

Article

Differentiating Interictal and Ictal States in Childhood Absence Epilepsy through Permutation Rényi Entropy

Nadia Mammone ^{1,*}, Jonas Duun-Henriksen ², Troels W. Kjaer ³ and Francesco C. Morabito ⁴

¹ IRCCS Centro Neurolesi Bonino-Pulejo, Via Palermo c/da Casazza, SS. 113, Messina, Italy

² HypoSafe A/S, Diplomvej 381, 2800 Kgs. Lyngby, Denmark; E-Mail: jh@hyposafe.com

³ Neurophysiology Center, Department of Neurology, Roskilde University Hospital, Koegevej 7-13, DK-4000 Roskilde, Denmark; E-Mail: neurology@dadlnet.dk

⁴ DICEAM Department, Mediterranean University of Reggio Calabria, Via Graziella Feo di Vito, 89060 Reggio Calabria, Italy; E-Mail: morabito@unirc.it

* Author to whom correspondence should be addressed; E-Mail: nadiamammone@tiscali.it; Tel.: +39-96-5875224; Fax: +39-96-5875220.

Academic Editor: Wassim M. Haddad

Received: 31 March 2015 / Accepted: 25 June 2015 / Published: 2 July 2015

Abstract: Permutation entropy (PE) has been widely exploited to measure the complexity of the electroencephalogram (EEG), especially when complexity is linked to diagnostic information embedded in the EEG. Recently, the authors proposed a spatial-temporal analysis of the EEG recordings of absence epilepsy patients based on PE. The goal here is to improve the ability of PE in discriminating interictal states from ictal states in absence seizure EEG. For this purpose, a parametrical definition of permutation entropy is introduced here in the field of epileptic EEG analysis: the permutation Rényi entropy (PER). PER has been extensively tested against PE by tuning the involved parameters (order, delay time and alpha). The achieved results demonstrate that PER outperforms PE, as there is a statistically-significant, wider gap between the PER levels during the interictal states and PER levels observed in the ictal states compared to PE. PER also outperformed PE as the input to a classifier aimed at discriminating interictal from ictal states.

Keywords: permutation entropy; electroencephalogram; Rényi's entropy; absence seizures

1. Introduction

Electroencephalography (EEG) is an essential tool for the diagnosis of epilepsy and many other neurological diseases. It generally consists of a non-invasive recording of brain electrical activity by means of electrodes integrated into a cap, which is worn by the patient and connected to the acquisition system through a certain number of wires. Once the EEG has been recorded (which can last for minutes to days), a careful review of the entire recording is needed, in order to detect the presence of critical events, such as epileptic seizures or interictal spikes.

In this paper, we will focus in particular on childhood absence epilepsy (CAE), which is a common idiopathic generalized epilepsy syndrome [1,2]. The pediatric neurologist's main objective is to neutralize all absence seizures through appropriate medication, while inducing a few side effects. CAE affects the characteristics of the EEG and, in particular, its randomness. This is the reason why permutation entropy (PE) was introduced as a possible mathematical tool to process the EEG in order to extract diagnostic information by many researchers. PE was introduced by Bandt and Pompe in 2002 [3] as a very fast and robust algorithm to detect dynamic complexity changes in real-world time series. Subsequently, many authors showed that PE can be successfully used to analyze epileptic EEG. Cao *et al.* [4] proposed the use of PE to detect qualitative and quantitative dynamical changes in EEG data from epileptic patients. Li *et al.* [5] carried out a predictability analysis of absence seizures with PE. They tested the quality of PE as a tool to predict the absence seizures in Genetic Absence Epilepsy Rats from Strasbourg (GAERS). In their work, PE was shown to outperform sample entropy (SampEn) in absence seizures prediction, and the authors concluded that the use of PE may shed new light on the mechanism of absence seizures. Bruzzo *et al.* [6] applied PE to detect vigilance changes and preictal states from scalp EEG in epileptic patients. They evaluated the separability of amplitude distributions of PE resulting from preictal and interictal phases. A good separability of the interictal and preictal phases was found, and PE was shown to be sensitive to changes in vigilance state. Zanin *et al.* [7] reviewed the theoretical foundations of the PE, as well as the main applications to the understanding of biomedical systems. Nicolau *et al.* [8] investigated the use of PE as a feature for automated epileptic seizure detection. Ouyang *et al.* [9] investigated by multiscale PE the transition of brain activities towards an absence seizure. Zhu *et al.* [10] demonstrated, by experimental results, that PE is able to identify epileptic seizures in EEG and intracranial EEG (iEEG). The same authors proposed an unsupervised multi-scale K-means (MSK-means) algorithm to distinguish epileptic EEG signals and to identify epileptic zones [11]. Mateos *et al.* [12] developed a method based on PE to characterize EEG from different stages in the treatment of a chronic epileptic patient. Li *et al.* [13] carried out statistical experiments to explore the utility of using relevance feedback on EEG data to distinguish between different activity states in human absence epilepsy. Yang *et al.* [14] analyzed the EEG of children with electrical status epilepticus during slow-wave sleep (ESES) syndrome and of control subjects. The complexity of background EEG was evaluated using PE and SampEn in combination with the ANOVA test. They could observe that the entropy measures of EEG were significantly different between the ESES patients and normal control subjects.

As regards other applications of PE to the EEG in the medical field, evaluating the effects of anesthetic drugs is another key topic. Silva *et al.* [15] studied the performance of EEG-derived parameters in

the prediction of depth of anesthesia in a rabbit model comparing the index of consciousness (IoC) to the approximate entropy (AE) and the PE. Li *et al.* [16] introduced multiscale PE analysis of EEG recordings during sevoflurane anesthesia comparing it to CMSPE (composite multi-scale permutation entropy). They showed that CMSPE outperforms the raw single-scale PE in reflecting the sevoflurane drug effect on the central nervous system. Anier *et al.* [17] studied the entropy of the EEG in the transition from continuous EEG to burst suppression in deep anesthesia through surrogate data analysis. They showed that PE was the most sensitive to phase information. Olofsen *et al.* [18] tested PE of the EEG as a measure of anesthetic drug effect, introducing the composite permutation entropy index (CPEI), which basically tries to average the effects of the selection of parameters in the computation of PE. Li *et al.* [19] exploited PE of EEG signals as a measure of anesthetic drug effect. They set the embedding dimension and the time lag according to the quantitative performance, to quantify the effect of a wide range of concentrations of isoflurane on the EEG.

In recent years, the authors proposed a spatial-temporal analysis of epileptic EEG recordings based on measures of complexity of the EEG [20–24]. The EEG of patients affected by absence epilepsy showed a particular PE topography over the scalp, especially during seizures. This result was inferred by revising movies that showed the evolution of PE mapping [23].

The present work aims to improve the ability of PE in discriminating interictal states (seizure-free EEG segments) from ictal states (seizure EEG segments) in CAE. For this purpose, a parametrical definition of permutation entropy, the permutation Rényi entropy (PEr), which was introduced in [25,26], is proposed here for the analysis of epileptic EEG. In this paper, PEr has been tested against PE tuning the different involved parameters (embedding dimension, delay time and α) in order to be optimized for the analysis of absence epilepsy EEG.

The paper is organized as follows: Section 2 will introduce the methodology, providing details of the EEG recording, PEr definition and theoretical comparison *vs.* PE, as well as parameter optimization, with particular attention on α . Section 3 will illustrate the results, and Section 4 will address the conclusions.

2. Methodology

The flowchart of the procedure (Figure 1) can be itemized as follows: (1) the n -channel EEG is recorded and stored on a computer; (2) the EEG is partitioned into m overlapping windows and processed window by window; and (3) given a window under analysis, PE is estimated channel by channel, and the n values are arranged in an $n \times 1$ PE vector (once the EEG is fully processed, we will have an $n \times m$ PE matrix).

2.1. EEG Recording

In order to carry out a parameter optimization, 23 EEG recordings of patients diagnosed with CAE were acquired and processed. The children's mean age was 7.44 years, with a standard deviation of 1.67 years, and 16 of them were female. The EEG were acquired according to the international 10/20 system [27] with Cadwell Easy II (Cadwell Laboratories, Inc., Kennewick, WA) (21 patients) or Stellate Harmonie (Stellate Systems, Inc., Montreal, Quebec, Canada) (2 patients). All channels were filtered with a pass band of 0.53–25 Hz and digitized at a rate of 200 Hz. In total, 161 paroxysms longer than 2

seconds were identified by a board-certified clinical neurophysiologist, and in total, 590.68 min of data were available. All data analysis was performed using MATLAB R2001b (The MathWorks, Inc., Natick, MA, USA).

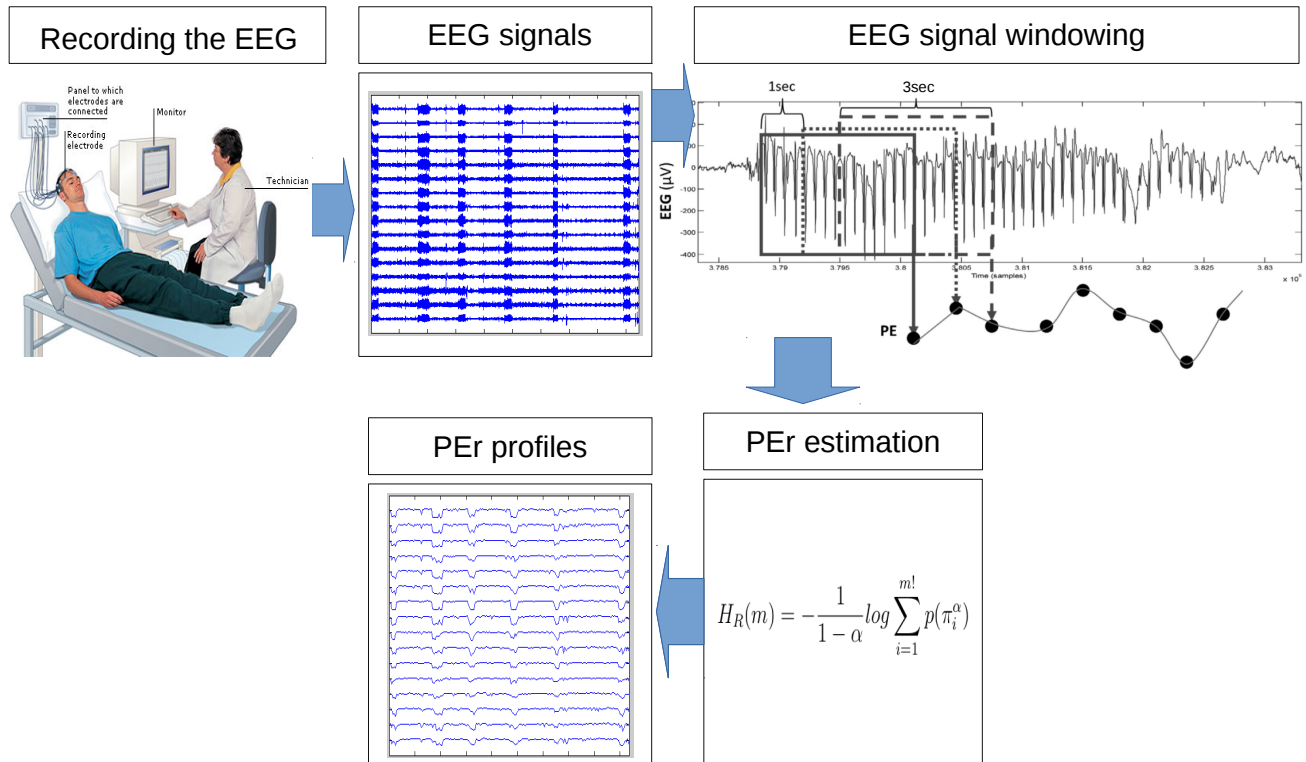


Figure 1. The flowchart of the procedure. (1) The n -channel EEG is recorded and stored on a computer; (2) the EEG is partitioned into m overlapping windows and processed window by window; (3) given a window under analysis, permutation entropy (PE) is estimated channel by channel, and the n values are arranged in an $n \times 1$ PE vector; (4) once the m PE vectors were estimated, the $n \times m$ matrix showing the PE trend of each channel can be displayed. PEr, permutation Rényi entropy.

2.2. Permutation Rényi Entropy

The n -channel EEG recording was processed by means of a sliding temporal window. The EEG samples were buffered in overlapping windows whose length was 3 s with a 2-s overlap (since the sampling rate is 200 Hz, one window includes $N = 600$ EEG samples). With a larger window, the PE profile would be smoother; nevertheless, the size was set at 3 s in order to detect even the shortest paroxysms (the paroxysms of clinical interest last for at least 2 s), so that they are not smoothed down. PE was estimated, channel by channel and window by window. Each EEG window includes n time series, where n is the number of EEG channels. This window is stored on the computer as an $n \times N$ matrix. Within each window, a sample of PE per channel is computed, and the n values are arranged in a $n \times 1$ PE vector; therefore, a $n \times N$ EEG matrix corresponds to an $n \times 1$ PE vector. In this way, a compressed temporal representation of the original time series is produced.

Going into the details of PE calculation, let us denote a given EEG time series as x . Each time series x was mapped into an m -dimensional space, with m being the embedding dimension and L being the time lag. Vectors X_t were constructed selecting m equally-spaced samples from x , starting from time point t :

$$X_t = [x(t), x(t+L), \dots, x(t+(m-1)L)]^T \quad (1)$$

The values of X_t are reshaped in an increasing order, and the time points are renamed yielding Xr_t , a reshaped version of X_t :

$$Xr_t = [x(t+(t_1-1)L), x(t+(t_2-1)L), \dots, x(t+(t_m-1)L)]^T \quad (2)$$

Therefore, each vector X_t can be considered uniquely mapped onto a symbol vector $\pi = [t_1, t_2, \dots, t_m]$. The vector π is a sequence of time points, hence a symbol. The frequency of occurrence of each possible π is indicated as $p(\pi)$, which represents the frequency of occurrence of the specific vector π in the time series under analysis, normalized by $N - (m-1)L$, where N is the number of samples of the time series x . PE is finally computed as:

$$H(m) = - \sum_{i=1}^{m!} p(\pi_i) \log(p(\pi_i)) \quad (3)$$

where \log is the natural logarithm and $m!$ is the number of the possible permutations. Since $H(m)$ can maximally reach $\ln(m!)$, PE is generally normalized as:

$$H_n(m) = - \frac{\sum_{i=1}^{m!} p(\pi_i) \log(p(\pi_i))}{\ln(m!)} \quad (4)$$

In this work, a new definition of permutation entropy, according to Rényi's theory, will be exploited [25]:

$$H_R(m) = - \frac{1}{1-\alpha} \log \sum_{i=1}^{m!} p(\pi_i)^\alpha \quad (5)$$

It is named permutation Rényi entropy (PER). A new parameter, α , is now involved, and the goal of the present work is to assess if this parameter can tune the sensitivity of PER in measuring EEG signal randomness.

Rényi's entropy is linked to the distribution of the time series through α . High α emphasizes the super-Gaussian distributions, whereas low-level α emphasizes the sub-Gaussian ones. Therefore, alpha can tune the sensitivity to the sub-Gaussianity, the super-Gaussianity of the pdf distribution. Higher alpha values ($\alpha > 4$) emphasize the super-Gaussianity, whereas average alpha values ($\alpha = 2-3$) emphasize both.

We tested the computational time over a simulated random time series (20000 samples); computing Equation (3) took 0.13 ms, whereas computing Equation (5) with $\alpha = 7$ took only 0.076 ms (Workstation: Intel Core 2 Quad CPU Q6600 @2.40GHzx4, 4GB RAM, with Ubuntu 14.04 LTSOS). This means that multiplying $p(\pi_i)$ by $\log(p(\pi_i))$ is more computationally intensive than calculating $p(\pi_i)^\alpha$, even for high α values, and that Rényi's PE calculation is faster than standard PE.

In particular, we aim to test the ability of PER, against PE, in discriminating interictal EEG from ictal EEG in absence epilepsy patients. For this purpose, we will compare the average PER of the interictal EEG segments and the average PER of the ictal EEG ones, against the same indices estimated with PE. First of all, we will carry out a theoretical comparison between PE and PER.

2.2.1. PEr vs. PE

In order to test the ability of PEr against PE in detecting dynamical changes in a time series, we first applied it to the transient logistic map. We generated a transient time series x , consisting of 1000 points, by starting from $x(0) = 0.5$ and $r = 3.5$, incrementing r up to $r = 4$, in steps of 0.01 at each iteration [3,4].

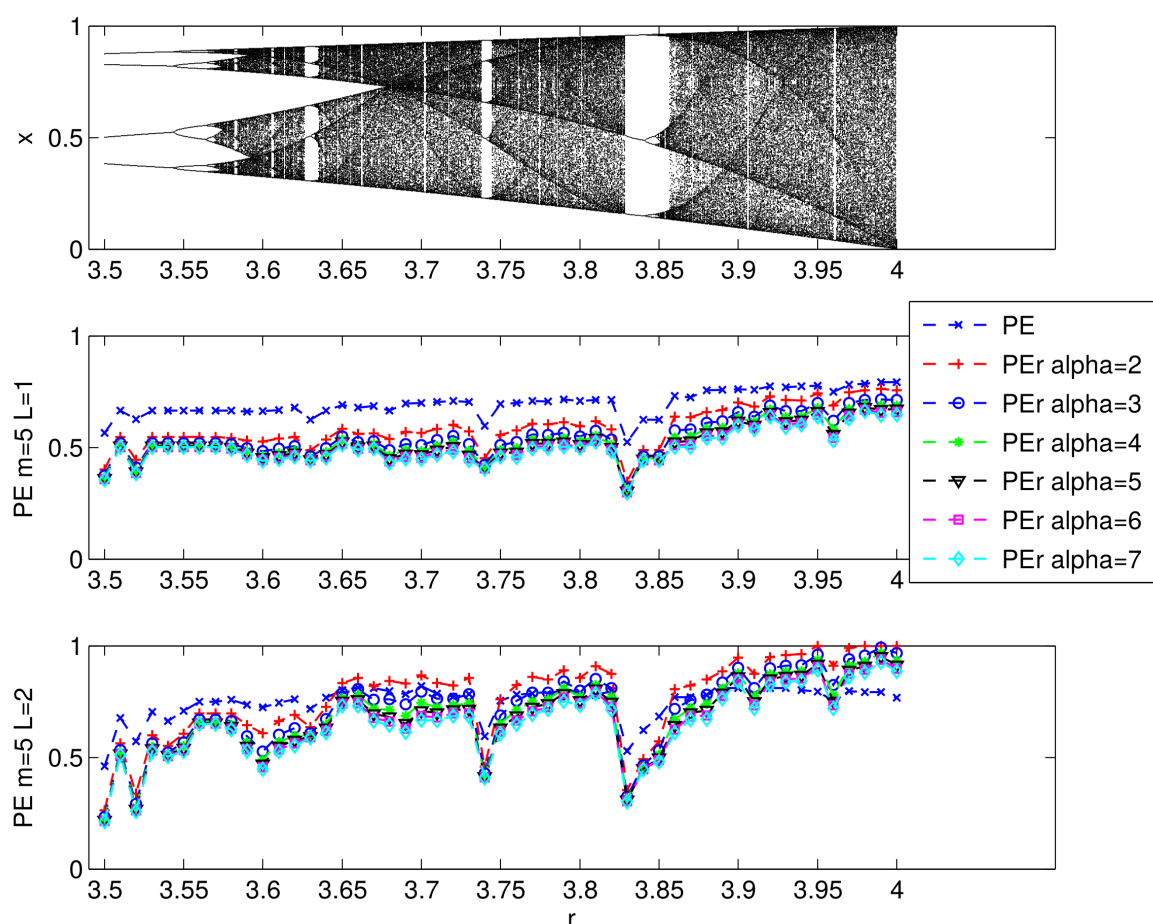


Figure 2. The transient logistic map data x (top), variations of PE and PEr with r for $m = 5$, $L = 1$ (middle), and $m = 5$, $L = 2$ (bottom).

Figure 2 shows the x time series and shows how well PE and PEr can detect dynamical changes from it. The variation of the permutation entropy with the parameter r captures the transitions better when $L = 2$ than $L = 1$ either for PE or PEr. It is noteworthy that the PEr *versus* r curve correlates with the original time series similarly to the PE *versus* r curve when $L = 1$, whereas it correlates better when $L = 2$.

2.2.2. Selecting the Optimal Parameter Configuration for PEr

Our aim was to find out what the optimal m - L - α configuration for PEr is and what the optimal m - L configuration for PE is, in order to optimize their ability to discriminate between interictal and ictal EEG. Several trials were carried out under different parameter settings:

- Order m ranging from 3–7;
- Lag L ranging from 1–10;
- α ranging from 2–7 (for PEr only).

Therefore, there are 50 possible parameter settings for PE and 300 for PEr.

Once the EEG recording is converted into PE and PEr matrices (see Section 2.2), it is partitioned into interictal and ictal sub-matrices. The two sub-matrices are then averaged with respect to the rows (channels) and the columns (PE sample vectors), so that an average interictal PE value ($avgPE(interictal)$) and an average ictal PE ($avgPE(ictal)$) value are estimated for the recording under analysis, in each specific parameter configuration. In order to compare the behavior of PE and PEr during the interictal and ictal states, we computed two indices:

- the difference $D = avgPE(interictal) - avgPE(ictal)$;
- the ratio $R = avgPE(interictal) / avgPE(ictal)$.

3. Results

The optimization process was carried out for every patient of the dataset described in Section 2.1. In order to clarify how the optimal parameter configuration was selected, we will now show the details of the procedure for one patient (Patient 4) and then summarize the results obtained for the remaining patients in Section 3.2, showing the optimal parameters selected patient by patient.

3.1. Analysis for Patient 4: Pilot Trial

3.1.1. Optimization for Patient 4

According to the procedure described in Section 2.2.2, PE and PEr are estimated in every possible parameter configuration. Each i -th parameter configuration provides a pair (D_i, R_i) that can be represented as a point (see Figure 3), and in conclusion, there are 50 points (blue circles in Figure 3) for PE and 300 for PEr (red circles in Figure 3).

First of all, we can observe that PEr (red circles) provided better results, because its (D_i, R_i) points lie in the right-upper part of the 2D plot, whereas PE's points lie in the lower-left part, thus showing smaller D and R . In order to identify the optimal parameter settings for PE and for PEr, the parameters that ensured the highest D_i and R_i were selected.

As regards Patient 4, the overall optimal configuration was $m = 4$, $L = 8$ (bold blue circle in Figure 3). As regards PEr, the configuration that provided the highest D was $m = 4$, $L = 8$, $\alpha = 7$, whereas the configuration that ensured the highest R was $m = 5$, $L = 5$ and $\alpha = 7$. In order to come up with a single overall optimal configuration also for PEr, we sought for the configuration that provided a relatively high performance with respect to both parameters R and D . We found that the configuration $m = 4$, $L = 7$ and $\alpha = 7$ ensured 96.5% of both the highest D and the highest R (red bold circle in Figure 3).

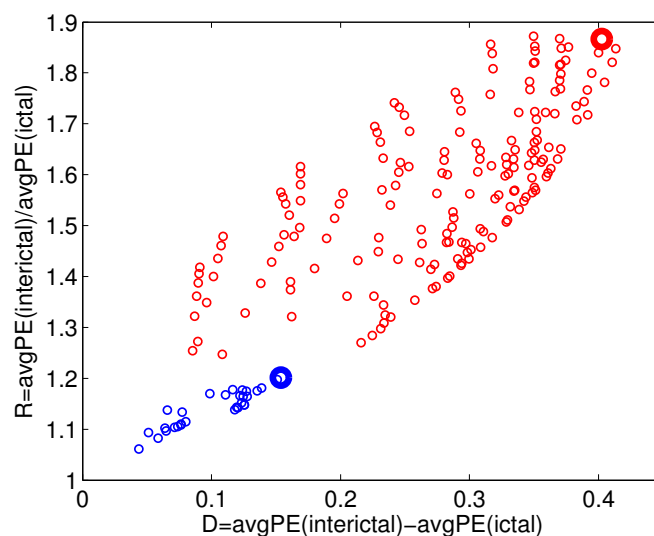


Figure 3. R vs. D in every possible parameter setting for either PE (blue circles) or PEr (red circles). The parameter setting that ensured the highest D_i and R_i for PE was $m = 4$, $L = 8$ (bold blue circle), whereas the parameters that ensured the best performance for PEr with respect to either D or R was $m = 4$, $L = 7$ and $\alpha = 7$ (red bold circle; the configuration ensured 96.5% of both the highest D and the highest R).

3.1.2. Evaluating the Effects of Alpha

In order to visually evaluate the effects of α on the behavior of PEr, we plotted the PE profile and the corresponding PEr profiles of a preictal EEG, see Figure 4. PE and PEr were estimated in the same $m = 4$ and $L = 8$ parameter configuration.

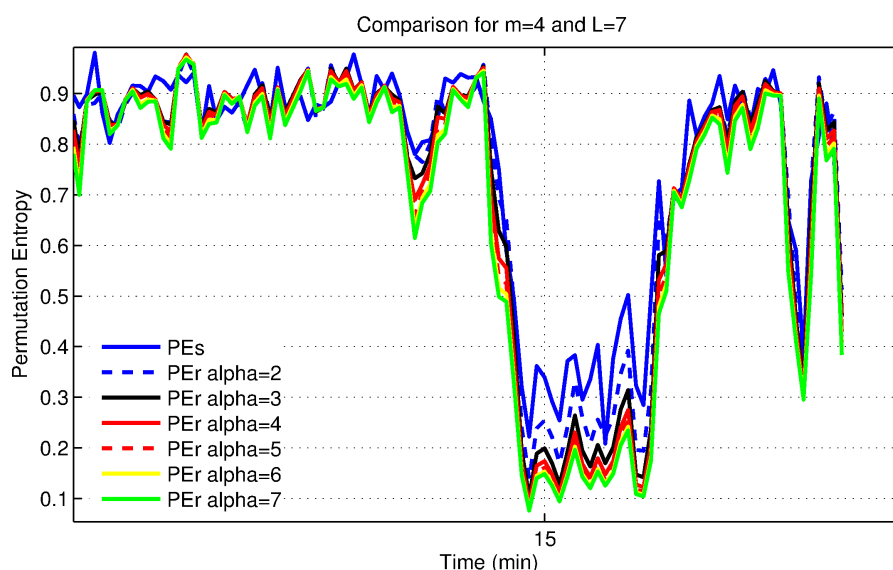


Figure 4. Effect of alpha on the behavior of PEr. The Figure shows the PE and PEr profiles of an EEG trace (from Channel Fp2), estimated in the same parameter configuration ($m = 4$ and $L = 8$, which is optimal for PE, but not for PEr). In the PEr estimation, α ranged from 2 to 7. The seizure onset is marked with a vertical dashed line.

In PEr estimation, α ranged from 2 to 7. The seizure onset is marked with a vertical dashed line. Even though the m - L configuration was optimal for PE, but not for PEr, we can observe that the gap between interictal levels and ictal levels is larger with PEr than with PE; thus, PEr appears more sensitive to the differences between interictal EEG and ictal EEG.

In order to give an overall view of the PEr profiles of the entire EEG recording, estimated for each channel, Figure 5 shows the PEr trend in the configuration $m = 4$, $L = 7$ and $\alpha = 7$ (seizure onsets are marked with vertical dashed lines).

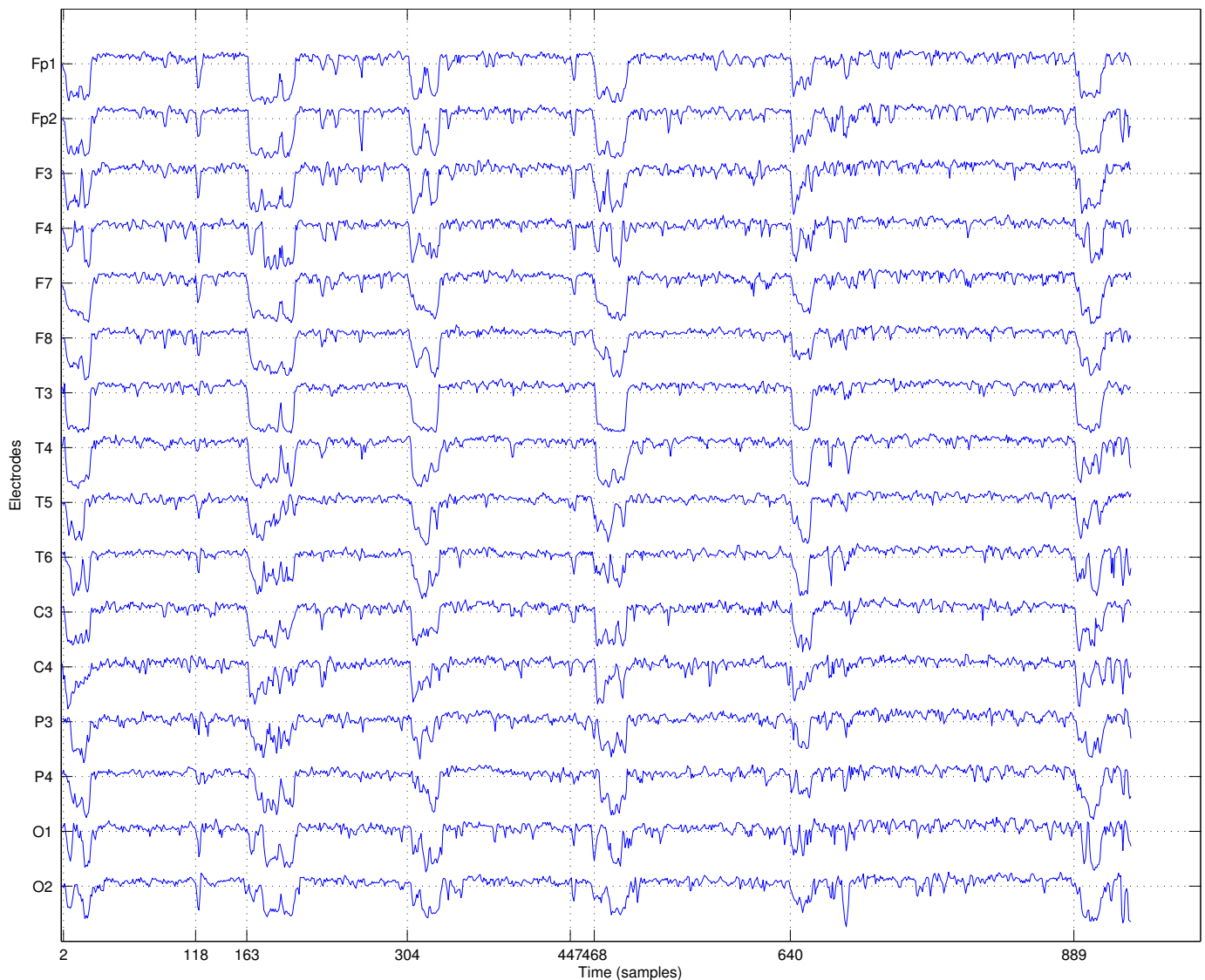


Figure 5. PEr profiles estimated for all of the channels, computed with $m = 4$, $L = 7$ and $\alpha = 7$.

3.2. Analysis of the Entire Dataset

The same optimization analysis was extended to the remaining 22 patients. Therefore, for every patient, 350 simulations were carried out (50 for PE estimation and 300 for PEr estimation, as described in Section 2.2.2), for a total of 8050 simulations. Table 1 summarizes the optimal parameters selected patient by patient. We can infer that the most recurrent optimal parameters for PE were $m = 4$ and

$L = 3$, whereas for PEr, $m = 5$ was the most recurrent embedding dimension, $L = 3$ was the most recurrent time delay and $\alpha = 7$ was the most recurrent alpha. As regards PEr, there was not an absolute overall most recurrent parameter configuration.

Table 1. Optimal PE and PEr parameter settings for each patient.

Pt	PE		PEr		
	m	L	m	L	α
1	4	3	5	4	6
2	4	2	5	3	3
3	4	10	4	7	7
4	4	8	4	7	7
5	3	10	3	5	7
6	4	3	5	5	7
7	4	2	5	3	7
8	4	10	5	4	7
9	4	3	5	6	7
10	4	3	5	5	6
11	4	3	4	3	5
12	4	3	5	5	5
13	4	8	4	7	7
14	4	3	5	4	7
15	4	2	4	3	4
16	4	3	5	4	6
17	5	3	4	6	7
18	4	2	4	2	2
19	4	3	5	4	6
20	4	3	4	3	4
21	4	3	5	5	6
22	4	3	5	3	4
23	4	3	4	7	7

For every EEG recording, both PE and PEr time series were estimated, according to the personalized parameter configuration, and stored as two matrices with n rows (where n is the number of channels). As described in Section 2.2.2, each entropy matrix was partitioned into two interictal and ictal sub-matrices. Then, the sub-matrices are averaged with respect to the rows and the columns in order to come up with an average interictal entropy value (**avgPE(interictal)**) and an average ictal entropy value (**avgPE(ictal)**), for each patient. The two values are used to compute the difference $\mathbf{D}_{PE} = \mathbf{avgPE(interictal)} - \mathbf{avgPE(ictal)}$ and the ratio $\mathbf{R}_{PE} = \mathbf{avgPE(interictal)}/\mathbf{avgPE(ictal)}$. In this way, we came up with a vector **avgPE(interictal)**, a vector **avgPE(ictal)**, a vector **D** and a vector **R**, each one with N elements, where N is the number of analyzed EEG recordings ($N = 23$).

In order to compare the ability of PEr and PE in discriminating interictal from ictal, the boxplots of either **R** or **D** vectors were plotted in Figure 6.

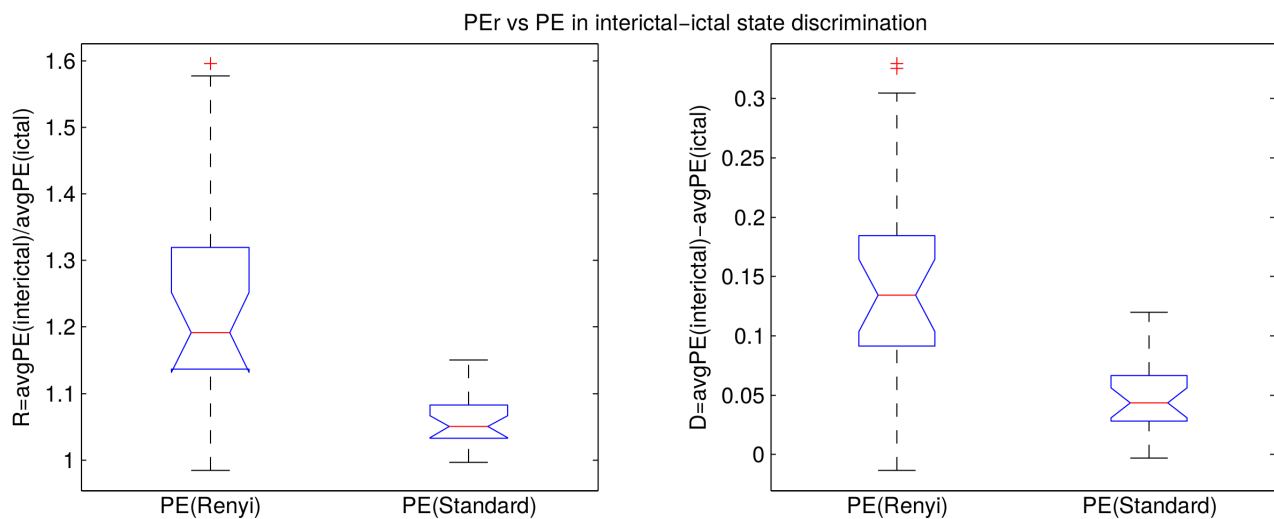


Figure 6. Boxplot of **R** and **D** vectors. On each box, the central mark is the median; the edges of the box are the 25th and 75th percentiles; the whiskers extend to the most extreme data points not considered as outliers.

PER showed a higher *R* and a higher *D*, therefore a better discrimination between interictal and ictal states, also using *R*. This result will be statistically tested in the Appendix (in Appendix A.2 and A.1, respectively).

Since PER was shown to outperform PE, we tested its general ability in discriminating interictal EEG from ictal EEG. The boxplots of the **avgPER(interictal)** and **avgPER(ictal)** vectors are plotted in Figure 7.

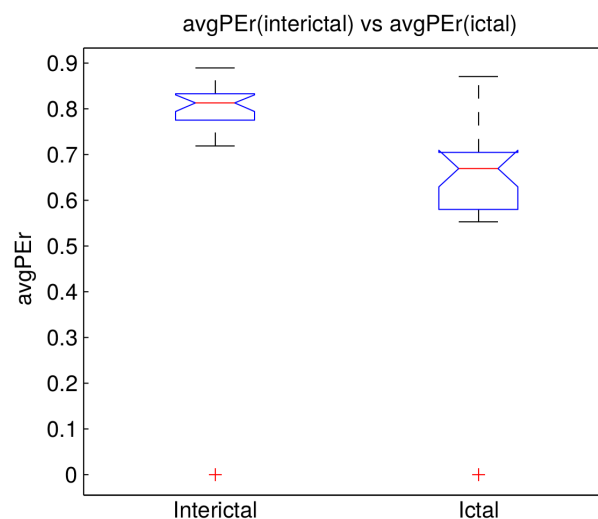


Figure 7. Boxplot of **avgPER(interictal)** and **avgPER(ictal)** vectors. On each box, the central mark is the median; the edges of the box are the 25th and 75th percentiles; the whiskers extend to the most extreme data points not considered as outliers.

On average, the PER extracted from interictal EEG was higher than the PER extracted from ictal EEG. This result will be statistically tested in the Appendix A.3.

Finally, in order to investigate what was the effect of the EEG channel on the discrimination ability of PEr for ictal/interictal segments, for every patient, we computed an average PEr profile for every cerebral area of interest (frontal, temporal, parietal, central, occipital). We averaged the PEr profiles of the EEG channels belonging to each specific area and then we plotted the boxplot of the **avgPEr(interictal)** and **avgPEr(ictal)** vectors for each specific area.

Looking at Figure 8, we can infer that no area looks significantly more sensitive to the transition from the interictal to the ictal state, from either the **R** or the **D** perspective.

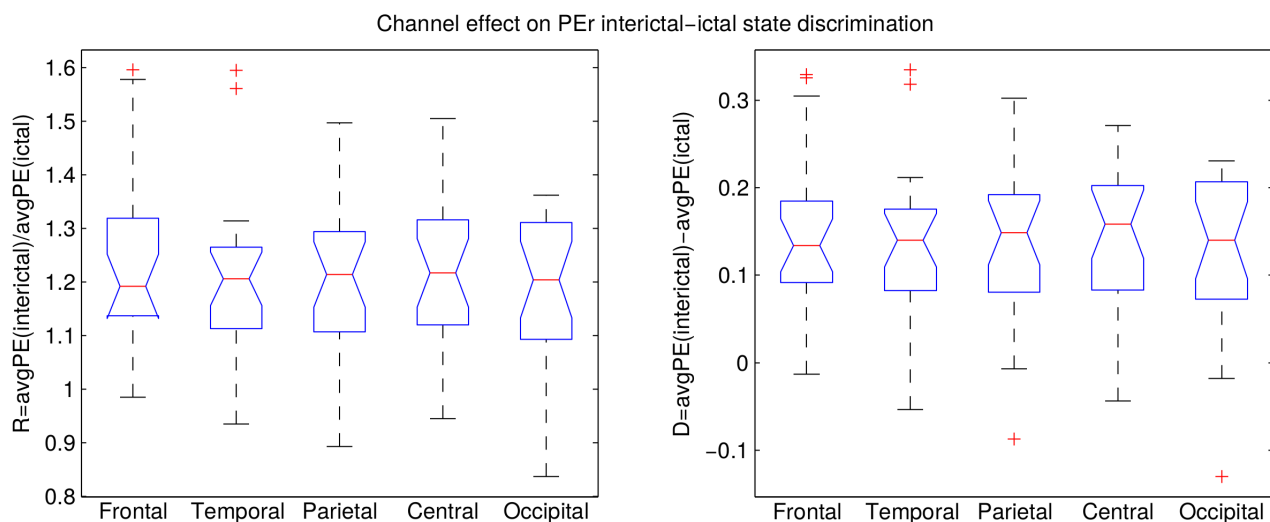


Figure 8. Investigation on the effect of the EEG channel on the discrimination ability of PEr for ictal/interictal segments. Given a patient, PEr profiles are averaged over every cerebral area of interest (frontal, temporal, parietal, central, occipital), then the elements of **R** vector are calculated as $\text{avgPEr}(\text{interictal})/\text{avgPEr}(\text{ictal})$. At the end of the process, we have one **R** vectors with n elements, where n is the number of patients, for every specific area. The boxplot of such **R** vectors is shown. On each box, the central mark is the median; the edges of the box are the 25th and 75th percentiles; the whiskers extend to the most extreme data points not considered as outliers.

3.3. Classifying the Brain States according to PE and PEr

In order to assess whether this increased gap might result in an improved classification of the cerebral state (interictal or ictal), we decided to test PE and PEr as possible input to a classifier. In recent years, the authors proposed a spatial-temporal analysis of epileptic EEG recordings based on measures of complexity of the EEG [20–24]. During the ictal state, the electrodes tended to cluster together according to their PE levels, exhibiting a particular PE topography. This peculiarity will be here exploited for classification purpose.

As here, the goal is to assess whether PEr overcomes PE in differentiating interictal and ictal states, we decided to associate the particular ictal PE topography with the ictal state. In this way, we will find out how the spatial distribution of the entropy extracted from the EEG matches the brain state. As this can be interpreted as a problem of clustering preliminary to supervised classification, we decided to classify

through learning vector quantization (LVQ) [28]. LVQ is a method for training competitive layers in a supervised manner. A competitive layer automatically learns to classify input vectors according to the distance between them. Given an entropy vector, the LVQ network can be trained to cluster the electrodes in a competitive way (so that the electrodes showing relatively similar entropy values are clustered together) and then to associate the vector to a specific target (the cerebral state: ictal or interictal).

In our experiment, once the PE and PEr profiles were estimated from the EEG (a $n \times 1$ entropy vector contains the n entropy values extracted from the n EEG channels within the specific EEG window under analysis), such entropy vectors were presented as input to the network, and the corresponding target was the cerebral state ictal (= 1) and interictal (= 0).

Seventy-five percent of ictal entropy vectors and of interictal entropy vectors were used for training, and the remaining 25% of both was used for testing. The LVQ network was trained with PE vectors and with PEr vectors separately. The LVQ network was trained with three hidden neurons, because, looking at the topographical distribution of PE values [23], three clusters of activity could be identified: high-, average- and low-PE. The learning function adopted was *lvql* [28].

Figure 9 shows the comparison of the accuracy provided by PE + LVQ and by PEr + LVQ. PEr + LVQ provided an improved or at least equal accuracy of classification in every patient. In particular, the accuracy of PEr resulted in being higher than the accuracy of PE in 12 out 23 patient and equal in 11 out of 23. Some patients (Patients 10, 12, 13) exhibited a very high accuracy (100%), whereas Patients 2, 6, 11, 15 and 20 exhibited an accuracy lower than 50%. The goal of the present analysis was to assess whether PEr outperformed PE in classifying the cerebral state. In the future, a classifier will be designed and trained through an appropriate and more sophisticated set of features extracted either from the PEr profiles or from other key features of the EEG.

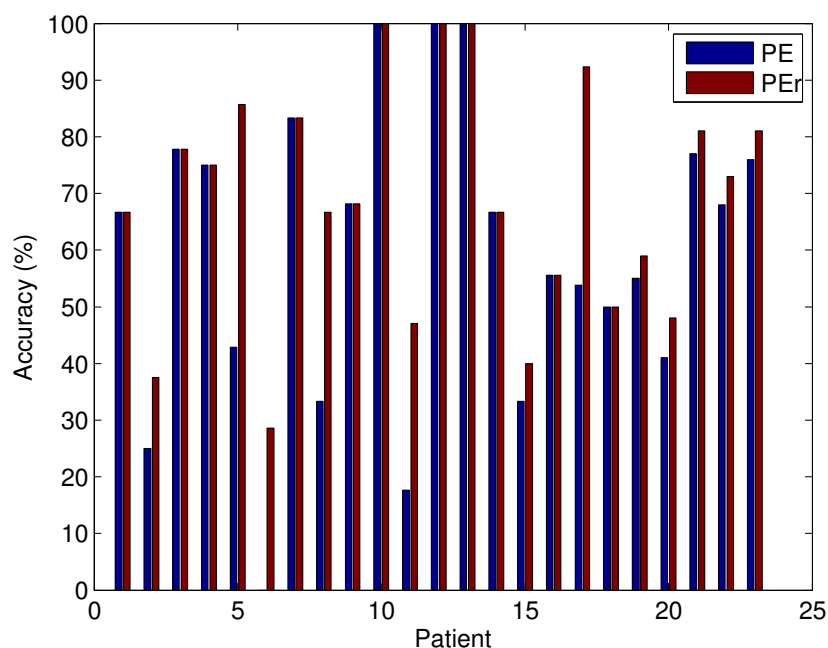


Figure 9. Comparison of the accuracy in the cerebral state classification (interictal or ictal) provided by PE + learning vector quantization (LVQ) and by PEr + LVQ.

4. Conclusions

Permutation entropy can measure the complexity of the electroencephalogram, which is a key feature when the aim is to diagnose neurological pathologies that affect the complexity of the EEG. This is the case with patients affected by absence seizures. The goal of the present research was to improve the ability of PE in discriminating interictal (seizure free) states from ictal (seizure) states in absence epilepsy EEG. For this purpose, a parametrical definition of permutation entropy and the permutation Rényi entropy is here introduced in the field of epileptic EEG analysis and tested against PE, tuning the different parameters involved (order, delay time and alpha). The results show that PEr performs better than the standard PE, as it provides a wider gap between the average interictal and ictal entropy levels. After the optimization of the parameter setting, PEr was tested over 23 EEG recordings in order to assess its ability in discriminating interictal from ictal states. The ability of PEr to discriminate between interictal and ictal states was statistically significant, and it outperformed PE, as endorsed by the statistical hypothesis tests and by the classification of the cerebral states based on PEr. In the future, PEr could replace PE in the techniques mentioned in the Introduction to assess if such methodologies for the analysis of the EEG may benefit from PEr's superior sensitivity.

Acknowledgments

This work was co-funded by the Italian Ministry of Health, Project Code GR-2011-02351397.

Author Contributions

Nadia Mammone designed the research, implemented the algorithm and ran the simulations. Francesco C. Morabito carried out an intensive research on the state-of-the-art in this topic and provided suggestions about algorithm design and results presentation. Jonas Duun-Henriksen and Troels Wesenberg Kjaer recorded the EEG data, set up the database and helped with setting the goals of the research. Every author contributed in co-writing the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

Jonas Duun-Henriksen is employed by HypoSafe A/S, a company that develops EEG recording equipment.

Appendix

A. Statistical Analysis

A.1. D_{PEr} is Higher than D_{PE}

The result $D_{PEr} > D_{PE}$ was tested statistically considering the two \mathbf{D}_{PEr} and \mathbf{D}_{PE} vectors as the two populations to be compared: $X1 = \mathbf{D}_{PEr}$ and $X2 = \mathbf{D}_{PE}$. The hypothesis to be tested is that PEr outperforms PE in discriminating interictal from ictal EEG in absence seizure EEG. The alternative

hypothesis, H1, can be formulated as follows: the difference between the average entropy of the interictal EEG and the average entropy of the ictal EEG is higher with PEr than with PE. The null hypothesis, H0, thus can be formulated as: the difference between the average entropy of the interictal EEG and the average entropy of the ictal EEG is the same for PEr and PE. First of all, we tested the normality of the two populations $X1 = \mathbf{D}_{PEr}$ and $X2 = \mathbf{D}_{PE}$ by means of the Shapiro–Wilk test [29]. Both populations were estimated to be normally distributed, and hence, the one-tailed t -test was chosen. The decision rule is to reject H0 if:

$$\frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}_{\bar{X}_1 - \bar{X}_2}} > t_{n1+n2-2, \alpha} \quad (6)$$

where:

$$\hat{\sigma}_{\bar{X}_1 - \bar{X}_2} = \sigma_P \sqrt{\frac{1}{n1} + \frac{1}{n2}} \quad (7)$$

and:

$$\sigma_P = \sqrt{\frac{(n1 - 1)\sigma_1^2 + (n2 - 1)\sigma_2^2}{n1 + n2 - 2}} \quad (8)$$

where $\alpha = 0.05$. The critical value $t_{n1 + n2 - 2, \alpha}$ was 1.6808, and as the result was $t\hat{a}\check{L}\check{S}test = 4.26$, according to the decision rule, the null hypothesis was rejected.

A.2. R_{PEr} is Higher than R_{PE}

This result was tested statistically considering the two \mathbf{R} vectors as the two populations to be compared: $X1 = \mathbf{R}_{PEr}$ and $X2 = \mathbf{R}_{PE}$. The alternative hypothesis, H1, can be formulated as follows: the ratio between the average entropy of the interictal EEG and the average entropy of the ictal EEG is higher with PEr than with PE. The null hypothesis, H0, can be formulated as: the ratio between the average entropy of the interictal EEG and the average entropy of the ictal EEG is the same for PEr and PE. First of all, we tested the normality of the two populations $X1 = \mathbf{R}_{PEr}$ and $X2 = \mathbf{R}_{PE}$ by means of the Shapiro–Wilk test again. Both populations were estimated to be normally distributed, and hence, the one-tailed t -test was chosen again. The critical value $t_{n1 + n2 - 2, \alpha}$ was 1.6808, and as the result was $t\hat{a}\check{L}\check{S}test = 4.52$, according to the decision rule, the null hypothesis was rejected.

A.3. $avgPER_{interictal}$ is Higher than $avgPER_{ictal}$

This result was tested statistically considering the two **avgPER** vectors as the two populations to be compared: $X1 = \mathbf{avgPER}(\mathbf{interictal})$ and $X2 = \mathbf{avgPER}(\mathbf{ictal})$. The alternative hypothesis, H1, can be formulated as follows: the average PEr in the interictal EEG is higher than the average PEr in the ictal EEG. The null hypothesis, H0, can be formulated as: there is no difference between the average PEr in the interictal EEG and the average PEr in the ictal EEG. First of all, we tested the normality of the two populations $\mathbf{X1}$ and $\mathbf{X2}$ by means of the Shapiro–Wilk test. Both populations were estimated not to be normally distributed. Hence, the Mann–Whitney U-test was chosen as the method for the hypothesis test [30]. The result of the test was $p = 1.5214e - 05$. The null hypothesis could thus be rejected at the significance level of 0.05.

References

1. Duun-Henriksen, J.; Madsen, R.E.; Remvig, L.S.; Thomsen, C.E.; Sorensen, H.B.; Kjaer, T.W. Automatic detection of childhood absence epilepsy seizures: Toward a monitoring device. *Pediatr. Neurol.* **2012**, *46*, 287–292.
2. Duun-Henriksen, J.; Kjaer, T.W.; Madsen, R.E.; Remvig, L.S.; Thomsen, C.E.; Sorensen, H.B. Channel selection for automatic seizure detection. *Clin. Neurophysiol.* **2012**, *123*, 84–92.
3. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102.
4. Cao, Y.; Tung, W-w; Gao, J.B.; Protopopescu, V.A.; Hively, L.M. Detecting dynamical changes in time series using the permutation entropy. *Phys. Rev. E* **2004**, *70*, 046217.
5. Li, X.; Ouyang, G.; Douglas, A.R. Predictability analysis of absence seizures with permutation entropy. *Epilepsy Research* **2007**, *77*, 70–74,.
6. Bruzzo, A.A.; Gesierich, B.; Santi, M.; Tassinari, C.A.; Birbaumer, N.; Rubboli, G. Permutation entropy to detect vigilance changes and preictal states from scalp EEG in epileptic patients. a preliminary study. *Neurol Sci.* **2008**, *29*, 3–9.
7. Zanin, M.; Zunino, L.; Rosso, O.A.; Papo, D. Permutation entropy and its main biomedical and econophysics applications: A review. *Entropy* **2012**, *14*, 1553–1577.
8. Nicolaou, N.; Georgiou, J. Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. *Expert Syst. Appl.* **2012**, *39*, 202–209.
9. Ouyang, G.; Li, J.; Liu, X.; Li, X. Dynamic characteristics of absence EEG recordings with multiscale permutation entropy analysis. *Epilepsy Res.* **2013**, *104*, 246–252.
10. Zhu, G.; Li, Y.; Wen, P.P.; Wang, S.; Xi, Epileptogenic focus detection in intracranial EEG based on delay permutation entropy. In Proceedings of 2013 International Symposium on Computational Models for Life Sciences (CMLS 2013), Sydney, Australia, 27–29 November 2013; pp. 31–36.
11. Zhu, G.; Li, Y.; Wen, P.P.; Wang, S. Classifying epileptic EEG signals with delay permutation entropy and multi-scale k-means. *Adv. Exp. Med. Biol.* **2015**, *823*, 143–157.
12. Mateos, D.; Diaz, J.M.; Lamberti, P.W. Permutation entropy applied to the characterization of the clinical evolution of epileptic patients under pharmacological treatment. *Entropy* **2014**, *16*, 5668–5676.
13. Li, J.; Liu, X.; Ouyang, G. Using relevance feedback to distinguish the changes in EEG during different absence seizure phases. *Clin. EEG Neurosci.* **2014**, doi:10.1177/1550059414548721.
14. Yang, Z.; Wang, Y.; Ouyang, G. Adaptive neuro-fuzzy inference system for classification of background EEG signals from eses patients and controls. *Sci. World J.* **2014**, *2014*, 140863.
15. Silva, A.; Ferreira, D.A.; Venâncio, C.; Souza, A.P.; Antunes, L.M. Performance of electroencephalogram-derived parameters in prediction of depth of anaesthesia in a rabbit model. *Br. J. Anaesth.* **2011**, *106*, 540–547.
16. Li, D.; Li, X.; Liang, Z.; Voss, L.J.; Sleigh, J.W. Multiscale permutation entropy analysis of EEG recordings during sevoflurane anesthesia. *J. Neural Eng.* **2010**, *7*, 046010.

17. Anier, A.; Lipping, T.; Jantti, V.; Puumala, P.; Huotari, A.M. Entropy of the EEG in transition to burst suppression in deep anesthesia: Surrogate analysis. In Proceedings of 32nd Annual International Conference on IEEE Engineering in Medicine and Biology Society (EMBC), Buenos Aires, Argentina, 1–4 September 2010; pp. 2790–2793.
18. Olofsen, E.; Sleigh, J.W.; Dahanĳĳĳĳ A. Permutation entropy of the electroencephalogram: a measure of anesthetic drug effect. *Br. J. Anaesth.* **2008**, *101*, 810–821.
19. Li, D.; Liang, Z.; Wang, Y.; Hagihira, S.; Sleigh, J.W.; Li, X. Parameter selection in permutation entropy for an electroencephalographic measure of isoflurane anesthetic drug effect. *J. Clin. Monit. Comput.* **2013**, *27*, 113–123.
20. Mammone, N.; Morabito, F.C.; Principe, J.C. Visualization of the short term maximum lyapunov exponent topography in the epileptic brain. In Proceedings of 28th Annual International Conference on IEEE Engineering in Medicine and Biology Society (EMBC), New York City, NY, USA, 31 August–3 September 2006; pages 4257–4260.
21. Mammone, N.; Principe, J.C.; Morabito, F.C.; Shiau, D.S.; Sackellares, J.C. Visualization and modelling of stlmax topographic brain activity maps. *J. Neurosci. Methods* **2010**, *189*, 281–294.
22. Mammone, N.; Morabito, F.C. Analysis of absence seizure EEG via permutation entropy spatio-temporal clustering. In Proceedings of the 2011 International Joint Conference on Neural Networks (IJCNN), San Jose, CA, USA, 31 July–5 August 2011; pp. 1417–1422.
23. Mammone, N.; Labate, D.; Lay-Ekuakille, A.; Morabito, F.C. Analysis of absence seizure generation using EEG spatial-temporal regularity measures. *Int. J. Neural Syst.* **2012**, *22*, 1250024.
24. Ferlazzo, E.; Mammone, N.; Cianci, V.; Gasparini, S.; Gambardella, A.; Labate, A.; Latella, M.A.; Sofia, V.; Elia, M.; Morabito, F.C.; Aguglia, U. Permutation entropy of scalp EEG: A tool to investigate epilepsies: Suggestions from absence epilepsies. *Clin. Neurophysiol.* **2014**, *125*, 13–20.
25. Zhao, X.; Shang, P.; Huang, J. Permutation complexity and dependence measures of time series. *Europhys. Lett.* **2013**, *102*, 40005.
26. Liang, Z.; Wang, Y.; Sun, X.; Li, D.; Voss, L.J.; Sleigh, J.W.; Hagihira, S.; Li, X. EEG entropy measures in anesthesia. *Front. Comput. Neurosci.* **2015**, *9*, doi:10.3389/fncom.2015.00016.
27. Malmivuo, J.; Plonsey, R. *Bioelectromagnetism—Principles and Applications of Bioelectric and Biomagnetic Fields*; Oxford University Press: New York, NY, USA, 1995.
28. Kohonen, T. Learning vector quantization. In *The Handbook of Brain Theory and Neural Networks*; MIT Press: Cambridge, MA, USA, 1995; pp. 537–540.
29. Shapiro, S.S.; Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika* **1965**, *52*, 591–611.
30. Mann, H.B.; Whitney, D.R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **1947**, *18*, 50–60.