EEG Signal Features Extraction based on Fractal Dimension

Francesca Finotello, Fabio Scarpa and Mattia Zanon

***Abstract*— The spread of electroencephalography (EEG) in countless applications has fostered the development of new techniques for extracting synthetic and informative features from EEG signals. However, the definition of an effective feature set depends on the specific problem to be addressed and is currently an active field of research. In this work, we investigated the application of features based on fractal dimension to a problem of sleep identification from EEG data. We demonstrated that features based on fractal dimension, including two novel indices defined in this work, add valuable information to standard EEG features and significantly improve sleep identification performance.**

I. INTRODUCTION

THE increasingly widespread use of electroencephalography (EEG) applied to different medical and research fields (e.g. in the study of epilepsy and sleep-related disorders, for the measurement of evoked potentials and for brain-computer interface) has led to the development of new methodologies for features extraction, namely for extracting quantitative information from the recorded signals. The EEG characteristics are thus described by features reflecting a distinguishing property, a recognizable measurement or a functional component obtained from a section of the recorded signal [1], so to highlight important information embedded in the signal. Features are meant to compress the information contained in the original signal, lowering the amount of resources needed to describe the original data, while conveying the discriminative characteristics of the phenomenon under study. In the last decades, a variety of methods have been widely used to extract the features from EEG signals, such as time-frequency distributions, fast Fourier transform, eigenvector methods, wavelet transform and auto regressive method [2]. More recently, promising features based on the field of mathematics also known as "chaos theory" have been introduced. Indeed, the complexity and limited predictability of EEG signals have motivated the application of non-linear indexes originally developed for the representation of chaotic systems, such as the fractal dimension [3, 4, 5, 6, 7]. In this work we defined two novel features based on the Higuchi's fractal dimension [8] and applied them to a sleep identification problem, employing a literature EEG data set. The comparison with indexes widely used in EEG data analysis shows that the features based on fractal dimension convey valuable information and that the proposed novel features improve the discrimination between the awake and asleep states on the considered data set.

F. F. is with the Department of Information Engineering, University of Padova, Padova, 35131 Italy, and also with the Biocenter, Division of Bioinformatics, Medical University of Innsbruck, Innsbruck, 6020, Austria (e-mail: `francesca.finotello@dei.unipd.it`).

F. S. is with the Department of Information Engineering, University of Padova, Padova, 35131 Italy (corresponding author, phone: 0039-049-827-7758; fax: 0039-049-827-7689; e-mail: `fabio.scarpa@unipd.it`).

M. Z. was with the Department of Information Engineering, University of Padova, Padova, 35131 Italy (e-mail: `zanon.mattia@gmail.com`).

II. FRACTAL DIMENSION FOR EEG ANALYSIS

The fractal dimension provides a complexity index that describes how the measure of the length of a curve changes depending on the scale k used as unit of measurement. Different approximation techniques can be used to estimate fractal dimension [9], although the Higuchi's algorithm is the most commonly employed.

A. Higuchi's Fractal Dimension

Higuchi's algorithm [8] is one of the most used approaches to estimate the fractal dimension D of an EEG signal. The Higuchi's standard approach consists in the following steps:

- For each sample i of the EEG segment S_j, absolute differences between the values $S_j(i)$ and $S_j(i+k)$ (i.e. samples at distance k) are computed, considering $k = 1, \dots, k_{lin}$.
- These absolute differences are multiplied by a normalization coefficient that takes into account the number of samples available for each value of k.
- For each k, $L(k)$ is computed by summing the obtained values along all samples of the EEG segment and dividing by k.
- By definition, if $L(k)$ value is proportional to k^{-D}, then the curve is fractal with dimension D. If this condition is verified for $k = 1, \dots, k_{lin}$, then $log(k)$ and $log(L(k))$ have a linear relationship. In particular, from the $log(L(k))$ *vs.* $log(k)$ curve, referred in the following as ℓ_k, D can be estimated by ordinary least squares as the linear coefficient of the regression line.
- k_{lin} is chosen as the maximum k for which $L(k)$ is proportional to k^{-D}.

The fractal dimension D with the Higuchi's method is calculated in the linear region of the ℓ_k curve, i.e. for $k \leq k_{lin}$, while the non-linear region of the ℓ_k curve, i.e. for $k > k_{lin}$, is usually not considered.

B. Effect of signal periodicity on Higuchi's curve

Employing simulated signals we show that, for periodic signals, the non-linear part of the Higuchi's curve ℓ_k presents an oscillatory behaviour whose characteristics depend on the periodicity of the signal itself (Fig. 1). We considered four periodic signals oscillating in the $[-1, 1]$ range (arbitrary unit), and their corresponding Higuchi's curve computed for $k = 1, \dots, 80$. It can be seen that the first part of the

978-1-4244-9270-1/15/$31.00 ©2015 IEEE

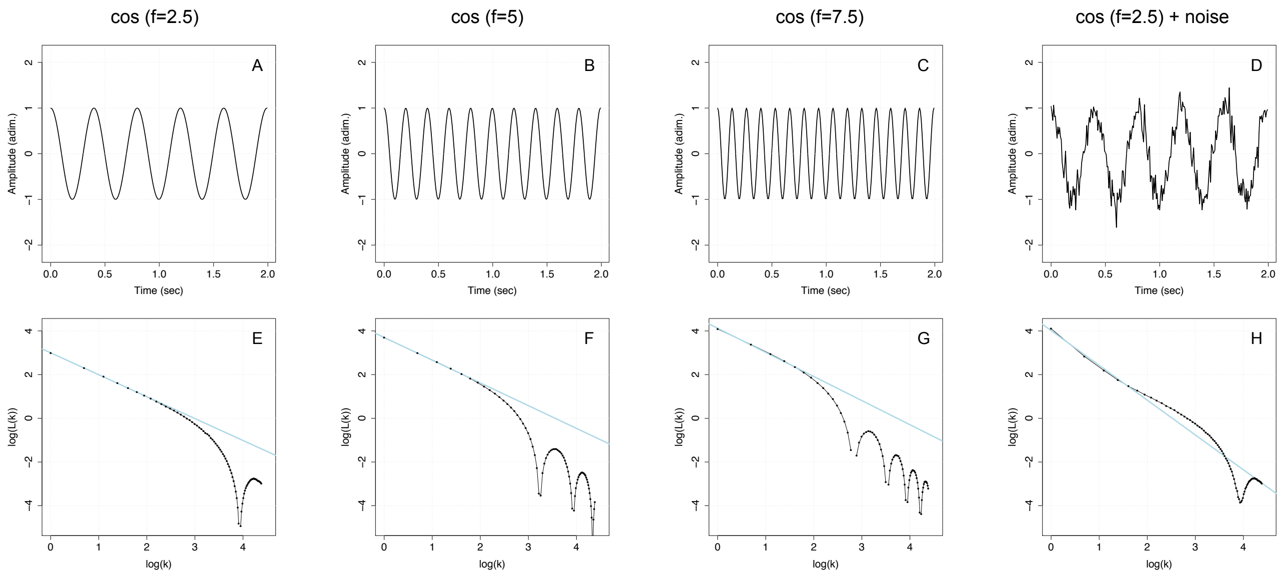

Fig. 1. Examples showing the effect of signal periodicity on Higuchi's curve. The upper panels show the original signals, generated as cosine waves with frequencies $2.5Hz$ (A and D), $5Hz$ (B) and $7.5Hz$ (C). The lower panels (E-H) show the corresponding Higuchi's curve (black) and linear fit (light-blue). The last signal (D and H) is generated adding pseudo-normal random noise, with standard deviation equal to 0.2, to the cosine wave with frequency $2.5Hz$.

ℓ_k curve is approximately linear, while the remaining part presents oscillations (i.e. negative spikes with respect to the linear fit). Oscillating patterns in the ℓ_k curve depend on the frequency of the original signal, with higher frequency resulting in a larger number of negative spikes on ℓ_k. These patterns can be explained by considering the definition of ℓ_k, which is computed as the sum of the absolute differences between all possible pairs of signal samples at distance k. When k is comparable to the number of samples Z comprised in one oscillation of the original signal (namely, a period), the considered sample values will be very similar, possibly equal. Consequently, samples differences will be small and will result in a low value of ℓ_k for that particular k. The same effect is present when k is a multiple of Z, thus explaining the occurrence of several, periodic negative spikes. The decrease in both minimum values and distances of subsequent spikes are due to the normalization by k (see ℓ_k definition in [8]) and the effect of the logarithmic scale, respectively. The presence of additive noise (Fig. 1.H) in the original periodic signal slightly affect the ℓ_k, but the presence and latency of its negative oscillations is preserved. In this example a cosine wave is considered, but the discussion can be generalized to any generic periodic signal.

C. Novel Features based on Higuchi's curve

Given the relationship between a periodic signal and its ℓ_k curve illustrated above, here we considered a wider domain of the ℓ_k curve with respect to the standard approach, so to investigate its non-linear characteristics. In particular, in addition to the standard Higuchi's fractal dimension (called simply "Higuchi" from now on) computed in the linear region of ℓ_k, we defined two novel features computed considering also the non-linear region. The first feature, here called "FD_residual", evaluates the deviation from the linear fit of ℓ_k, calculated as the sum of squares of the residuals in the non-linear region. The regression line is estimated considering only the linear region of ℓ_k. The second feature, here called "FD_tortuosity", gives a tortuosity measure τ of the ℓ_k curve. It computes a measure of the rate at which the curve is changing with respect to its coordinates changes ($x = log(k)$ and $y = log(L(k))$), by using their first and second partial differences:

$$\tau(\ell_k) = \sum_{n=3}^{N} \left| \frac{\Delta x(n)\Delta^2 y(n) - \Delta^2 x(n)\Delta y(n)}{\left[(\Delta x(n))^2 + (\Delta y(n))^2\right]^{3/2}} \right| \qquad (1)$$

where N is the number of points of the curve ℓ_k, and $\Delta x(n) = x(n) - x(n-1)$, $\Delta^2 x(n) = \Delta x(n) - \Delta x(n-1)$, $\Delta y(n) = y(n) - y(n-1), \Delta^2 y(n) = \Delta y(n) - \Delta y(n-1)$. A complete description of the tortuosity measure can be found in [10]. The linear region is here defined considering $k_{lin} = 6$, according to [4], while the two novel features are computed considering the non-linear region up to $k = 18$.

III. FREQUENCY DOMAIN FEATURES

In standard EEG data analysis, EEG signals are often reported to a frequency domain and analysed in terms of spectral power, i.e. decomposed into a sum of pure frequency components. The spectral power describes the rhythmic activity present in EEG signals and is usually divided into characteristic frequency bands that can be related to different brain states. Here we considered them as a reference for assessing the performance of the novel fractal dimension indexes. We computed the spectral power for each frequency in the range $1-30Hz$ using the non-parametric Welch's periodogram method [11]. We considered four frequency bands: delta δ $(1-4Hz)$, theta θ $(4-8Hz)$, alpha α $(8-12Hz)$ and beta β $(12-30Hz)$, and computed the relative spectral power, namely band-specific spectral power (P_{band}) normalized by the total spectral power in $1-30Hz$ (P_{1-30Hz}): P'_{δ}, P'_{θ}, P'_{α}, and P'_{β}, where $P'_{band} = P_{band}/P_{1-30Hz}$. In the following, we refer to these features as "Freq_delta", "Freq_theta", "Freq_alpha" and "Freq_beta", respectively.

IV. DATA SET

To test the ability of the considered features to discriminate between the "awake" and "asleep" state in healthy subjects, we considered EEG data from the Sleep-EDF Database [12]. They consist in 24-hours EEG signals acquired from adult individuals using standard EEG laboratory equipment. EEG signals were recorded from FpzCz and PzOz derivations at $100Hz$ sampling rate and each 30-seconds epoch was manually scored either as one of the five sleep stages or as the "awake" state [13]. For our purpose, we considered only the "sc*" recordings (PzOz derivation), acquired from individuals without sleep-related disorders, for a total of four individuals and 339 413 acquisition seconds. We grouped all sleep stages in a single "asleep" state. We split the original EEG signal into segments of two-seconds length, with one-second overlap, and performed linear de-trending to ensure the stationarity property for the analysis.

V. RESULTS

We investigated whether the original and novel indexes of fractal dimension defined in this study allow identifying variations in the EEG signals reflecting true physiological changes related to sleep. To this end, we first visually inspected the ℓ_k curve of EEG segments belonging to different states. Fig. 2 shows that changes in EEG signals induced by sleep result in oscillations in the non-linear region of the ℓ_k curve. In particular, the state of sleep is characterized by a non-linear shape of the ℓ_k curve, which reflects a more oscillatory behaviour of the corresponding EEG signal with respect to the awake condition. In addition, visual inspection of features values for all subjects and EEG segments (Fig. 3) highlights different distributions in the awake and asleep states obtained for Freq_delta, Freq_beta, Higuchi and FD_tortuosity.

TABLE I

RESULTS OF THE ROC ANALYSIS. AREA UNDER THE CURVE (AUC) COMPUTED FOR EACH FEATURE.

Feature	AUC
Higuchi	0.97
Freq_beta	0.95
FD_tortuosity	0.88
Freq_delta	0.86
FD_residual	0.81
Freq_alfa	0.78
Freq_theta	0.62

A. Univariate ROC Analysis

Receiver operating characteristic (ROC) analysis was performed to test the capability of each feature to identify sleep. ROC curve and the associated area under the curve

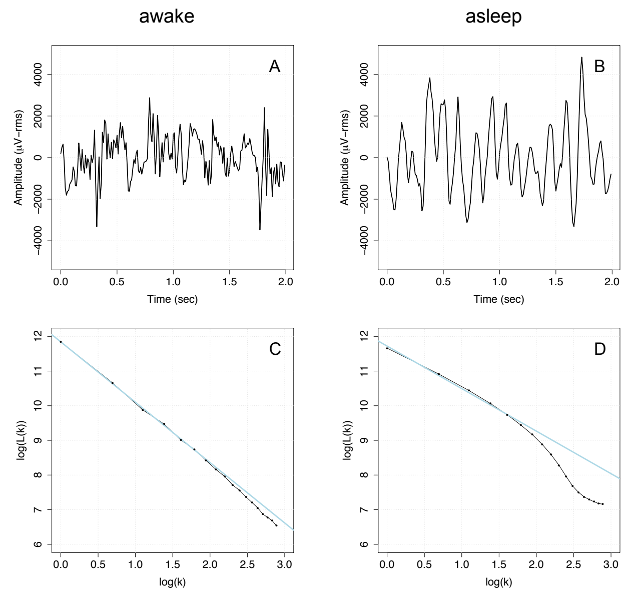

Fig. 2. Examples of EEG signals from the Sleep-EDF data set and corresponding Higuchi's curves ($k = 1, \ldots, 18$). The upper panels show EEG signals for the awake (A) and asleep (B) states. The lower panels (C and D) show the Higuchi's curve (black) and linear fit (light blue).

(AUC) were calculated for each feature (see Table 1). An AUC of 1 indicates perfect discrimination of the two classes. High AUC scores (Table 1) for Freq_delta and Freq_beta confirm the class separation shown in Fig. 3 and are in agreement with well-known findings in literature [14, 15]. The high AUC scores obtained by Higuchi (the highest AUC value amongst the considered features) and by FD_tortuosity demonstrate their effectiveness in extracting useful information with high discriminatory power from the EEG signal. FD_residual also obtained a high AUC score, but FD_tortuosity provided a more accurate class separation.

B. Multivariate Analysis

While univariate analysis reveals the ability of a single feature to discriminate between the asleep and awake states, a multivariate analysis can be used to measure the discrimination ability of the features combined together. A logistic regression model was here used to combine the considered features and classify between the asleep and awake states. For a fair model evaluation the EEG data set were split into a train set, for the identification of the logistic model, and a test set, for the assessment of its classification ability on "unseen" data. We repeated the assignment of data to the train and test set 100 times as in [16], to avoid bias in train/test set definition, and measured the classification performance through AUC. Fig. 3.H shows the AUC distributions obtained considering only the features defined in the frequency domain ("Freq"), adding the Higuchi's fractal dimension ("Higuchi") and also the two novel indices ("FDnovel"). Both models considering the features based on fractal dimension obtained higher AUC with respect to the one based only on the frequency-domain features (one-sided t-test, p-values always < 0.0001). In particular, the model

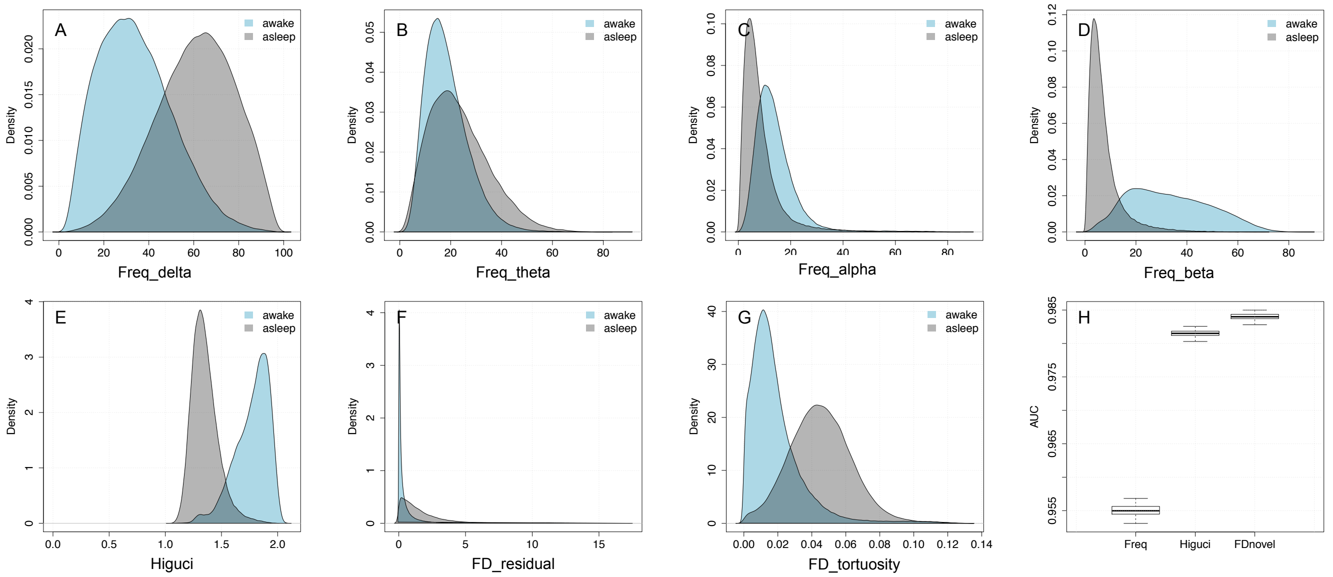

Fig. 3. (A-G) Feature distributions and results of the multivariate analysis. Density plots showing the distribution of the considered features in the awake and asleep states, considering all subjects. (H) Area under curve (AUC) scoring the classification performance obtained considering only the frequency-domain features ("Freq"), adding the Higuchi's fractal dimension ("Higuchi") or adding also the novel features based on fractal dimension ("FDnovel").

considering also the novel features obtained the highest AUC (p-values always < 0.0001), with an improvement from 0.955 median AUC value, for the basic "Freq" model, to 0.984.

VI. Conclusion and Discussion

Features extraction is an important step for the analysis of EEG data, aimed at mining and summarising useful information from EEG signals. In this work, we assessed the performance of techniques based on the fractal dimension to extract informative features to be used in sleep identification. First, we illustrated the relationship between EEG signals and the Higuchi's curve of fractal dimension and highlighted the importance of considering also its non-linear region. Second, we demonstrated the effectiveness of features based on fractal dimension for extracting valuable information form EEG data to be used in sleep identification. Finally, we defined two novel indices based on fractal dimension and demonstrated they improve sleep identification performance with respect to commonly used features for sleep classification. We are currently working on the implementation of a classifier based on artificial neural networks to consider non-linear feature combinations, to be employed in sleep identification and in a more challenging application of drowsiness monitoring.

References

[1] A. Y. Kaplan and S. L. Shishkin, "Application of the change-point analysis to the investigation of the brain's electrical activity," in *Non-Parametric Statistical Diagnosis*, pp. 333–388, Springer, 2000.

[2] A. S. Al-Fahoum and A. A. Al-Fraihat, "Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains," *International Scholarly Research Notices*, vol. 2014, 2014.

[3] M. Ronzhina, O. Janoušek, J. Kolářová, M. Nováková, P. Honzík, and I. Provazník, "Sleep scoring using artificial neural networks," *Sleep Medicine Reviews*, vol. 16, no. 3, pp. 251–263, 2012.

[4] A. Accardo, M. Affinito, M. Carrozzi, and F. Bouquet, "Use of the fractal dimension for the analysis of electroencephalographic time series," *Biological Cybernetics*, vol. 77, no. 5, pp. 339–350, 1997.

[5] G. Polychronaki, P. Ktonas, S. Gatzonis, A. Siatouni, P. Asvestas, H. Tsekou, D. Sakas, and K. Nikita, "Comparison of fractal dimension estimation algorithms for epileptic seizure onset detection," *Journal of neural engineering*, vol. 7, no. 4, p. 046007, 2010.

[6] M. Bachmann, J. Lass, A. Suhhova, and H. Hinrikus, "Spectral asymmetry and higuchi's fractal dimension measures of depression electroencephalogram," *Computational and mathematical methods in medicine*, vol. 2013, 2013.

[7] W. S. Pritchard and D. W. Duke, "Measuring chaos in the brain - a tutorial review of EEG dimension estimation," *Brain and Cognition*, vol. 27, no. 3, pp. 353–397, 1995.

[8] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena*, vol. 31, no. 2, pp. 277–283, 1988.

[9] B. Klinkenberg, "A review of methods used to determine the fractal dimension of linear features," *Mathematical Geology*, vol. 26, no. 1, pp. 23–46, 1994.

[10] F. Scarpa, X. Zheng, Y. Ohashi, and A. Ruggeri, "Automatic evaluation of corneal nerve tortuosity in images from in vivo confocal microscopy," *Investigative ophthalmology & visual science*, vol. 52, no. 9, pp. 6404–6408, 2011.

[11] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on audio and electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[12] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[13] A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," 1968.

[14] D. L. Schomer and F. L. Da Silva, *Niedermeyer's electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2012.

[15] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neuroscience & Biobehavioral Reviews*, vol. 44, pp. 58–75, 2014.

[16] M. Zanon, G. A. Susto, and S. McLoone, "Root cause analysis by a combined sparse classification and Monte Carlo approach," in *World Congress*, vol. 19, pp. 1947–1952, 2014.