimate the Markov process orders. Two other model selection methods, the Akaike Information Criterion (AIC) [10] and the Bayesian Information Criterion (BIC) [11], have been also applied in Vectorial AutoRegressive (VAR) model order estimation and to estimate the order of Markov chains [12]. To compute transfer entropy on electrophysiological signals, AIC and BIC are currently used [13] to select the orders in a VAR model (including  $N$  channels  $X_1, \dots, X_N$ ) under the constraint that, when regressing  $X_j$  on  $X_i$ , the same number  $q_{ij} = q$  of lags is used for any  $(i, j)$  (where  $q$  must be determined and  $q_{ij}$  is named hereafter the regression order from  $X_i$  to  $X_j$  even if no explicit regression procedure is introduced to compute TE). In this case, this procedure returns only one value for the number of past samples retained on  $X$  and on  $Y$  to compute TE. In many cases, the "true" regression orders for the driven and driving systems may be different (this is typically the case for the cross regression order when there is no influence from one system to another). Consequently, it can lead to erroneous estimation in TE when the regression orders for the driven and driving systems are chosen artificially identical. To deal with this issue, we released the constraint  $\forall(i, j): q_{ij} = q$  by applying the generalized Akaike information criterion (gAIC) or the generalized Bayesian information criterion (gBIC). In Section II, after recalling the transfer entropy approach, we develop the methodology to estimate the regression orders for the driven and driving systems. We also introduce a transfer entropy based measure (called PTE) to quantify the contribution of an observation to another one and we give some details on the physiology-based model used as ground-truth. Section III is devoted to experimental results: two linear AR models were first introduced to verify the relevance of the two new order-estimated measures. Then, TE was tested on a physiological model and on real iEEG signals involving the new criteria. Results are discussed and a conclusion is drawn in Section IV.

## II. METHODS AND MATERIALS

### A. Transfer Entropy

Let  $u_n^{(p)}$  denote a  $p \times 1$  vector, such that:

$$u_n^{(p)} = [u_n, u_{n-1}, \dots, u_{n-p+1}]^T. \quad (1)$$

Considering a  $k^{th}$ -order Markov process  $X$ , the conditional probability measures (continuous or discrete) are such that:

$$\forall k' > k: P_{X_{n+1}/X_n^{(k)}=x_n^{(k)}}(dx_{n+1}) = P_{X_{n+1}/X_n^{(k')}=x_n^{(k')}}(dx_{n+1}). \quad (2)$$

The relation in (2) can be extended by considering two random processes  $X$  and  $Y$ . The absence of information flow from  $Y$  to  $X$  is then formalized by

$$P_{X_{n+1}/X_n^{(k)}=x_n^{(k)}}(dx_{n+1}) = P_{X_{n+1}/X_n^{(k)}=x_n^{(k)}, Y_n^{(l)}=y_n^{(l)}}(dx_{n+1}). \quad (3)$$

So, in this absence of information flow, the  $Y$  process values  $y_n^{(l)}$  have no influence on the  $X$  transition probabilities from the state value  $x_n^{(k)}$  to the state value  $x_{n+1}$  ( $X_{n+1}$  and  $Y_n^{(l)}$  are independent conditionally to  $X_n^{(k)}$ ). The deviation from this assumption can be quantified using the Kullback entropy, which leads to the definition of the transfer entropy

$$TE_{Y \rightarrow X} = \int_{\mathbb{R}^{k+l+1}} P_{X_{n+1}, X_n^{(k)}, Y_n^{(l)}}(dx_{n+1}, dx_n^{(k)}, dy_n^{(l)}) \times \log_2 \left( \frac{dP_{X_{n+1}/X_n^{(k)}=x_n^{(k)}, Y_n^{(l)}=y_n^{(l)}}}{dP_{X_{n+1}/X_n^{(k)}=x_n^{(k)}}} (x_{n+1}) \right) \quad (4)$$

where the ratio in (4) corresponds to a derivative of the conditional measure in the numerator with respect to the conditional measure in the denominator. This measure is not symmetric ( $TE_{X \rightarrow Y} \neq TE_{Y \rightarrow X}$ ). The values of the parameters  $k$  and  $l$  are the regression orders to predict  $X_{n+1}$  from  $X$  past and  $Y$  past. The estimation of TE from the observations  $(x_n, y_n)$ ,  $n=1, \dots, N$ , is obtained using a discrete kernel estimation of  $(X_{n+1}, X_n^{(k)}, Y_n^{(l)})$  probability distribution [2]

$$\hat{P}(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \propto \sum_m \Theta \left( r - \left\| \begin{pmatrix} x_{n+1} - x_{m+1} \\ x_n^{(k)} - x_m^{(k)} \\ y_n^{(l)} - y_m^{(l)} \end{pmatrix} \right\| \right) \quad (5)$$

$$\triangleq C_{n,r}, \quad n = k+1, \dots, N-1$$

which depends on a neighborhood size (radius  $r$ ), where  $x_n^{(k)}$  and  $x_n^{(l)}$  are defined as in (1). The kernel function is formalized with the step function  $\Theta(x > 0) = 1$ ,  $\Theta(x \leq 0) = 0$  and the norm  $\|\bullet\|$  is the max norm. The summation is performed for  $m \in [k+1, \dots, N-1]$  by excluding some evident redundancies. The value of  $r$  is chosen in the linear region of the curve  $\ln C(r)$  vs  $\ln r$  obtained as an average of  $n$  over  $C_{n,r}$ . Then, it is used to compute the estimation

$$\widehat{TE}_{Y \rightarrow X} = \sum_n \hat{P}(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log_2 \frac{\hat{P}(x_{n+1} | x_n^{(k)}, y_n^{(l)})}{\hat{P}(x_{n+1} | x_n^{(k)})}. \quad (6)$$

In practice, the estimations of the conditional probabilities in (6) are computed from the estimated joint probabilities in (5)

using Bayes' rule  $P(A|B) = P(A, B) / P(B)$ . Thereby, the definition in (6) is rewritten as

$$\widehat{TE}_{Y \rightarrow X} = \sum_n \hat{P}(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \times \log_2 \frac{\hat{P}(x_{n+1}, x_n^{(k)}, y_n^{(l)}) P(x_n^{(k)})}{\hat{P}(x_{n+1}, x_n^{(k)}) P(x_n^{(k)}, y_n^{(l)})}. \quad (7)$$

Considering the more complicated issue of real data [14], we define a selective index of the flow direction, named PTE, by

$$PTE_{Y \rightarrow X} = \frac{|TE_{Y \rightarrow X}|}{|TE_{X \rightarrow Y}| + |TE_{Y \rightarrow X}|}, \quad (8)$$

$$PTE_{X \rightarrow Y} = \frac{|TE_{X \rightarrow Y}|}{|TE_{X \rightarrow Y}| + |TE_{Y \rightarrow X}|} \quad (9)$$

to help in quantifying the contribution of a signal to another one compared to the global information flow induced by these two signals. Note that theoretically (i) when only  $Y$  drives  $X$  (resp.  $X$  drives  $Y$ ),  $PTE_{Y \rightarrow X}$  (resp.  $PTE_{X \rightarrow Y}$ ) equals 1, (ii) if there exist two equivalent flows for the two directions, values of  $PTE_{Y \rightarrow X}$  and  $PTE_{X \rightarrow Y}$  equal 0.5, and (iii) a null value of  $PTE_{Y \rightarrow X}$  (resp.  $PTE_{X \rightarrow Y}$ ) reveals that no information flow exists from  $Y$  to  $X$  (resp. from  $X$  to  $Y$ ). Clearly, when estimations of PTE are substituted for theoretical PTE, some caution must be introduced. In practice, the selective index PTE is calculated only when the variables are first proven not to be independent.

### B. Improved Computation of Transfer Entropy

- *Selection of the parameters  $k$  and  $l$*

For the two processes  $X$  and  $Y$ , the selection of the regression orders  $k$  and  $l$  plays experimentally a primary role to acquire dependable values for transfer entropy. As suggested before, to get the models' orders, AIC and BIC are often applied under the hypothesis that  $k$  and  $l$  are identical, which not only produces extra computation but also can result in erroneous estimation in TE. To solve this problem, we propose two measures, gAIC and gBIC, to choose independently different values for the parameters  $k$  and  $l$ . The procedures of gAIC and gBIC are detailed hereafter.

Given a real  $m$ -dimensional vectorial autoregressive process  $z_n = [z_n(1), z_n(2), \dots, z_n(m)]^T$  with zero mean

$$z_n = \Phi_1 z_{n-1} + \Phi_2 z_{n-2} + \dots + \Phi_q z_{n-q} + w_n \quad (10)$$

where  $\Phi_i, i=1, 2, \dots, q$ , are  $m \times m$  coefficient matrices and  $q$  the model order. The process  $w_n$  is a zero mean independent identically distributed vector process with covariance matrix  $\Sigma$ . It is assumed that  $w_n$  and  $z_{n-i}$  are independent for each variable  $i > 0$ . In addition, it is assumed that  $z_n$  is mean and covariance ergodic.

Suppose that the observations  $z_1, z_2, \dots, z_N$  are generated by the VAR process given in (10). The least-squares method [15] is used to fit a VAR model with order  $q$  to the observations. Suppose that the estimated model is as follows:

$$z_n = \hat{\Phi}_1 z_{n-1} + \hat{\Phi}_2 z_{n-2} + \dots + \hat{\Phi}_q z_{n-q} + \hat{w}_n. \quad (11)$$

The coefficient matrices  $\hat{\Phi}_i, i=1, 2, \dots, q$ , of this estimated model are obtained by solving the following set of equations:

$$Z A_j = z_{q+1,j}; j=1, 2, \dots, m \quad (12)$$

$$\text{where } Z = \begin{bmatrix} z_q^T & z_{q-1}^T & \dots & z_1^T \\ z_{q+1}^T & z_q^T & \dots & z_2^T \\ \vdots & \vdots & \ddots & \vdots \\ z_{N-1}^T & z_{N-2}^T & \dots & z_{N-q}^T \end{bmatrix}, \quad A_j = \begin{bmatrix} \Phi_1^T(j, :) \\ \Phi_2^T(j, :) \\ \vdots \\ \Phi_q^T(j, :) \end{bmatrix},$$

$$\text{and } z_{q+1,j} = \begin{bmatrix} z_{q+1}(j) \\ z_{q+2}(j) \\ \vdots \\ z_N(j) \end{bmatrix}$$

where  $\Phi_i(j, :)$  is the  $j^{\text{th}}$  row of  $\Phi_i$ . The least-squares solution of (12) is given by

$$\hat{A}_j = (Z^T Z)^{-1} Z^T z_{q+1,j}; j=1, 2, \dots, m. \quad (13)$$

Then, the residual covariance matrix of the VAR model is estimated as follows:

$$\hat{\Sigma} = \frac{1}{N-q} \sum_{n=q+1}^N (z_n - \hat{z}_n)(z_n - \hat{z}_n)^T \quad (14)$$

where  $\hat{z}_n = \hat{\Phi}_1 z_{n-1} + \hat{\Phi}_2 z_{n-2} + \dots + \hat{\Phi}_q z_{n-q}$ .

To select the order parameter  $q$ , the Akaike Information Criterion is given by:

$$AIC(q) = N \ln(\det(\hat{\Sigma})) + 2m^2 q, \quad (15)$$

and the Bayesian Information Criterion is:

$$BIC(q) = N \ln(\det(\hat{\Sigma})) + m^2 q \ln(N). \quad (16)$$

Rewriting (10) with a 2-dimensional ( $m=2$ ) VAR process  $z_n = [z_n(1), z_n(2)]^T = [x_n, y_n]^T$ , we have

$$\begin{cases} x_n = \sum_{i=1}^{q_{xx}} \alpha_{xx}(i) x_{n-i} + \sum_{i=1}^{q_{yx}} \alpha_{yx}(i) y_{n-i} + w_{n,x} \\ y_n = \sum_{i=1}^{q_{xy}} \alpha_{xy}(i) x_{n-i} + \sum_{i=1}^{q_{yy}} \alpha_{yy}(i) y_{n-i} + w_{n,y} \end{cases} \quad (17)$$

where  $\alpha$  are the coefficients of the VAR process. AIC and BIC can be used to determine a common summation range for the four sums in (17):  $q_{xx} = q_{yx} = q_{xy} = q_{yy} = q$ . Since we do not impose  $k=l$ , we estimate independently the four orders parameters so that we may have  $k(=q_{xx}) \neq l(=q_{yx})$  to

compute  $TE_{Y \rightarrow X}$  and  $k(=q_{yy}) \neq l(=q_{xy})$  to compute  $TE_{X \rightarrow Y}$ . In fact, for most of physical cases, these four orders are different, and the standard approach (AIC or BIC) only gives the maximal value among them. Moreover, if we choose the maximal value for the parameters  $k$  and  $l$ , TE computational time is drastically increased. Focusing on this point, we propose to minimize the following criteria, namely gAIC and gBIC, with respect to the four orders  $\{q_{xx}, q_{yx}, q_{xy}, q_{yy}\}$

$$\text{gAIC}(\tilde{q}) = N \ln(\det(\hat{\Sigma})) + 2(q_{xx} + q_{yx} + q_{xy} + q_{yy}), \quad (18)$$

$$\text{gBIC}(\tilde{q}) = N \ln(\det(\hat{\Sigma})) + (q_{xx} + q_{yx} + q_{xy} + q_{yy}) \ln(N), \quad (19)$$

where  $\tilde{q}$  is a set of estimated orders  $\{\tilde{q}_{xx}, \tilde{q}_{yx}, \tilde{q}_{xy}, \tilde{q}_{yy}\}$ . Using gAIC and gBIC criteria is justified by considering that AIC and BIC criteria have been developed under less restrictive constraints [11], [16], [17] (i.e. not only for linear models such as VAR models). In this widened context, the log-likelihood function (which corresponds in (16) to the first term on the right hand side) must only belong to an exponential family, and the tested models do not need to be nested. When applied to VAR order selection, even if the search of candidate models may be limited to the subclass defined by  $q_{xx} = q_{yx} = q_{xy} = q_{yy}$ , as it is generally done in most applications, this constraint can be interestingly released to expect a better fitting without losing optimality relatively to the type of criterion.

Using gAIC and gBIC to get the four orders, the transfer entropy computation time is reduced, even though the computation time to evaluate the four orders is increased compared to the time needed to compute the orders with AIC and BIC. For example, given  $q \in [1, 20]$ , procedures (12) to (14) merely need to be implemented 20 times to minimize AIC and BIC. For gAIC and gBIC, this step needs to be implemented  $21^4 - 1$  times since  $\{q_{xx}, q_{yx}, q_{xy}, q_{yy}\} \in [0, 20]^4$  respectively and, therefore, it becomes prohibitive. To solve this problem, we propose a “greedy” strategy [18] to optimize the computation. This procedure using gAIC is detailed below (let us note that this procedure would be the same using gBIC):

- 1) Use AIC to obtain a model order, namely the maximum order  $q_{\max}$ , and select this order to decide the range for gAIC, i.e.  $q_{ij} \in [0, q_{\max}]$ ,  $i, j \in \{x, y\}$ ;
- 2) Considering the values of  $q_{ij}$  for the first signal (i.e.  $q_{xx}$  and  $q_{yx}$ ), while setting the values of  $q_{ij}$  for the second signal equal to  $q_{\max}$  (i.e.  $q_{xy}$  and  $q_{yy}$ ), implement the procedures from (12) to (14) to get the optimal orders  $\hat{q}_{xx}$  and  $\hat{q}_{yx}$  using the definition of (18);
- 3) Considering the values of  $q_{ij}$  for the second signal (i.e.

$q_{xy}$  and  $q_{yy}$ ), while setting the values of  $q_{ij}$  for the first signal equal to the optimal order  $\hat{q}_{xx}$  and  $\hat{q}_{yx}$ , implement the procedures from (12) to (14) to get the optimal orders  $\hat{q}_{xy}$  and  $\hat{q}_{yy}$  using the definition of (18).

According to this “greedy” strategy, the computation time is decreased significantly. As an example, the global procedure only needs to be implemented  $20 + 16^2 \times 2 - 1$  times (0.27% of the computation of AIC) when  $q_{\max} = 15$ .

- *Selection of the Radius  $r$*

For the selection of the radius  $r$ , we take the similar notion as in [3], [4], [19]. Firstly, the data are normalized to zero mean and unit variance in order to establish a common radius  $r$  in the state space  $X$  and  $Y$ . Then, the joint probability in log scale ( $\ln C(r)$  obtained as an average on  $n$  over  $C_{n,r}$  in (5)) is calculated and plotted vs the corresponding radius  $r$  in log scale (e.g. in Fig. 2(a)). Finally, some optimal value  $r^*$  is chosen in the linear region of the curve  $\ln C(r)$  vs  $\ln r$  so as to estimate reliable values of TE at specific values  $r^*$ .

### C. Model of iEEG Signals Generation

We used a physiology-based time continuous model to represent the electrical field activity of some distant - and possibly coupled - neuronal populations. Each of them generates a local field activity that is converted to an iEEG signal in a proximal electrode using a quasi-static transfer function [20]. In the model, each population contains three subpopulations of neurons that mutually interact via excitatory or inhibitory feedback linking main pyramidal cells and two other types of local interneurons. Since pyramidal cells are excitatory neurons that project their axons to other areas of the brain, the model accounts for this organization by using the average pulse rate of action potentials from the main cells of one population  $i$  as an excitatory input to main cells inputs of another population  $j$ . In addition, this connection from population  $i$  to population  $j$  is represented by a parameter  $K^{ij}$  which is proportional to the number of corresponding axonal links. Appropriate setting of this parameter allows for building systems where the neuronal populations are unidirectionally or bidirectionally coupled. Other introduced parameters are intra-population parameters. They include excitatory and inhibitory gains in feedback loops as well as coefficients related to the number of synaptic contacts between subpopulations. These parameters are adjusted to control the intrinsic activity of each population (normal background activity vs epileptic activity). This population model [20] is viewed as a stochastic state variable model. For each population  $j$ , this model comprises  $N = 11$  state components  $s_j^k$ ,  $k = 1, \dots, 11$ , grouped in a state vector  $s_j$ . Considering two populations, each generating a simulated EEG signal, the pair

$(s_1, s_2)$  is driven by two coupled Stochastic Differential Equations:

$$ds_1(t) = A(s_1(t))dt + Bd w_1(t) + C_{21}z_2(t)dt, \quad (20)$$

$$ds_2(t) = A(s_2(t))dt + Bd w_2(t) + C_{12}z_1(t)dt, \quad (21)$$

from which are derived the two outputs  $z_1$  and  $z_2$

$$z_1(t) = f(Ds_1(t)), z_2(t) = f(Ds_2(t)). \quad (22)$$

The quantities  $w_1$  and  $w_2$  are internal Brownian noise sources representing distal influences from other neural populations not explicitly modeled,  $B$  and  $C_{ij}$  are 11-dimensional vectors and the  $A$  operator is a mapping from  $\mathbb{R}^{11}$  to  $\mathbb{R}^{11}$  including four instantiations of a nonlinear mapping from  $\mathbb{R}$  to  $\mathbb{R}^+$  defined as a smooth nonlinear sigmoid function  $f$ . This type of function is necessary to take into account effects of thresholding and saturation in neurons [20]. In (22),  $D$  is an 11-dimensional row vector and the populations' outputs (which are not directly measured) are obtained through the sigmoid function. The time discretization of (20) and (21), needed to get a computational model, cannot lead to a linear representation as in (17) and so the resulting model induces nonlinear dynamics. Finally, both observations  $x$  and  $y$  are obtained as a linear combination of three of the eleven state variables.

#### D. Real Signals

The tested signals correspond to ictal epileptiform activities obtained from an animal model (epilepsy induced in guinea pig brain with bicuculline). Simultaneous signals were recorded in the medial entorhinal cortex and in the perirhinal cortex. The recorded EEG signals are plotted in Fig. 5. In the following section, a particular phase of these signals, the FOA (Fast Onset Activity), corresponding to the time interval  $[30s, 40s]$ , is tested by the measures we developed.

### III. EXPERIMENTS AND RESULTS

Simulations were carried out 200 times on 2048-point signals (corresponding to 4-s length EEG signals in current acquisition systems) both for linear and physiology-based models. For each computation of TE, the orders are estimated by the 4 different measures (AIC, BIC, gAIC, and gBIC) discussed in section II, and the values of  $r$  is chosen from  $e^{-1.5}$  to  $e^{2.0}$  with a step of  $e^{0.05}$ .

#### A. Linear Models

For the first linear stochastic system we considered, the following signals with unidirectional relation were generated:

$$\begin{cases} x_t = 0.95\sqrt{2}x_{t-1} - 0.9025x_{t-2} + u_t \\ y_t = 0.5x_{t-3} - 0.4y_{t-2} + v_t \end{cases} \quad (23)$$

where  $u_t$  and  $v_t$  were independent white Gaussian noises with zero means and unit variances. To compare the performance of

the 4 criteria (AIC, BIC, gAIC, and gBIC), we indicated in Table I the number of times the orders obtained with these 4 measures equal the expected optimal orders over the 200 trials.

As discussed in Section II. B, the expected optimal order values of the model described by (23) are  $\hat{q}_{xx} = 2$ ,  $\hat{q}_{yx} = 0$ ,  $\hat{q}_{xy} = 3$ , and  $\hat{q}_{yy} = 2$ . In this way, considering TE, we must obtain  $k = \hat{q}_{xx} = 2$ ,  $l = \hat{q}_{yx} = 0$  for  $\widehat{TE}_{Y \rightarrow X}$ , and  $k = \hat{q}_{yy} = 2$ ,  $l = \hat{q}_{xy} = 3$  for  $\widehat{TE}_{X \rightarrow Y}$ . As recalled before, the estimated order values using AIC and BIC are all the same, and are equal to the maximum order value, i.e.  $q_{xx} = q_{yx} = q_{xy} = q_{yy} = 3$  in the present case (see Table I). So, in this experiment, 3 parameters out of 4 are not correctly estimated. The gAIC measure is a reasonable estimator since most of the estimated orders are the optimal ones (for example, 189 correct estimations over a total of 200 (94.5%) for  $q_{xx}$ ). In this experiment, the gBIC measure behaves the best (100% of accuracy estimation for the 4 parameters).

Given the orders estimated by the 4 techniques, we computed the means of the corresponding TE values over the 200 simulations. Results are shown in Fig. 1 for different values of the radius  $r$ . Using gBIC,  $\widehat{TE}_{X \rightarrow Y}$  is the greatest in some range of  $\ln r$  (around the vertical dotted line) when there is effective information flow from  $x$  to  $y$  while  $\widehat{TE}_{Y \rightarrow X}$  is zero when there is no information flow from  $y$  to  $x$  (see Fig. 1(d)). Using AIC and BIC (see Figs. 1(a) and 1(b)), we note some non-zero values for  $\widehat{TE}_{Y \rightarrow X}$  due to some erroneous model orders: the estimated orders  $k$  ( $= q_{xx}$ ) and  $l$  ( $= q_{yx}$ ) are always equal to 3 which are not the optimal orders. Using gAIC, the transfer entropy from  $x$  to  $y$  is close to the same measure using gBIC, while the reciprocal quantity  $\widehat{TE}_{Y \rightarrow X}$  is mostly close to zero (see Fig. 1(c)) thanks to only a few incorrect estimated orders. For each technique, to ensure a correct analysis of the information flow direction, a specific radius  $r^*$  is chosen in the linear region of the curve  $\ln C(r)$  vs  $\ln r$  (Fig. 2(a)). Then, the means and standard deviations (shown in parentheses) of the corresponding TE values at this specific radius  $r^*$  are given in Table II as well as those of the PTE index. For sake of clarity, the notation  $\widehat{TE}_{X \rightarrow Y}$  (M) represents the  $\widehat{TE}_{X \rightarrow Y}$  results corresponding to the technique noted M (the technique refers here to the model order criterion). From this table, we observe that  $\widehat{TE}_{X \rightarrow Y}$  (gBIC) gives the greatest values when the direction of information flow is from  $x$  to  $y$ , and  $\widehat{TE}_{Y \rightarrow X}$  (gBIC) is zero when no information flow exists from  $y$  to  $x$  since the estimated orders are always identical to the optimal orders. On the opposite,  $\widehat{TE}_{Y \rightarrow X}$  (AIC),  $\widehat{TE}_{Y \rightarrow X}$  (BIC), and  $\widehat{TE}_{Y \rightarrow X}$  (gAIC) still provide some non zero values due to different values of the estimated order.

To test the reliability of the new criteria, some parameters values of our linear models were modified and the corresponding effects summarized in Annex A. Finally, to complete these results and test the relevance of the approach, an extended range of model orders values around the best ones chosen by the algorithm was tested and subsequent impacts on TE estimation are reported in Annex B.

For the second linear stochastic system, the following signals with bidirectional relation were generated:

$$\begin{cases} x_t = 0.95\sqrt{2}x_{t-1} - 0.9025x_{t-2} + 0.45y_{t-2} + u_t \\ y_t = 0.5x_{t-3} - 0.4y_{t-2} + v_t \end{cases} \quad (24)$$

where  $u_t$  and  $v_t$  were independent white Gaussian noises with zero means and unit variances. The numbers of trials (over 200 trials) for which the model orders obtained by the 4 criteria equal the expected optimal order are presented in Table III. Given these orders, TE is computed on 200 trials and the averaged values are displayed in Fig. 3 when the radius  $r$  in log scale is in the range  $[-1.5; 2]$ . In the same way as previously, a specific radius  $r^*$  is retained, and the means and standard deviations (shown in parentheses) of the corresponding TE and PTE values are given in Table IV. From Table III, the gBIC order estimator is still the more accurate. This criterion leads to the highest values for  $\widehat{TE}_{X \rightarrow Y}$  (gBIC) and  $\widehat{TE}_{Y \rightarrow X}$  (gBIC) (see Table IV).

If we focus on the PTE index (see Tables II and IV), both  $PTE(AIC)$  and  $PTE(BIC)$  conclude to a bidirectional information flow whatever the actual situation (unidirectional or bidirectional flow) whereas the two other indices  $PTE(gAIC)$  and  $PTE(gBIC)$  allow us to illustrate the two different situations: unidirectional vs bidirectional flow. For example, when the propagation is unidirectional (see Table II), the index  $\widehat{PTE}_{X \rightarrow Y}$  (gBIC) is always 1 and the index  $\widehat{PTE}_{Y \rightarrow X}$  (gBIC) zero, concluding to a unidirectional flow from  $x$  to  $y$ . When the propagation is bidirectional (see Table IV), the  $\widehat{PTE}_{X \rightarrow Y}$  (gBIC) index equals 0.6205 and  $\widehat{PTE}_{Y \rightarrow X}$  (gBIC) is 0.3795 which is coherent with the propagation graph. Note that PTE cannot be solely interpreted without verifying that  $\widehat{TE}_{X \rightarrow Y}$  or  $\widehat{TE}_{Y \rightarrow X}$  is significantly high.

### B. Physiology-based Models

The model described in Section II.C was used to simulate long duration signals (800 s) for a fixed connectivity pattern (“ground truth”) among neuronal populations. Sampling rate was equal to 512 Hz. Two simulated physiology-based models with uni- and bi-directionally coupled connectivities are considered. Model parameters were such that a fast quasi-sinusoidal (25 Hz) activity (similar to that observed during seizure onset) was generated by the two populations when they were uni- or bi-directionally coupled. The coupling parameters were such as:  $K^{12} = 1500$ ,  $K^{21} = 0$  for unidirectionally coupled populations 1 and 2, and

$K^{12} = K^{21} = 1500$  for bidirectionally coupled populations 1 and 2.

In this section, according to the results obtained on the simulated linear models, we only tested the performance of the gBIC technique and compared it to the BIC. Given the orders estimated by gBIC, we computed the corresponding TE values and averaged them on 200 trials for different values of the radius  $r$  as shown in Fig. 4. As previously, a specific radius  $r^*$  is chosen for each pair (1, 2), and the means and standard deviations (in parentheses) of the corresponding TE and PTE values are displayed in Table V.

As expected, the important values of  $\widehat{TE}_{1 \rightarrow 2}$  (gBIC) (both for unidirectional and bidirectional connectivities) and  $\widehat{TE}_{2 \rightarrow 1}$  (gBIC) (for reciprocal connectivity) indicate that an actual information flow exists from one population to the other. The non significant value of  $\widehat{TE}_{2 \rightarrow 1}$  (gBIC) for unidirectional connectivity indicates that there is no information flow, whereas the strong value for  $\widehat{TE}_{2 \rightarrow 1}$  (BIC) is erroneous.

Focusing on the PTE index, the value of  $\widehat{PTE}_{2 \rightarrow 1}$  (BIC) (0.4034) is incorrect for unidirectional connectivity since there is no causal influence from population 2 to population 1. In other words,  $PTE(BIC)$  fails in identifying unidirectional connectivity. On the contrary,  $PTE(gBIC)$  performs well and allows distinguishing unidirectional and bidirectional connectivities. For one-way propagation,  $\widehat{PTE}_{1 \rightarrow 2}$  (gBIC) is 1 and  $\widehat{PTE}_{2 \rightarrow 1}$  (gBIC) is zero. For two-way propagation, where  $K^{12} = K^{21} = 1500$ , the values of  $\widehat{PTE}_{1 \rightarrow 2}$  (gBIC) and  $\widehat{PTE}_{2 \rightarrow 1}$  (gBIC) are close to 0.5.

To confirm these results, we conducted a statistical test. Two independent populations were first simulated (using  $K^{12} = K^{21} = 0$ ). Then, we computed TE with the order-estimated methods of BIC and gBIC when these populations are coupled or not. The results of the Mann-Whitney test are summarized in Table VI ( $K^{12} = 1500, K^{21} = 0$  vs  $K^{12} = K^{21} = 0$  for unidirectional relation and  $K^{12} = K^{21} = 1500$  vs  $K^{12} = K^{21} = 0$  for bidirectional relation). In this table, the statistical analysis is presented in two columns, the first one corresponding to the  $p$ -value and the second one to the result  $h$  of the hypothesis test (a value of  $h = 1$  indicates that the null-hypothesis tested on the 200 trials is rejected). The expected value according to our simulation is also given into brackets. A value of  $h = 1$  proves the presence of an information flow. For the statistical analysis on TE (gBIC), all values are identical to the expected values whatever unidirectional or bidirectional propagation flows. As for  $\widehat{TE}_{2 \rightarrow 1}$  (BIC), in the case of unidirectional flow, the null-hypothesis is rejected which indicates that a propagation flow from population 2 to population 1 is erroneously detected.

### C. Real Signals

As discussed in Section II.D, we analyzed two real signals recorded during ictal onset. This phase is characterized by a peculiar pattern of fast activity that originates from the entorhinal/hippocampal region and, secondarily, visually propagates to the perirhinal cortex. Signal 1 was recorded in the medial entorhinal cortex (m-EC) and signal 2 in the perirhinal cortex (PRC). The time duration of the fast onset activity phase was 10 seconds. To detect the relations at different time instants, a 4-s sliding window with a step of 0.125 s was chosen for the estimation of the TE index, leading to a total of 49 TE values. In Fig. 6, we represented the evolution of this index with the orders estimated by gBIC and BIC over time during the analyzed period corresponding to this fast onset activity phase. To estimate the actual influence of signal  $i$  on signal  $j$ , and to evaluate the statistical significance of the TE measure, we tested this index against the null hypothesis of independent time series. To this end, we needed surrogate data for which we destroyed the causal dependency. We used box plots corresponding to this null hypothesis represented respectively at the bottom-left for the relation from signal 1 to signal 2, and, at the bottom-right for the relation from signal 2 to signal 1 in Fig. 6. Let us indicate that, in Fig. 6, the boxes correspond to TE estimation using gBIC knowing that results were quite comparable using BIC in this case of independence. These boxes have been obtained by randomizing the phase [21] of signal 2 without modifying signal 1, to simulate the independence hypothesis  $H_0$  (between signal 1 and signal 2). Note that, if shuffling the phase destroyed the dependence between the two signals, it also destroyed the nonlinearity in signal 2. Even if this effect was not desired, it had only a small influence on the statistical dispersion in TE estimation. Another approach would consist in shuffling the samples time instants. In this case, to test the  $H_0$  hypothesis, the gBIC procedure and TE estimation should be performed using the pairs of signals  $(x_{n+1}, x_n^{(k)})$  and  $(y_{n'+1}, y_n^{(l)})$  with a difference in time,  $n - n'$ , sufficiently large to ensure statistical independence between these signals (instead of the pairs  $(x_{n+1}, x_n^{(k)})$  and  $(y_{n+1}, y_n^{(l)})$ ). Both approaches (shuffle phase based approach and shuffle time based approach) leave invariant the second-order properties of the modified signal. In our experiment, we considered the whole duration, i.e. 10 seconds of observation, and we chose 4 windows of 4 seconds each, with an overlap of 2 seconds. For each window, we performed 50 realizations by randomizing the phase so as to get 200 values ( $4 \times 50$ ) of the index under  $H_0$  merged in these box plots. From the results in Fig. 6, using gBIC, it is obvious that a strong relation exists from signal 1 to signal 2 in the former part of the interval whereas the relation from signal 2 to signal 1 is completely non-existent. In the latter part of the sequence, the

propagation from the entorhinal region to the perirhinal cortex still holds at a non negligible level, this result being consistent with currently accepted propagation pathways. Compared to BIC, the gBIC criterion allowed for getting larger dynamics on the whole sequence and helped for concluding on the a priori predominant flow direction. On the first part of the sequence, it is clear that the m-EC structure drove the PRC one whatever the criterion. For the second part of the sequence, some estimated orders were the same using either BIC or gBIC, which justifies the superposition of the two curves of TE for the relation from signal 1 to signal 2. Since some orders differed, even if the TE index remained low when estimated in the opposite direction (from signal 2 to signal 1), its values revealed higher when using the classical criterion instead of the new one.

### IV. DISCUSSION AND CONCLUSION

In this paper, we improved the estimation of transfer entropy by introducing a greedy strategy to derive a suitable order-estimated measure which is decisive in acquiring coherent and trustworthy results, as shown in bivariate linear AR models displaying either unidirectional or bidirectional flows. The approach including the new order-estimated measure was successfully applied to detect effective connectivity in iEEG signals simulated by physiology-based models. Moreover, the introduction of a second index, namely  $PTE_{i \rightarrow j}$ , allowed us to quantify the contribution of population  $i$  towards population  $j$  in the global information flow between the two populations under study. When tested on real signals, the proposed approach appeared robust and helpful in the determination of the organization of the epileptogenic zone.

On account of computational efficiency, we just investigated bivariate signals in this study and TE was only employed to identify flow direction using pairs of signals. We are aware that the problem of direct and indirect interactions may arise when considering multivariate signals [22]. We therefore plan to extend this work on TE for multivariate signals to solve this issue.

### REFERENCES

- [1] J. Engel, P. C. Van Ness, T. B. Rasmussen, and L. M. Ojemann, *Outcome with respect to epileptic seizures*. New York: Raven Press, 1993.
- [2] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, pp. 461-464, 2000.
- [3] S. Sabesan, K. Narayanan, A. Prasad, L. D. Iasemidis, A. Spanias, and K. Tsakalis, *Information flow in coupled nonlinear systems: application to the epileptic human brain*. New York: Springer, 2007.
- [4] S. Sabesan, L. Good, K. Tsakalis, A. Spanias, D. Treiman, and L. Iasemidis, "Information flow and application to epileptogenic focus localization from intracranial EEG," *IEEE Trans. on Neural Syst. Rehabil. Eng.: A Publication of the IEEE EMBS*, vol. 17, pp. 244-253, 2009.
- [5] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy - a model-free measure of effective connectivity for the neurosciences," *Journal of Computational Neuroscience*, vol. 30, pp. 45-67, 2011.
- [6] M. Wibral, B. Rahm, M. Rieder, M. Lindner, R. Vicente, and J. Kaiser, "Transfer entropy in magnetoencephalographic data: Quantifying information flow in cortical and cerebellar networks," *Progress in Biophysics and Molecular Biology*, vol. 105, pp. 80-97, 2011.



- [7] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, pp. 1134, 1986.
- [8] J. M. Martinerie, A. M. Albano, A. I. Mees, and P. E. Rapp, "Mutual information, strange attractors, and the optimal estimation of dimension," *Physical Review A*, vol. 45, pp. 7058-7064, 1992.
- [9] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, pp. 43-50, 1997.
- [10] H. Akaike, "Information theory and an extension of the maximum likelihood principle," presented at Proceedings of the Second International Symposium on Information Theory, Budapest, Hungary, 1973.
- [11] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, pp. 461-464, 1978.
- [12] R. W. Katz, "On some criteria for estimating the order of a Markov chain," *Technometrics*, vol. 23, pp. 243-249, 1981.
- [13] M. Ding, Y. Chen, and S. L. Bressler, *Granger causality: basic theory and application to neuroscience*. Berlin: Wiley-VCH, 2006.
- [14] J. W. H. Schnupp, T. M. Hall, R. F. Kokelaar, and B. Ahmed, "Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex," *Journal of Neuroscience*, vol. 26, pp. 4785-4795, 2006.
- [15] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, second ed: Springer Verlag, 1991.
- [16] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, pp. 36-47, 2004.
- [17] A. A. Neath and J. E. Cavanaugh, "Regression and time series model selection using variants of the Schwarz information criterion," *Communications in Statistics, Theory and Methods*, vol. 26, pp. 559-580, 1997.
- [18] H. Z. Shu, L. M. Luo, J. D. Zhou, and X. D. Bao, "Moment-based methods for polygonal approximation of digitized curves," *Pattern Recognition*, vol. 35, pp. 421-434, 2002.
- [19] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D: Nonlinear Phenomena*, vol. 9, pp. 189-208, 1983.
- [20] F. Wendling, A. Hernandez, J. J. Bellanger, P. Chauvel, and F. Bartolomei, "Interictal to ictal transition in human temporal lobe epilepsy: insights from a computational model of intracerebral EEG," *Journal of Clinical Neurophysiology*, vol. 22, pp. 343-356, 2005.
- [21] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Doynne Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D: Nonlinear Phenomena*, vol. 58, pp. 77-94, 1992.
- [22] C. Yang, R. Le Bouquin Jeannès, G. Faucon, and H. Shu, "Extracting information on flow direction in multivariate time series," *IEEE Signal Process. Lett.*, vol. 18, pp. 251-254, 2011.

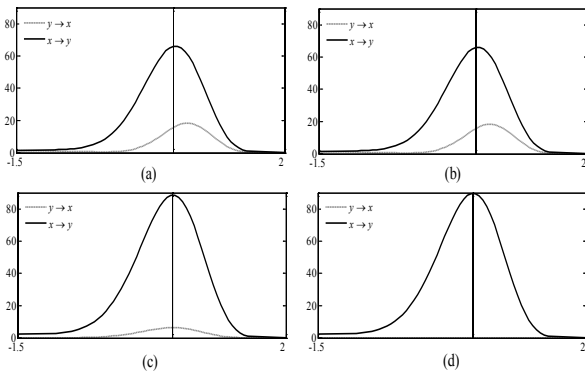


Fig. 1. Means of TE values in the first linear model. The x-axis is the value of radius  $r$  (in  $\ln$  scale), and the y-axis is the value of TE (bits). (a) The orders are estimated by AIC. (b) The orders are estimated by BIC. (c) The orders are estimated by gAIC. (d) The orders are estimated by gBIC. Solid line: flow direction  $x \rightarrow y$ , dotted line: flow direction  $y \rightarrow x$ .

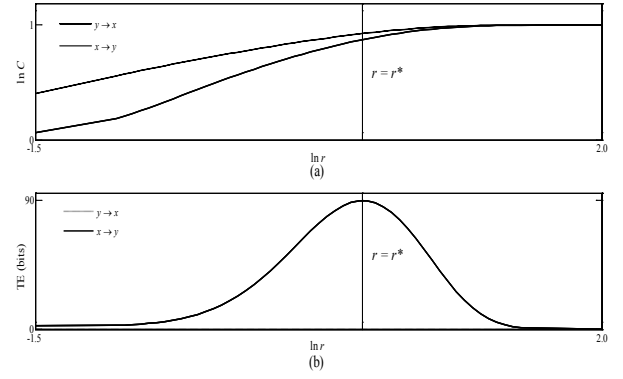


Fig. 2. Results on TE with the orders estimated by gBIC between signals  $x$  and  $y$  in the first linear model. (a) Plot of the mean of  $C_{n,r}$  with respect to  $n$  vs  $r$  (in  $\ln$  scale), where  $C$  and  $r$  denote average joint probability and radius in the state space respectively. (b) Plot of TE (bits) vs radius  $r$  (in  $\ln$  scale) (same as Fig. 1(d)). Solid line: flow direction  $x \rightarrow y$ , dotted line: flow direction  $y \rightarrow x$ .

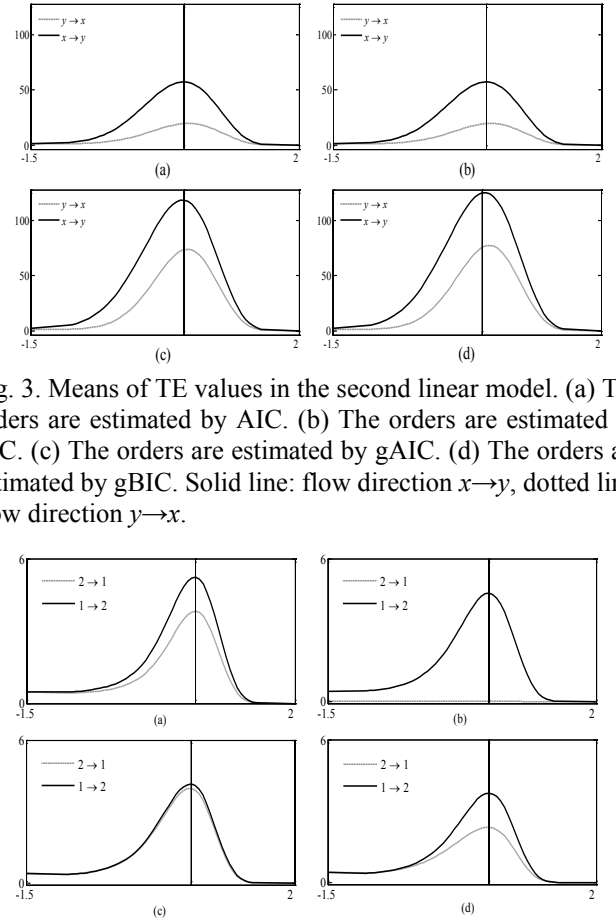


Fig. 3. Means of TE values in the second linear model. (a) The orders are estimated by AIC. (b) The orders are estimated by BIC. (c) The orders are estimated by gAIC. (d) The orders are estimated by gBIC. Solid line: flow direction  $x \rightarrow y$ , dotted line: flow direction  $y \rightarrow x$ .

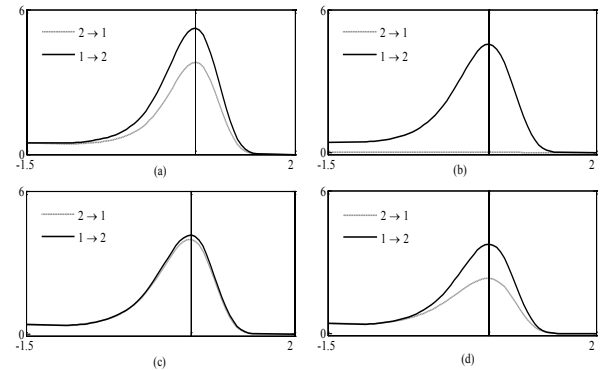


Fig. 4. Means of TE values in physiology-based models where the models' orders are estimated by BIC and gBIC. (a) Populations 1 and 2 are unidirectionally coupled and the orders are estimated by BIC. (b) Populations 1 and 2 are unidirectionally coupled and the orders are estimated by gBIC. (c) Populations 1 and 2 are bidirectionally coupled and the orders are estimated by BIC. (d) Populations 1 and 2 are bidirectionally coupled and the orders are estimated by gBIC.

Solid line: flow direction 1→2, dotted line: flow direction 2→1.

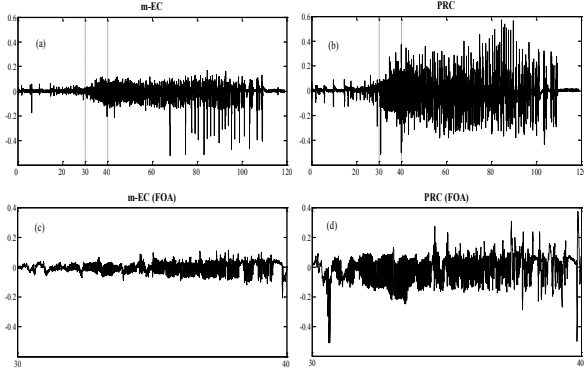


Fig.5. Typical pattern of interictal-to-ictal transition induced by bicuculline in the guinea pig brain. (a) in medial entorhinal cortex (m-EC), (b) in perirhinal cortex (PRC). (c) and (d): fast onset activity on the interval [30s,40s] extracted from m-EC and PRC, respectively.

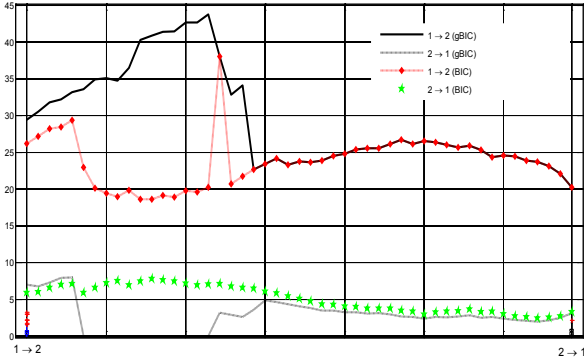


Fig. 6. Results on TE with the orders estimated by gBIC and BIC for 2 real iEEG signals. Signal 1 corresponds to a signal recorded in m-EC and signal 2 to a signal recorded in PRC.

TABLE I  
RESULTS ON THE ORDERS ESTIMATED BY THE 4 DIFFERENT MEASURES IN THE FIRST LINEAR MODEL

	$q_{yx}=2$	$q_{yx}=0$	$q_{yx}=3$	$q_{yx}=2$
AIC	0/200	0/200	200/200	0/200
BIC	0/200	0/200	200/200	0/200
gAIC	189/200	174/200	200/200	190/200
gBIC	200/200	200/200	200/200	200/200

TABLE II  
RESULTS ON TE AND PTE FOR THE FIRST LINEAR MODEL

	TE		PTE	
	$x \rightarrow y$	$y \rightarrow x$	$x \rightarrow y$	$y \rightarrow x$
AIC	65.1007 (2.9843)	14.3390 (2.1793)	0.8203 (0.0179)	0.1797 (0.0179)
BIC	65.1007 (2.9843)	14.3390 (2.1793)	0.8203 (0.0179)	0.1797 (0.0179)
gAIC	87.1233 (7.5735)	6.3522 (17.2288)	0.9542 (0.1225)	0.0458 (0.1225)
gBIC	88.4243 (4.9077)	0.0000 (0.0000)	1.0000 (0.0000)	0.0000 (0.0000)

TABLE III  
RESULTS ON THE ORDERS ESTIMATED BY THE 4 DIFFERENT MEASURES IN THE SECOND LINEAR MODEL

	$q_{yx}=2$	$q_{yx}=2$	$q_{yx}=3$	$q_{yx}=2$
AIC	0/200	0/200	200/200	0/200
BIC	0/200	0/200	200/200	0/200
gAIC	187/200	189/200	200/200	180/200
gBIC	200/200	200/200	200/200	200/200

TABLE IV  
RESULTS ON TE AND PTE FOR THE SECOND LINEAR MODEL

	TE		PTE	
	$x \rightarrow y$	$y \rightarrow x$	$x \rightarrow y$	$y \rightarrow x$
AIC	57.2852 (2.8155)	19.5017 (0.9358)	0.7459 (0.0097)	0.2541 (0.0097)
BIC	57.2852 (2.8155)	19.5017 (0.9358)	0.7459 (0.0097)	0.2541 (0.0097)
gAIC	118.7760 (22.1809)	73.6934 (19.0563)	0.6182 (0.0991)	0.3818 (0.0991)
gBIC	125.5434 (8.2244)	77.1255 (9.0390)	0.6205 (0.0147)	0.3795 (0.0147)

TABLE V  
RESULTS ON TE AND PTE IN PHYSIOLOGY-BASED MODELS

		TE		PTE	
		1→2	2→1	1→2	2→1
Unidirectional	BIC	5.0743 (2.6330)	3.6785 (2.3692)	0.5966 (0.1002)	0.4034 (0.1002)
	gBIC	5.5393 (3.0851)	0.0000 (0.0000)	1.0000 (0.0000)	0.0000 (0.0000)
Bidirectional	BIC	4.1299 (3.3139)	3.9693 (3.1104)	0.5017 (0.1444)	0.4983 (0.1444)
	gBIC	4.2004 (3.8541)	2.8840 (2.7001)	0.5535 (0.2447)	0.4465 (0.2447)

TABLE VI  
MANN-WHITNEY TEST ON TE IN PHYSIOLOGY-BASED MODELS

		1→2		2→1	
		$p$	$h$ , [expected value]	$p$	$h$ , [expected value]
Uni-dir.	BIC	8.01e-63	1, [1]	1.39e-56	1, [0]
	gBIC	2.74e-66	1, [1]	0.3197	0, [0]
Bi-dir.	BIC	6.35e-37	1, [1]	2.72e-34	1, [1]
	gBIC	3.41e-72	1, [1]	1.12e-65	1, [1]

## Annex A

When adding different regressors in the model (see equation (A.1)) or modifying the values of the weighted coefficients (equation (A.2)), or increasing the delay from one signal to the other one (equation (A.3)), the new gBIC criterion remained robust and allowed for a better estimation of the transfer entropy (Tables A.1 to A.6) compared to the standard criteria. In all of these cases, the unidirectional flow was detected with accuracy.

$$\begin{cases} x_t = 0.95\sqrt{2}x_{t-1} - 0.9025x_{t-2} + u_t \\ y_t = -0.35x_{t-2} + 0.5x_{t-3} - 0.4y_{t-2} + v_t \end{cases} \quad (\text{A.1})$$

TABLE A.1

RESULTS ON THE ORDERS ESTIMATED BY THE 4 DIFFERENT MEASURES IN EQUATION (A.1)

	$q_{xx}=2$	$q_{yx}=0$	$q_{xy}=3$	$q_{yy}=2$
AIC	0/200	0/200	200/200	0/200
BIC	0/200	0/200	200/200	0/200
gAIC	190/200	182/200	200/200	183/200
gBIC	200/200	200/200	200/200	199/200

TABLE A.2

RESULTS ON TE AND PTE FOR THE MODEL IN EQUATION (A.1)

	TE		PTE	
	$x \rightarrow y$	$y \rightarrow x$	$x \rightarrow y$	$y \rightarrow x$
AIC	34.5670 (1.8184)	5.9612 (0.9828)	0.8536 (0.0176)	0.1464 (0.0176)
BIC	34.5670 (1.8184)	5.9612 (0.9828)	0.8536 (0.0176)	0.1464 (0.0176)
gAIC	54.6408 (5.1603)	2.0719 (5.6763)	0.9711 (0.0778)	0.0289 (0.0778)
gBIC	55.7624 (3.8396)	0.0000 (0.0000)	1.0000 (0.0000)	0.0000 (0.0000)

$$\begin{cases} x_t = 0.95\sqrt{2}x_{t-1} - 0.9025x_{t-2} + u_t \\ y_t = 0.7x_{t-3} - 0.7y_{t-2} + v_t \end{cases} \quad (\text{A.2})$$

TABLE A.3

RESULTS ON THE ORDERS ESTIMATED BY THE 4 DIFFERENT MEASURES IN EQUATION (A.2)

	$q_{xx}=2$	$q_{yx}=0$	$q_{xy}=3$	$q_{yy}=2$
AIC	0/200	0/200	200/200	0/200
BIC	0/200	0/200	200/200	0/200
gAIC	188/200	183/200	200/200	193/200
gBIC	200/200	200/200	200/200	200/200

TABLE A.4

RESULTS ON TE AND PTE FOR THE MODEL IN EQUATION (A.2)

	TE		PTE	
	$x \rightarrow y$	$y \rightarrow x$	$x \rightarrow y$	$y \rightarrow x$
AIC	62.9057 (2.6545)	19.4879 (2.1043)	0.7640 (0.0148)	0.2360 (0.0148)
BIC	62.9057 (2.6545)	19.4879 (2.1043)	0.7640 (0.0148)	0.2360 (0.0148)
gAIC	84.4181 (6.9006)	5.6985 (15.6086)	0.9565 (0.1154)	0.0435 (0.1154)
gBIC	85.9249 (4.4645)	0.0000 (0.0000)	1.0000 (0.0000)	0.0000 (0.0000)

$$\begin{cases} x_t = 0.95\sqrt{2}x_{t-1} - 0.9025x_{t-2} + u_t \\ y_t = 0.7x_{t-5} - 0.7y_{t-2} + v_t \end{cases} \quad (\text{A.3})$$

TABLE A.5

RESULTS ON THE ORDERS ESTIMATED BY THE 4 DIFFERENT MEASURES IN EQUATION (A.3)

	$q_{xx}=2$	$q_{yx}=0$	$q_{xy}=5$	$q_{yy}=2$
AIC	0/200	0/200	200/200	0/200
BIC	0/200	0/200	200/200	0/200
gAIC	188/200	183/200	200/200	186/200
gBIC	200/200	200/200	200/200	200/200

TABLE A.6

RESULTS ON TE AND PTE FOR THE MODEL IN EQUATION (A.3)

	TE		PTE	
	$x \rightarrow y$	$y \rightarrow x$	$x \rightarrow y$	$y \rightarrow x$
AIC	34.8387 (1.2853)	3.8929 (0.8332)	0.8999 (0.0192)	0.1001 (0.0192)
BIC	34.8387 (1.2853)	3.8929 (0.8332)	0.8999 (0.0192)	0.1001 (0.0192)
gAIC	76.6072 (8.1165)	0.6591 (3.3145)	0.9918 (0.0400)	0.0082 (0.0400)
gBIC	78.6284 (4.4936)	0.0000 (0.0000)	1.0000 (0.0000)	0.0000 (0.0000)

## Annex B

In the model described by (23), we extended the range of model orders by changing successively the values of one and only one of the four orders  $q_{xx}$ ,  $q_{yx}$ ,  $q_{xy}$  and  $q_{yy}$  in the range [1; 6]. As discussed in Section III.A, the true model orders were  $q_{xx} = 2$ ,  $q_{yx} = 0$ ,  $q_{xy} = 3$ , and  $q_{yy} = 2$ . For TE calculation, when  $q_{xx}$  or  $q_{yx}$  is modified, only  $TE_{Y \rightarrow X}$  can be theoretically impacted. In the same way, when  $q_{xy}$  or  $q_{yy}$  is modified, only  $TE_{X \rightarrow Y}$  can be influenced. Consequently, we represented in Table B the means and standard deviations (shown in parentheses) of the corresponding TE estimated values at the specific radius to compare them with results displayed in Table II. As can be seen from this table, the unidirectional flow from signal  $x$  to signal  $y$  was perfectly

detected whatever the value of  $q_{xx}$  as  $\widehat{TE}_{Y \rightarrow X}$  remained null. On the other hand, the transfer entropy from signal  $y$  to signal  $x$  was erroneous as soon as the order  $q_{yx}$  used for TE estimation differed from the true value,  $q_{yx} = 0$ , not reported in the table, and leading to  $\widehat{TE}_{Y \rightarrow X} = 0$ . Now, when analyzing the transfer entropy from signal  $x$  to signal  $y$  with a false value of  $q_{xy}$ , the value of  $\widehat{TE}_{X \rightarrow Y}$  was not as high as expected (all the more since the order was lower than the true one). In the same manner,  $\widehat{TE}_{X \rightarrow Y}$  decreased drastically when the order  $q_{yy}$  was larger than the true value. To summarize, these examples illustrate the benefits that can be expected when selecting correctly the orders.

TABLE B  
RESULTS ON TE WHEN  $r^* = e^{0.5}$

order	1	2	3	4	5	6
$\widehat{TE}_{Y \rightarrow X}$ , $q_{xx}$ in the range [1; 6]	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
$\widehat{TE}_{Y \rightarrow X}$ , $q_{yx}$ in the range [1; 6]	55.1835 (5.7597)	57.9085 (6.5732)	48.0994 (5.9302)	43.1869 (5.4855)	42.2638 (5.6660)	37.8527 (5.4543)
$\widehat{TE}_{X \rightarrow Y}$ , $q_{xy}$ in the range [1; 6]	15.3217 (2.1962)	67.2140 (4.5509)	90.4821 (4.8040)	86.8531 (4.7981)	82.5186 (4.8169)	81.7844 (5.2656)
$\widehat{TE}_{X \rightarrow Y}$ , $q_{yy}$ in the range [1; 6]	113.2201 (5.6794)	90.4821 (4.8040)	65.6312 (2.9434)	44.4543 (1.5272)	33.0680 (1.2150)	28.2080 (1.1906)