Non-linear Analysis of Intracranial Electroencephalogram Recordings with Approximate Entropy and Lempel-Ziv Complexity for Epileptic Seizure Detection

Daniel Abásolo*, *Member, IEEE,* Christopher J. James, *Senior Member, IEEE,* and Roberto Hornero, *Member, IEEE*

Abstract—**Epileptic seizures are generated by an abnormal synchronization of neurons unforeseeable for the patients. In this study we analyzed invasive electroencephalogram (EEG) recordings in patients suffering from medically intractable focal epilepsy with two non-linear methods, Approximate Entropy (*ApEn*) and Lempel-Ziv (*LZ*) complexity. *ApEn* and *LZ* complexity quantify the regularity and complexity of a time series, respectively, and are well suited to the analysis of non-stationary biomedical signals of short length. Our results show an increase in *ApEn* and *LZ* complexity values during seizures at the focal electrodes. These changes could also be seen at some extra focal electrodes. After the seizure ends, the values of both non-linear metrics return to values lower than those before the seizure. Moreover, we quantified the changes in *LZ* complexity, showing the complexity increase during the seizure and its notable decrease after its end. Our results suggest that these techniques are useful to detect changes due to epileptic seizures in the EEG.**

I. Introduction

EPILEPSY is characterized by a sudden and recurrent malfunction of the brain that is termed "seizure" [1]. Epileptic seizures reflect the clinical signs of an excessive and hyper-synchronous activity of neurons in the brain. Approximately 1% of the world's population suffers from epilepsy. With current antiepileptic drugs, seizures can be controlled satisfactorily in about two thirds of the affected individuals and another 8% could benefit from epilepsy surgery. However, the remaining 25% of epilepsy patients cannot be treated sufficiently by any available therapy [1].

The electroencephalogram (EEG) has been used as a tool for investigating the brain electrical activity in different physiological and pathological states for several decades. When an epileptic focal seizure is generated, synchronized epileptic brain activity is initially observed in a small area of the brain. From this focus, the activity spreads to other brain areas [2]. This process should reflect in the recorded EEG. Visual inspection of the EEG data has not yet led to the detection of any characteristic changes preceding seizure onsets. However, prediction of seizures will considerably increase the quality of life of epileptic patients who cannot be treated successfully [2].

Traditional linear analysis has been used for seizure prediction but it has its limits [3]. As non-linearity in the brain is introduced even at the cellular level, non-linear dynamical analysis techniques may be a better approach than traditional linear methods to understand the EEG dynamics. In fact, different non-linear techniques have been applied to seizure prediction [2]–[4].

The present pilot study was undertaken to examine intracranial EEG recordings of patients suffering from medically intractable focal epilepsy by means of two different non-linear methods: Approximate Entropy (*ApEn*), and Lempel-Ziv (*LZ*) complexity. *ApEn* quantifies the regularity in data [5], while *LZ* complexity is a non-parametric measure of complexity for one-dimensional signals [6], [7].

II. Material and Methods

A. Invasive EEG recordings

We analyzed invasive EEGs, recorded during presurgical monitoring at the Epilepsy Center of the University Hospital of Freiburg, Germany. Intracranial grid-, strip-, and depth-electrodes were used. The EEG data were acquired using a Neurofile NT digital video EEG system with 128 channels, 256 Hz sampling rate, and a 16 bit analogue-to-digital converter. 23 EEG recordings with simple partial, complex partial and generalized tonic-clonic seizures from 5 patients (3 men and 2 women, mean age: 35.4 ± 14.2 years, mean ± standard deviation, SD) with focal epilepsy originated in the temporal region were selected. All recordings were digitally filtered with a band-pass filter with cut-off frequencies at 0.5 Hz and at 40 Hz prior to the computation of *ApEn* and *LZ* complexity.

B. Approximate entropy

ApEn is a family of statistics introduced to provide a widely applicable, statistically valid formula that will distinguish data sets by a measure of regularity [5]. *ApEn*

*D. Abásolo and R. Hornero are with the Biomedical Engineering Group, Department of Signal Theory and Communications, E.T.S.I. de Telecomunicación, University of Valladolid, Camino del Cementerio s/n, 47011, Valladolid, Spain (e-mail: daniel.abasolo@tel.uva.es).

C.J. James is with the Signal Processing and Control Group, Institute of Sound and Vibration Research, University of Southampton, University Road, Southampton, SO17 1BJ, United Kingdom.

evaluates dominant and subordinated patterns in data, and it discriminates series for which clear feature recognition is difficult [8]. Moreover, it can be applied to short time series and it is finite for stochastic, noisy deterministic and composite processes [8]. Larger *ApEn* values correspond to more irregularity in the data.

Two input parameters must be specified to compute *ApEn*: a run length *m* and a tolerance window *r*. *ApEn* measures the logarithmic likelihood that runs of patterns that are close (within *r*) for *m* contiguous observations remain close (within the same *r*) on subsequent incremental comparisons. Given *N* data points from a time series $\{x(n)\} = x(1), x(2), \dots, x(N)$, one can compute *ApEn* as follows [8]:

1. Form $N{-}m{+}1$ vectors $X(1)\dots X(N{-}m{+}1)$ defined by: $X(i) = [x(i), x(i{+}1),\dots, x(i{+}m{-}1)]$, $i = 1\dots N{-}m{+}1$.
2. Define the distance between $X(i)$ and $X(j)$ as:

$$d[X(i),X(j)] = \max_{k=1,2,\dots m} |x(i{+}k{-}1) - x(j{+}k{-}1)|. \quad (1)$$

3. For a given $X(i)$, count the number of j ($j = 1\dots N{-}m{+}1$) so that $d[X(i), X(j)] \leq r$, denoted as $N^m(i)$. Then, for $i = 1\dots N{-}m{+}1$,

$$C_r^m(i) = N^m(i)/(N{-}m{+}1). \quad (2)$$

$C_r^m(i)$ measures, within a tolerance *r*, the frequency of patterns similar to a given one of window length *m*.

4. Compute the natural logarithm of each $C_r^m(i)$ and average it over *i*,

$$\phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_r^m(i). \quad (3)$$

5. Increase the run length and find $C_r^{m+1}(i)$ and $\phi^{m+1}(r)$.
6. We define *ApEn* by:

$$ApEn(m,r,N) = \phi^m(r) - \phi^{m+1}(r). \quad (4)$$

As suggested in [8], *ApEn* was estimated with $m = 1$ and $r = 0.25$ times the SD of the original data sequence. These values produce good statistical reproducibility in *ApEn* for time series of length $N \geq 60$, as considered herein [5], [9].

C. Lempel-Ziv complexity

LZ complexity analysis is based on a coarse-graining of the measurements. Before calculating the *LZ* complexity measure *c(n)*, the signal must be transformed into a finite symbol sequence. In the context of biomedical signal analysis, typically the discrete–time signal $\{x(n)\}$ is converted into a binary sequence. By comparison with a threshold T_d, the original signal samples are converted into a 0–1 sequence $P = s(1), s(2),\dots, s(n)$, with $s(i)$ defined by:

$$s(i) = \begin{cases} 0 & if\ x(i) < T_d \\ 1 & if\ x(i) \geq T_d \end{cases}. \quad (5)$$

We used the median as the threshold T_d because of its robustness to outliers [7]. Previous studies [10]–[13] have shown that 0–1 conversion is adequate to estimate the *LZ* complexity in biomedical signals.

To compute *LZ* complexity, the sequence *P* has to be scanned from left to right and a complexity counter *c(n)* is increased by one unit every time a new subsequence of consecutive characters is encountered. The complexity measure can be estimated as follows [13]:

1. Let *S* and *Q* denote two subsequences of *P* and *SQ* be the concatenation of *S* and *Q*, while *SQπ* is derived from *SQ* after its last character is deleted. Let *v(SQπ)* denote the vocabulary of all different subsequences of *SQπ*. Initially, $c(n) = 1$, $S = s(1)$, $Q = s(2)$ and $SQπ = s(1)$.
2. In general, $S = s(1), s(2),\dots, s(r)$, $Q = s(r{+}1)$, then $SQπ = s(1), s(2),\dots, s(r)$; if *Q* belongs to *v(SQπ)*, then *Q* is a subsequence of *SQπ*, not a new sequence.
3. Renew *Q* to be $s(r{+}1), s(r{+}2)$ and judge if *Q* belongs to *v(SQπ)* or not.
4. Repeat the previous steps until *Q* does not belong to *v(SQπ)*. Now $Q = s(r{+}1), s(r{+}2),\dots, s(r{+}i)$ is not a subsequence of $SQπ = s(1), s(2),\dots, s(r{+}i{-}1)$, so increase *c(n)* by one.
5. Thereafter, *S* is renewed to be $S = s(1), s(2),\dots, s(r{+}i)$, and $Q = s(r{+}i{+}1)$.

These procedures have to be repeated until *Q* is the last character. At this time the number of different subsequences in *P* is *c(n)*. To obtain a complexity measure independent of the sequence length, *c(n)* should be normalized. In the case of a 0–1 sequence, *c(n)* can be normalized as [13]:

$$C(n) = \frac{c(n)}{\dfrac{n}{\log_2(n)}}. \quad (6)$$

The normalized *LZ* complexity reflects the arising rate of new patterns along with the sequence, capturing its temporal structure. Larger values correspond to more complexity.

III. RESULTS

Initially we computed *ApEn* and *LZ* complexity of five different seizures in one patient suffering from temporal-lobe epilepsy. We analyzed the evolution over time before, during and after the seizure with a sliding window of 10s, and a 1s delay for each window. Thus, we have one non-linear metric for each second of the recording. Fig. 1 shows the changes in these non-linear metrics with time during one of these seizures. The *ApEn* and *LZ* complexity values increase during the seizure at the focal electrodes, where a slight drop in their values before the onset is followed by an increase during the seizure. This is quite clear at electrodes 2 and 3. After the seizure ends, the values of both non-linear metrics return to lower values. It is interesting to note that the changes in *ApEn* and *LZ* complexity can also be seen at the extra focal electrode 6. These results suggest that seizures are reflected in an increase of EEG complexity and irregularity. Furthermore, the peak in the non-linear metrics does not happen at the same time. It appears first at focal electrode 1, then at focal electrodes 3 and 2 and extra focal electrode 6, respectively. This shows that these non-linear metrics can detect how the seizure spreads from the focal area to the extra focal electrodes.
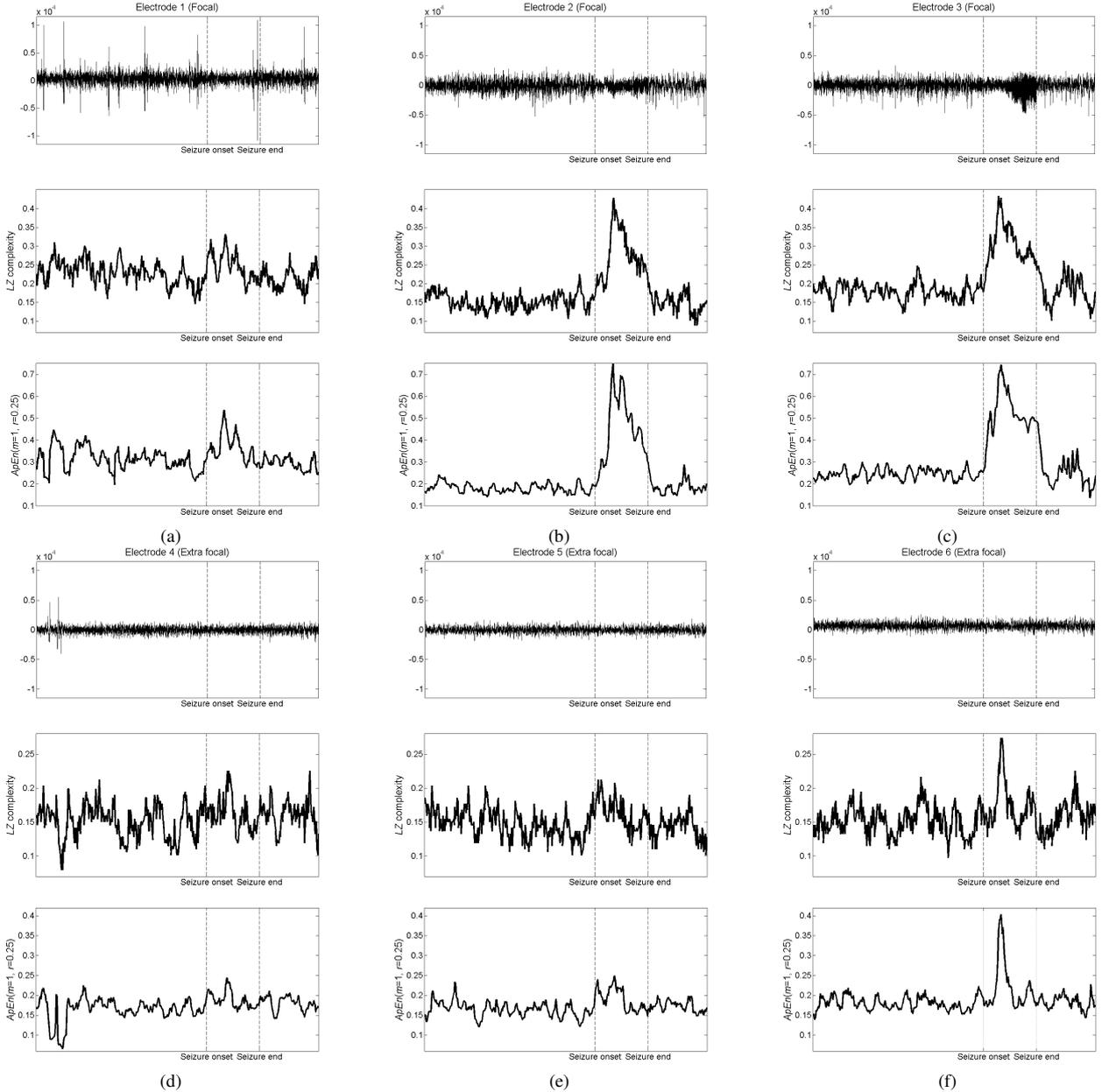
Fig. 1. Evolution of the invasive EEG recording and *ApEn* and *LZ* complexity values over time before the epileptic seizure onset, during the epileptic seizure and after its end; (a) to (c) focal electrodes, (d) to (f) extra focal electrodes.

Although *ApEn* curves are smoother, the changes over time are quite similar in both methods. Given the fact that estimating *LZ* complexity has a much lower computational cost, we quantified changes due to seizures with this method in 23 EEG recordings from 5 patients with focal temporal-lobe epilepsy. We have averaged our results in three different time intervals: before the seizure onset (interval of 5 minutes), during the seizure and after the seizure ends (2 minutes, or until the recording finishes if there were not enough samples). Table 1 summarizes the average *LZ* complexity values for all seizures in each subject and the average *LZ* complexity values for all seizures. With the exception of patient 15, *LZ* complexity clearly increases at the focal electrodes during the seizure. However, this

increase is not clear at the extra focal electrodes. Nevertheless, the *LZ* complexity after the seizure is notably lower than before the seizure at the focal and some extra focal electrodes. This might be due to a general slowing of the EEG associated with the end of the seizure.

IV. DISCUSSION AND CONCLUSION

In this pilot study, we analyzed invasive EEG recordings from patients with medically intractable focal epilepsy by means of *ApEn* and *LZ* complexity. *ApEn* quantifies the regularity in time series, while *LZ* complexity quantifies the complexity as the rate of recurrence of patterns in a one-dimensional signal. We studied the evolution of *ApEn* and *LZ* complexity over time in recordings with seizures from

Subj.	Elect.	Before seizure (mean±SD)	During seizure (mean±SD)	After seizure end (mean±SD)
12(4)	F	0.2096 ± 0.0314	0.3157 ± 0.0367	0.1979 ± 0.0200
12(4)	EF	0.1851 ± 0.0184	0.1925 ± 0.0162	0.1880 ± 0.0248
15(4)	F	0.2104 ± 0.0243	0.2102 ± 0.0591	0.1339 ± 0.0454
15(4)	EF	0.2215 ± 0.0246	0.2058 ± 0.0635	0.1430 ± 0.0567
16(5)	F	0.2942 ± 0.0184	0.3213 ± 0.0455	0.2131 ± 0.0420
16(5)	EF	0.2832 ± 0.0817	0.2570 ± 0.0861	0.2638 ± 0.1020
17(5)	F	0.2463 ± 0.0415	0.3024 ± 0.0331	0.1067 ± 0.0144
17(5)	EF	0.2766 ± 0.0418	0.3159 ± 0.0452	0.1153 ± 0.0151
21(5)	F	0.2212 ± 0.0439	0.2961 ± 0.0516	0.2098 ± 0.0414
21(5)	EF	0.2007 ± 0.0352	0.2287 ± 0.0361	0.2134 ± 0.0383
All	F	0.2400 ± 0.0463	0.2914 ± 0.0590	0.1728 ± 0.0562
All	EF	0.2384 ± 0.0616	0.2435 ± 0.0697	0.1863 ± 0.0777

F: Focal electrodes; EF: Extra focal electrodes.

one patient. Our results show an increase in *ApEn* and *LZ* complexity during the seizure at the focal electrodes. These changes could also be seen at some extra focal electrodes. Given the fact that both methods are related and the lower computational cost of *LZ* complexity, we quantified the EEG changes due to the seizures with this method. Our results show that *LZ* complexity increases during the seizures, especially at focal electrodes, and that it decreases after the seizures end to values lower than before the seizure.

Despite the usefulness of non-linear analysis techniques to characterize the EEG activity, little attention has been paid to the fulfillment of the basic conditions under which these methods can be applied. Non-linear techniques usually require stationarity in the time series, as is also the case with many linear measures. Moreover, they also require an amount of data for meaningful results in their computation beyond the experimental possibilities for physiological data. On the other hand, *ApEn* and *LZ* complexity do not show these problems. The former is finite for stochastic, noisy deterministic and composite processes [8], while the latter is model-independent: only those differences between activity patterns that make a difference to the underlying system itself are considered, no matter whether the system is dominated by deterministic chaos or a stochastic process [13]. Furthermore, while applying these non-linear methods to the EEG, we are not testing for a particular model form, but attempting to distinguish features in the EEG recordings.

Some limitations of our study merit consideration. First of all, the sample size was small to demonstrate the capability of the method. Further studies are currently under way with a larger sample size and comprehensive statistical analysis to evaluate the utility of our methodology in epileptic seizure detection. Moreover, although our preliminary results suggest that *ApEn* and *LZ* complexity can detect epileptic seizures from invasive EEG recordings, further work must be carried out to prove the possible usefulness of these techniques in seizure prediction. Future lines of research include the development of a seizure predictor with these non-linear metrics, the assessment of its performance with the seizure prediction characteristic and a larger patient population [5], and the evaluation of the interrelations among electrodes with non-linear techniques such as cross-*ApEn*, cross-Sample Entropy or cross-Mutual Information. Furthermore, it has been recently shown that new measures of complexity based on recurrence plots can detect and quantify changes preceding life-threatening cardiac arrhythmias thereby facilitating a prediction of such event [14]. Thus, the possible usefulness of these measures in epileptic seizure prediction should be checked.

REFERENCES

[1] K. Lehnertz, F. Mormann, T. Kreuz, R. G. Andrzejak, C. Rieke, P. David, and C. E. Elger, "Seizure prediction by nonlinear EEG analysis," *IEEE Eng. Med. Biol.*, vol. 22, pp. 57–63, Jan.–Feb. 2003.

[2] B. Schelter, M. Winterhalder, T. Maiwald, A. Brandt, A. Schad, A. Schulze-Bonhage, and J. Timmer, "Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction," *Chaos*, vol. 16, p. 013108, Jan. 2006.

[3] N. Päivinen, S. Lammi, A. Pitkänen, J. Nissinen, M. Pentonnen, and T. Grönfors, "Epileptic seizure detection: A nonlinear viewpoint," *Comput. Meth. Prog. Bio.*, vol. 79, pp. 151–159, Aug. 2005.

[4] T. Mainwald, M. Winterhalder, R. Aschenbrenner-Scheibe. H. U. Voss, A. Schulze-Bonhage, and J. Timmer, "Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic," *Physica D*, vol. 194, pp 357–368, Jul. 2004.

[5] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proc. Natl. Acad. Sci. USA*, vol. 88, pp. 2297–2301, Mar. 1991.

[6] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Trans. Inform. Theory*, vol. 22, pp. 75–81, Jan. 1976.

[7] R. Nagarajan, "Quantifying physiological data with Lempel-Ziv complexity–certain issues," *IEEE Trans. Biomed. Eng.*, vol. 49, pp. 1371–1373, Nov. 2002.

[8] S. M. Pincus, "Assessing serial irregularity and its implications for health," *Ann. NY Acad. Sci.*, vol. 954, pp. 245–267, Dec. 2001.

[9] S. M. Pincus and D. L. Keefe, "Quantification of hormone pulsatility via an approximate entropy algorithm," *Am. J. Physiol. (Endocrinol. Metab.)*, vol. 262, pp. E741–E754, May 1992.

[10] N. Radhakrishnan and B. N. Gangadhar, "Estimating regularity in epileptic seizure time-series data. A complexity-measure approach," *IEEE Eng. Med. Biol.*, vol. 17, pp. 89–94, May-Jun. 1998.

[11] J. Xu, Z. R. Liu, R. Liu, and Q. F. Yang, "Information transformation in human cerebral cortex," *Physica D*, vol. 106, pp. 363–374, Aug. 1997.

[12] X. S. Zhang, R. J. Roy, and E. W. Jensen, "EEG complexity as a measure of depth of anesthesia for patients," *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 1424–1433, Dec. 2001.

[13] X. S. Zhang, Y. S. Zhu, N. V. Thakor, and Z. Z. Wang, "Detecting ventricular tachycardia and fibrillation by complexity measure," *IEEE Trans. Biomed. Eng.*, vol. 46, pp. 548–555, May 1999.

[14] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths, "Recurrence-plot based measures of complexity and their application to heart-rate variability data," *Phys. Rev. E*, vol. 66, p. 026702, Aug. 2002.