

# On the Application of the Auto Mutual Information Rate of Decrease to Biomedical Signals

Javier Escudero\*, *Student Member, IEEE*, Roberto Hornero, *Member, IEEE*, Daniel Abásolo, *Member, IEEE*, and Miguel López, *Member, IEEE*

**Abstract**—The auto mutual information function (*AMIF*) evaluates the signal predictability by assessing linear and non-linear dependencies between two measurements taken from a single time series. Furthermore, the *AMIF* rate of decrease (*AMIFRD*) is correlated with signal entropy. This metric has been used to analyze biomedical data, including cardiac and brain activity recordings. Hence, the *AMIFRD* can be a relevant parameter in the context of biomedical signal analysis. Thus, in this pilot study, we have analyzed a synthetic sequence (a Lorenz system) and real biosignals (electroencephalograms recorded with eyes open and closed) with the *AMIFRD*. We aimed at illustrating the application of this parameter to biomedical time series. Our results show that the *AMIFRD* can detect changes in the non-linear dynamics of a sequence and that it can distinguish different physiological conditions.

## I. INTRODUCTION

SEVERAL physiological recordings exhibit a partially non-linear behavior. For instance, neural processes are non-linear phenomena [1] and non-linearity also appears in the cardiovascular system [2], [3]. Information Theory can be useful to characterize signals considering non-linear features, since it offers a relevant approach to analyze time series regarded as information sources [4], [5].

The concept of mutual information (*MI*) was derived from the Information Theory to estimate the information obtained from observations of one random event on another [6]. *MI* measures both linear and non-linear dependences between two time series [6]. Hence, it is a non-linear counterpart to the classical correlation statistic [6], [7]. Similarly to this linear metric, *MI* can be applied to time-delayed versions of two different signals – cross mutual information function (*CMIF*) – or from the same sequence – auto mutual information function (*AMIF*) [2], [6].

*MI* has been previously applied to several in biomedical recordings [2], [3], [5]–[7]. Since the *CMIF* quantifies the statistical coupling between signals [6], [8], it has been used to analyze physiological systems where certain synchronization is expected [2], [9]. The *CMIFs* of respiratory and heart rate variability recordings provide

useful information to detect cardiac and respiratory diseases [2], [10]. The statistical dependences between different brain regions have also been assessed computing the *CMIF* from pairs of electroencephalogram (EEG) channels in several brain states [11], [12] and neurological disorders, like Alzheimer’s disease [13] or schizophrenia [14].

The *AMIF* measures the predictability of a time series [3], [13]. The first relative minimum of this function can be used to estimate the time lag needed to reconstruct the signal attractor in the phase space [15]. Additionally, the *AMIF* rate of decrease (*AMIFRD*) with increasing time delay is related to signal entropy [7], [16]. It has been shown that the decay of the *AMIF* provides relevant information about the underlying physiological systems [7]. Due to this ability to characterize diverse biomedical signals, the *AMIFRD* has been previously applied to study several patho-physiological conditions. For instance, the *AMIFRD* of EEG and magnetoencephalogram (MEG) recordings characterized patients with Alzheimer’s disease in contrast to control subjects, showing that this dementia produces a more predictable brain activity [13], [17]. This metric has also been successfully applied to the EEG in schizophrenic patients [14]. In addition, different *AMIF* decay parameters can help to detect several cardiomyopathies from cardiac data [2], [3], [7].

Due to the relevance and the possible usefulness of the *AMIFRD* in certain biomedical analyses, it is important to understand and exemplify the behavior of this parameter for various kinds of signals. It is worth mentioning that several studies have illustrated the behavior of *MI* for a few test sequences in contrast to the correlation function [2], [6], [8], [9], [15]. Moreover, the decay of the *AMIF* was studied and compared with other entropic measures in [7]. Following these research works, this preliminary study aims at illustrating the ability of the *AMIFRD* to distinguish different non-linear dynamics in a time series and at exemplifying the application of the *AMIFRD* to biomedical signals by analyzing real surface EEG data.

This paper is organized as follows. In Section II, we describe the computation of the *AMIF* and its derived parameter, the *AMIFRD*. Section III introduces the synthetic and real signals to which the *AMIFRD* was applied. Our results are presented in Section IV and, finally, Section V discusses our findings and draws the conclusions of this study.

Manuscript received April 4, 2008. This work was supported in part by “Ministerio de Educación y Ciencia” and FEDER grant MTM 2005-08519-C02-01 and by the grant project VA108A06 from “Consejería de Educación de la Junta de Castilla y León.” *Asterisk indicates corresponding author.*

J. Escudero\*, R. Hornero, D. Abásolo, and Miguel López are with the Biomedical Engineering Group, E.T.S.I. Telecomunicación, University of Valladolid, Camino del Cementerio s/n, 47011 - Valladolid, Spain (phone: +34 983 423000, ext 5589; fax: +34 983 423667; e-mail: javier.escudero@ieee.org).

## II. AUTO MUTUAL INFORMATION FUNCTION

The *AMIF*, similarly to the *MI*, is derived from Information Theory [4]. *MI* provides a measure of both the linear and non-linear statistical dependencies between two time series [6].

The *AMIF* is based on the *MI* between two measurements taken from a single time series,  $X(t)$ , separated by a time index,  $\tau$  [7]. Let  $X(t) = \{x(1), x(2), \dots, x(N)\}$  be a sequence acquired from an ergodic stochastic process [10]. If the amplitude values of the variable  $X(t)$  are partitioned into  $I$  bins, a probability  $p_i^{X(t)} = n_i^{X(t)}/N$  can be assigned to each possible partition  $X_i (i = 1, \dots, I)$ , where  $n_i^{X(t)}$  is the number of samples in every bin [8], [17]. Similarly,  $p_j^{X(t+\tau)}$  denotes the probability derived from a time-delayed version of  $X(t)$ :  $X(t+\tau)$ . Finally,  $p_{ij}^{X(t)X(t+\tau)}$  is the joint probability for the measurements  $X(t)$  and  $X(t+\tau)$ , in that order. Then, the auto mutual information function for each  $\tau$ ,  $AMIF(\tau)$ , is computed as [13], [17]:

$$AMIF(\tau) = \sum_{i,j} p_{ij}^{X(t)X(t+\tau)} \log_2 \left( \frac{p_{ij}^{X(t)X(t+\tau)}}{p_i^{X(t)} p_j^{X(t+\tau)}} \right). \quad (1)$$

It is necessary to construct the amplitude distributions of  $X(t)$  for different time delays,  $\tau$  [8]. These distributions are estimated from histograms [10], [13], [17]. For a fixed sequence length, smaller partitions may enhance changes in the joint probability distribution over short distances, but they produce fluctuations due to the small sample size. Therefore, the *MI* may be overestimated. On the other hand, larger bins estimate the probabilities more accurately. Nevertheless, the joint probability distribution could be too flat and the *MI* may be underestimated [13]. Considering the window length used in this study (4500 data points for the synthetic signals and 4096 samples for the EEG recordings), we have employed 64 bins to build the histograms and estimate the distributions. This value was used in previous studies with similar epoch lengths [12]–[14], [17], since it provided stable estimations of the *MI*. The procedure to construct the histograms is detailed in [17].

The  $AMIF(\tau)$  was estimated over a time interval ranging from  $\tau = 0$  to  $\tau = 0.5$  s and it was normalized so that  $AMIF(\tau = 0) = 1$  [13], [14], [17]. The *AMIFRD* between  $\tau = 0$  and the first relative minimum value of the *AMIFRD* was calculated using a first-order least-squares fitting method [13], [17], so that different time scales were simultaneously taken into account when analyzing the signals. The *AMIFRD* measures the information loss versus  $\tau$  [2] and it is correlated with signal entropy [16].

## III. APPLICATION OF THE *AMIFRD* TO SYNTHETIC SEQUENCES AND REAL BIOMEDICAL SIGNALS

### A. Application of the *AMIFRD* to a Non-linear Synthetic Time Series

This section describes the simulated time series (a Lorenz system) used to exemplify the application of the *AMIFRD* and its ability to distinguish different kinds of signal dynamics. The test signal had a length of 150 s and its sampling frequency ( $f_s$ ) was 150 Hz (22500 sample points). For this signal, the  $AMIF(\tau)$  was computed using a moving window of 30 s (4500 data samples) with 90% overlap. The *AMIFRD* was estimated within each window.

This test is based on a non-linear system like the Lorenz attractor, which is given by:

$$\left. \begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= x(\rho - z) - y \\ \dot{z} &= xy - \beta z \end{aligned} \right\} \quad (2)$$

where  $\sigma$ ,  $\beta$ , and  $\rho$  are the system parameters [5]. The first segment of this synthetic signal had a length of 11250 samples and it was generated with  $\sigma = 10$ ,  $\beta = 8/3$ , and  $\rho = 28$ . Thus, it exhibited a chaotic behavior. The second segment also had 11250 points and was produced with  $\sigma = 10$ ,  $\beta = 8/3$ , and  $\rho = 99.96$ , which produced a torus knot [5]. Both parts were normalized so that their standard deviation was equal to 1. Fig. 1 displays the coordinate  $x$ , which was the time series analyzed in this study.

For this synthetic signal, 250 independent realizations of the time series were created with different random seeds and the *AMIFRD* was averaged for each window.

### B. Distribution of the *AMIFRD* for White Noises

We also studied the variability of the *AMIFRD* for two different kinds of white noise. We generated 10000 independent realizations of Gaussian white noise and uniform white noise. These time series had a length of 30 s with  $f_s = 150$  Hz (4500 sample points).

### C. Application to Real Biomedical Recordings

Finally, the *AMIFRD* was applied to two groups of real surface EEG signals in order to illustrate the ability of this parameter to distinguish different physiological states. These recordings belong to the EEG database made available online by Dr. Andrzejak of the Department of Epileptology at the University of Bonn [1]. The analyzed signals correspond to the datasets A and B of that database. Each dataset contains 100 single-channel surface EEG signals of 23.6 s recorded with  $f_s = 173.61$  Hz (4096 sample points). All signals were recorded from five healthy volunteers who were relaxed in an awake state. The subjects had their eyes open during the recording of the EEG in dataset A, whereas the EEG signals belonging to dataset B were acquired with eyes closed [1]. Additional details can be found in [1]. Prior to the computation of the *AMIFRD*, all EEG epochs were digitally filtered using a band-pass filter with cut-off

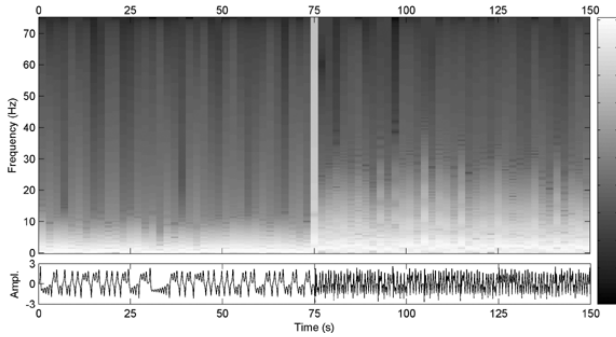


Fig. 1. Example of the spectrogram and time plot of the synthetic signal used to test the *AMIFRD*: Lorenz system with two different types of non-linear dynamics.

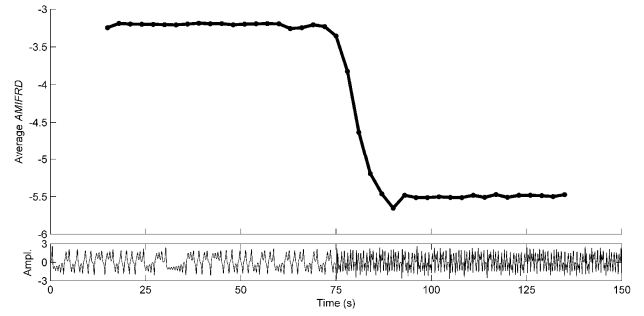


Fig. 2. Average results of the tests performed on the simulated signals showing the *AMIFRD* values versus two different types of non-linear dynamics.

frequencies at 0.5 Hz and 40 Hz.

#### IV. RESULTS

This preliminary study was performed to illustrate the application of the *AMIFRD* to different types of time series, with special focus on biomedical recordings.

By analyzing the Lorenz system, we wanted to determine whether the *AMIFRD* detects changes in the non-linear parameters that govern this system. Fig. 2 depicts the results. The *AMIFRD* values show a sudden change between both kinds of non-linear dynamics.

Fig. 3 illustrates the boxplots of the *AMIFRD* for Gaussian and uniform white noises. The distributions are relatively similar. These noises are characterized by very steep declines of the *AMIF* and both boxplots show a tail toward less negative values. It is remarkable that, although the *MI* is based on estimating histograms, the computations of the *AMIFRD* for these two noises with different probability distributions have a similar behavior due to the fact that the *AMIF* of white noise is a delta function [2].

Fig. 4 shows the boxplots of the *AMIFRD* for real surface EEG signals acquired with eyes open and closed. The closed-eyes condition modifies the EEG spectrum by increasing the power of the alpha rhythm (oscillations between 8 Hz and 13 Hz) [1]. Fig. 4 suggests that the closing of eyes is associated with less predictable EEG signals, as it can be inferred from the more negative *AMIFRD* values related to the eyes-closed state.

#### V. DISCUSSION AND CONCLUSIONS

The relationships between *CMIF* and other coupling measures have been previously illustrated [2], [6], [8], [9]. Additionally, the computation of the *AMIF* decay can help to identify several physiological states [2], [3], [7], [13], [14], [17]. Due to the relevance of the *AMIFRD* in the context of biomedical signal analysis, this preliminary study aimed at illustrating the application of this parameter in the analysis of biomedical recordings. Our simulations showed that the *AMIFRD* changes with the non-linear dynamics of a signal. Therefore, the application of this non-linear may provide useful information when certain non-linearity is expected in

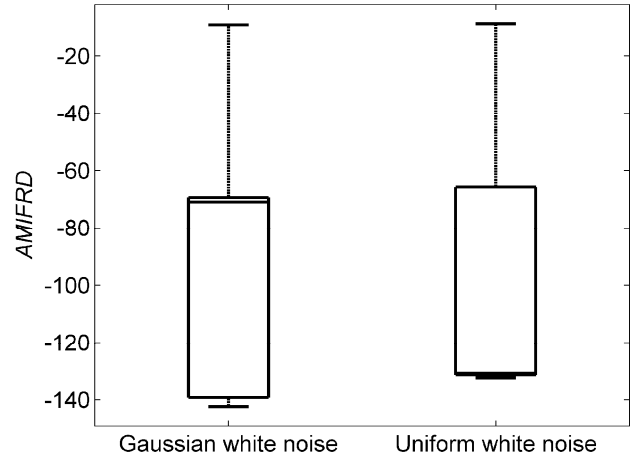


Fig. 3. Distributions of the *AMIFRD* for Gaussian white noise and uniform white noise.

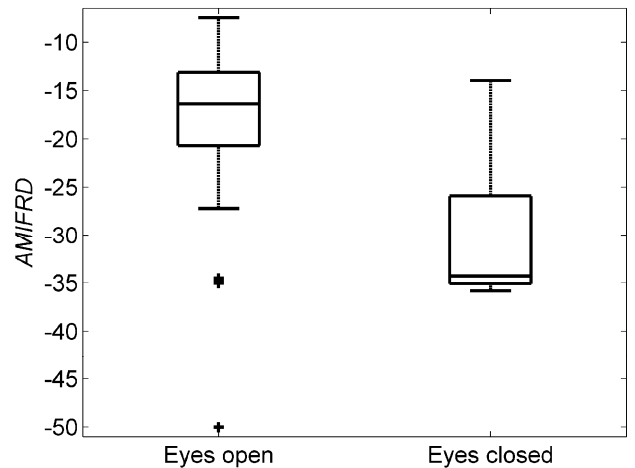


Fig. 4. Distributions of the *AMIFRD* for real surface EEG signals acquired with eyes open and eyes closed.

the underlying physiological system. We also illustrate that Gaussian and uniform white noises produce relatively similar distributions of the *AMIFRD* and that this statistic can differentiate different physiological states such as those produced by the closing of eyes in EEG signals.

A first-order least-squares fitting method was used to

compute the *AMIFRD* [13], [17]. Thus, we could assess the information loss on several time scales simultaneously. Diverse research works have suggested that studying several time scales when analyzing biomedical data provides an advantage in comparison to the use of other non-linear measures based on one time scale only [3], [7].

The *AMIFRD* has been applied to analyze biomedical recordings in several diseases, including Alzheimer's disease [13], [17], schizophrenia [14], myocardial infarction [2] and idiopathic dilated cardiomyopathy [3]. The use of this non-linear analysis method in the context of biomedical signal analysis has some advantages. Firstly, it can be applied to short biomedical time series in comparison to other non-linear analysis methods, like the correlation dimension [13], [17]. Secondly, it does not necessarily involve an embedding process, thus requiring shorter computation times. Moreover, the only input parameter for the *AMIF* is the number of histogram partitions [13], [14]. Finally, *MI* is invariant under strictly monotone transformations of the input sequences [6].

In contrast to other studies based on the computation of *MI* [7], [9], [11], signals were not embedded in a phase space before computing this statistic. The embedding process should be carried out in order to consider higher dimensional relationships [6], [8]. However, it would require longer recordings, something that is not always possible for biological data, and it would increase the computation time. Additionally, some advanced strategies can be applied to estimate *MI*, like a ranking transformation [6]. Nevertheless, we found that the straightforward computation of the histograms using 64 equal bins provided stable estimations of the *AMIF*.

To sum up, we performed a series of analyses to illustrate the application of the *AMIFRD* to the signal processing of biomedical recordings. The results suggested that this parameter detects changes in the non-linear dynamics of a system. Additionally, Gaussian and uniform white noises have relatively similar *AMIFRD* distributions. Finally, the analysis of EEG recordings showed that the *AMIFRD* can characterize different physiological conditions. Nevertheless, additional tests should be performed to properly understand and exemplify the properties and application of this and other non-linear analysis methods.

#### ACKNOWLEDGMENT

We are thankful to Dr. Andrzejak from the Department of Epileptology at the University of Bonn, Germany, for providing the EEG recordings described in Section III-C.

#### REFERENCES

[1] R. G. Andrzejak, K. Lehnertz, F. Moormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E*, vol. 64, no. 6, p. 061907, Dec. 2001.

[2] D. Hoyer, U. Leder, H. Hoyer, B. Pompe, M. Sommer, and U. Zwiener, "Mutual information and phase dependencies: measures of reduced nonlinear cardiorespiratory interactions after myocardial infarction," *Med. Eng. Phys.*, vol. 24, no. 1, pp. 33–43, Jan. 2002.

[3] M. Palacios, H. Friedrich, C. Götz, M. Vallverdú, A. Bayes de Luna, P. Caminal, and D. Hoyer, "Changes of autonomic information flow due to idiopathic dilated cardiomyopathy," *Physiol. Meas.*, vol. 28, no. 6, pp. 677–688, Jun. 2007.

[4] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press, 1949.

[5] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge: Cambridge University Press, 1997.

[6] B. Pompe, P. Blihdh, D. Hoyer, and M. Eiselt, "Using mutual information to measure coupling in the cardiorespiratory system," *IEEE Eng. Med. Biol. Mag.*, vol. 17, no. 6, pp. 32–39, Nov.–Dec. 1998.

[7] D. Hoyer, B. Pompe, K. H. Chon, H. Hardraht, C. Wicher, and U. Zwiener, "Mutual information function assess autonomic information flow of heart rate dynamics at different time scales," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 4, pp. 584–592, Apr. 2005.

[8] O. David, D. Cosmelli, and K. J. Friston, "Evaluation of different measures of functional connectivity using a neural mass model," *Neuroimage*, vol. 21, no. 2, pp. 659–673, Feb. 2004.

[9] R. Quian Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger, "Performance of different synchronization measures in real data: A case study on electroencephalographic signals," *Phys. Rev. E*, vol. 65, no. 4, p. 041903, Mar. 2002.

[10] J. F. Alonso, M. A. Mañanas, D. Hoyer, Z. L. Topor, and E. N. Bruce, "Evaluation of respiratory muscles activity by means of mutual information function at different levels of ventilatory effort," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 9, pp. 1573–1582, Sep. 2007.

[11] J. Xu, Z. Liu, R. Liu, and Q. Yang, "Information transmission in human cerebral cortex," *Physica D*, vol. 106, no. 3–4, pp. 363–374, Aug. 1997.

[12] B.-C. Min, S.-H. Jin, I.-H. Kang, D. H. Lee, J. K. Kang, S. T. Lee, and K. Sakamoto, "Analysis of mutual information content for EEG responses to odor stimulation for subjects classified by occupation," *Chem. Senses*, vol. 28, no. 9, pp. 741–749, Nov. 2003.

[13] J. Jeong, J. C. Gore, and B. S. Peterson, "Mutual information analysis of the EEG in patients with Alzheimer's disease," *Clin. Neurophysiol.*, vol. 112, no. 5, pp. 827–835, May 2001.

[14] S. H. Na, S.-H. Jin, S. Y. Kim, and B.-J. Ham, "EEG in schizophrenic patients: mutual information analysis," *Clin. Neurophysiol.*, vol. 113, no. 12, pp. 1954–1960, Dec. 2002.

[15] M. Paluš, V. Albrecht, and I. Dvořák, "Information theoretic test for nonlinearity in time series," *Phys. Lett. A*, vol. 175, no. 3–4, pp. 203–209, Apr. 1993.

[16] M. Paluš, "Coarse-grained entropy rates for characterization of complex time series," *Physica D*, vol. 93, no. 1–2, pp. 67–77, May 1996.

[17] C. Gómez, R. Hornero, D. Abásolo, A. Fernández, and J. Escudero, "Analysis of the magnetoencephalogram background activity in Alzheimer's disease patients with auto mutual information," *Comput. Meth. Programs Biomed.*, vol. 87, no. 3, pp. 239–247, Sep. 2007.