New approach in features extraction for EEG signal detection

Carlos Guerrero-Mosquera and Angel Navia Vazquez

Abstract— This paper describes a new approach in features extraction using time-frequency distributions (TFDs) for detecting epileptic seizures to identify abnormalities in electroencephalogram (EEG). Particularly, the method extracts features using the Smoothed Pseudo Wigner-Ville distribution combined with the McAulay-Quatieri sinusoidal model and identifies abnormal neural discharges. We propose a new feature based on the length of the track that, combined with energy and frequency features, allows to isolate a continuous energy trace from another oscillations when an epileptic seizure is beginning. We evaluate our approach using data consisting of 16 different seizures from 6 epileptic patients. The results show that our extraction method is a suitable approach for automatic seizure detection, and opens the possibility of formulating new criteria to detect and analyze abnormal EEGs.

I. INTRODUCTION

In the last few years, EEG epileptic detectors have evolved including new techniques such as neural networks [1], non-linear models [2], independent component analysis (ICA) [3], Bayesian methods [4], support vector machines [5] and variance-based methods. These methods reinforce the nonstationary EEG concept and call for the necessity of extracting more information to understand the brain signals and its dynamics.

Other group of methods potentially useful for detecting and analyzing nonstationary signals are time-frequency distributions (TFDs) [6]. These methods allow us to visualize the evolution of the frequency behavior during some nonstationary event by mapping a one dimensional (1-D) time signal into a two-dimensional (2-D) function of time and frequency. Cohen [8] provided an overall class of TFDs based on the Wigner distribution and emphasized its importance in signal processing. This formulation led other researchers to introduce the concept of "kernel", and also provided an important model to obtain many different types of time-frequency distributions and applications in EEG, evoked potentials (EP), electrocardiogram (ECG), electrogastrogram (EGG), as well as in other signals of different nature, such as sounds and vibrations [9].

The TFDs offer the possibility to analyze relatively long continuous segments of EEG data even when the dynamics of the signal are rapidly changing. Taking the most of these, we could extract features from the time frequency plane. However, three considerations have to be taken. Firstly, a TFD will need signals as clean as possible for good results. Secondly, a good resolution both in time and frequency is necessary and as the "uncertainty principle" states, it is not possible to have a good resolution in both variables simultaneously. Thirdly, it is also required to sufficiently eliminate the spurious information (i.e. cross-term artifacts) inherent in the TFDs [8], [10].

The first consideration implies a good pre-processing stage to eliminate artifacts and noise. To improve resolution in a TFD for feature extraction, we propose to use the McAulay-Quatieri (MQ) sinusoidal model [12]. This is a robust tool used in speech processing for characterizing nonstationary speech signals from the short-time Fourier transform (STFT). We can tailor this technique to our detection task in EEG signals.

We will evaluate here if the proposed method is able to improve the resolution of TFDs for EEG signals, and give an objective approach to detect the onset of a neural brain event.

This paper is organized as follows. Section II introduces the preprocessing method that provides a clean EEG without artifacts, background and noise. It also explains the different techniques that comprise the detection method: windowing, the SPWV distribution (Smooth Pseudo Wigner-Ville), the MQ model applied to SPWV distribution and the feature extraction method. Section III shows the results of the seizure detection method applied to real EEG data and in Section V the principal conclusions of the paper are given and further work is proposed.

II. PROCESSING METHODS

The design of an EEG detection system comprises several tasks: acquisition, preprocessing to obtain a cleaner EEG signal, feature extraction and decision. The analysis of the EEG usually relies on windowing the signal using an sliding window. Each resulting segment is processed using time-frequency analysis (e.g. SPWV) and MQ sinusoidal analysis with the objective of extracting features and using them for the task of detection. We assume the existence of some wave in epileptic seizures from results obtained by others authors [6], [13] that have shown tracks along the time-frequency plane when there is a seizure. Our approach can be summarized in Fig. 1 and is detailed along this section.

A. Materials and settings: Raw EEG

The EEG records of 6 adult epileptic patients were obtained in a restful wakefulness stage recorded at the Clinica Universitaria de Navarra, Department of Neurophysiology (Pamplona, Spain). All of them contained focal epileptiform activity, according to experienced neurologists.
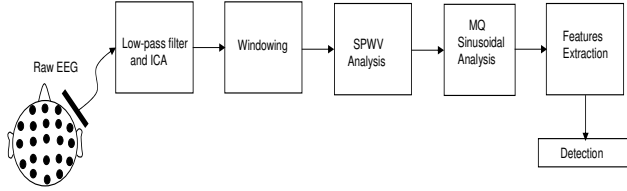
Fig. 1. A general detection scheme with the proposed method for detection of epileptic waves.

We used 16 EEG records of 24 min. length taken from 23-th and 25-th channels using the 10- 20 International System of Electrode Placement with additional anterotemporal electrodes T1/T2. In practice, raw EEG data were digitized at a sample rate of 200 Hz using a "DAD-32" equipment (La Mont Medical) and were filtered by a digital low-pass filter with cut-off frequency of 20 Hz. All computation has been carried out off-line in a Pentium III computer, using the Matlab (V.6) programming environment.

To get more generalization, we also used the dataset described in [7]. We created a problem detection called N1, consisting in two sets, normal and seizure, of EEG segments.

B. Low-pass filter and independent component analysis (ICA)

EEG activity can be severely contaminated by eye movements, blinking, muscle and heart artifacts, line noise, etc. The elimination of these artifacts demands a preprocessing stage. After sampling, the EEG signal can be modeled as a process $X(n)$ that relates the relevant activities as elementary waves, background activity, noise and artifacts [14]:

$$X(n) = F(n) + \sum_{i=1}^{n_p} P_i(n - t_{pi}) + \sum_{j=1}^{n_a} R_j(n - t_{aj}) + B(n) \quad (1)$$

where $F(n)$ is the background activity; the P_i terms represent brief duration potentials corresponding to abnormal neural discharges; the R_j terms are related to artifacts and $B(n)$ is the measurement noise which is modeled as a stationary process. Our goal is to obtain neural discharge information (i.e. P_i and t_{pi}) from the signal $X(n)$.

After preprocessing the EEG using such low-pass digital filtering it is necessary to separate artifacts such as muscle movements, eye blinks and other interfering activities without altering important information of abnormal activity. Taking these requirements into account, tools like ICA do permit to separate components in complex signals with the possibility of discriminating artifacts and brain waves. The technique used to calculate the independent components was the Joint Approximate Diagonalization of Eigen-matrices method (JADE). Artifacts were identified and visually eliminated on JADE components similarly as in a previous work [3], however we are also exploring automatic mechanisms for such elimination, such as those presented in [11].

If noise and artifacts are successfully eliminated, we can

approximate Eq.(1) as:

$$X(n) \approx F(n) + S(n) \quad (2)$$

where

$$S(n) = \sum_{i=1}^{n_p} P_i(n - t_{pi}) \quad (3)$$

C. Windowing

A quasi-stationary window refers to a period of time in which the EEG signal can be considered to be stationary. Although the time-frequency methods are oriented to deal with the concept of stationarity, increasing the data length has been found to reduce the degree of stationarity of EEG because in long windows we have more dynamic events come into play. As it is necessary to detect neural discharges with short time duration, the data length should be taken as short as possible. The preprocessed EEG signal was segmented using 5 secs. non-overlapping rectangular windows.

D. Smooth Pseudo Wigner-Ville distribution (SPWV) analysis

In a series of papers, Cohen generalized the definition of time-frequency distributions (TFDs) in such a way that a wide variety of distributions could be included in the same framework [8], [9]. Specifically the TFD of signal $x(n)$ is computed as:

$$P(t, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(\theta, \tau) \Phi(\theta, \tau) e^{-j\theta t - j\omega\tau} d\theta d\tau \quad (4)$$

where,

$$A(\theta, \tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(u + \frac{\tau}{2}) x^*(u - \frac{\tau}{2}) e^{j\theta u} du \quad (5)$$

is the so-called Ambiguity Function and the weighting function $\Phi(\theta, \tau)$ is called the kernel of the distribution. t and w are respectively time and frequency ("dummy") variables.

A variant called Smooth Pseudo Wigner-Ville (SPWV) distribution incorporates smoothing by independent windows in time and frequency, namely $W_w(\tau)$ and $W_t(t)$:

$$SPWV(t, \omega) = \int_{-\infty}^{\infty} W_w(\tau) \Big[\int_{-\infty}^{\infty} W_t(u - t) x(u + \frac{\tau}{2}) \\ x^*(u - \frac{\tau}{2}) du \Big] e^{-j\omega\tau} d\tau \quad (6)$$

Eq.(6) provides great flexibility in the choice of time and frequency smoothing, but the length of the windows should be determined empirically according to the type of signal analyzed and the required cross terms suppression.

E. MQ Sinusoidal analysis

In 1986, Robert McAulay and Thomas Quatieri proposed a new method for analysis/synthesis of continuous time speech signals which turned out to be a reconstruction process that provided a close approximation of the original signal [12]. EEG waves represent the combined activity of many neuronal cells which can generate sinusoidal-like oscillatory

waves. In this sense the EEG signal may be modeled as a collection of sinusoidal components of arbitrary amplitude, frequency and phase. Eq.(3) can then be written as:

$$S(n) = \sum_{\ell=1}^{L} A_\ell exp[jn\Psi_\ell] \qquad (7)$$

where A_ℓ and Ψ_ℓ respectively represent the amplitude and frequency of the ℓ-th component (out of L components (waves) conforming the EEG signal). Here amplitudes and frequencies are implicitly related to the P_i terms of Eq.(3). The problem now is to identify the terms A_ℓ and Ψ_ℓ in relation with neural discharges. The MQ algorithm performs this estimation by peak-matching based on the localization of peaks in energy on the SPWV plane. By linking peaks which occur at similar frequencies, we can define tracks along the t-f plane.

The concept of sinusoidal birth and death is used to account for the appearance or disappearance of spectral peaks between frames and the tracks are formed by connecting peaks between contiguous frames (see Fig. 2, upper). A new track is born if the frequency of a peak in the current frame does not appear in an interval $\pm\Delta$ of the peak frequency in the previous frame. Similarly, a track is dead when a peak in the current frame is not followed by another peak in the interval $\pm\Delta$ of frequency in the next frame. A magnitude condition is also imposed so that contiguous peaks at the same frequency which have large magnitude differences are proposed to belong to different tracks (the partials). The process of matching each frequency in frame t to some frequency in frame $t + 1$ is given in [12]. Fig. 2 shows the birth and death of frequency tracks formed by connecting peaks of similar frequencies between frames (upper) and the result of applying this method to an EEG seizure segment using the SPWV (bottom).

F. Feature extraction

The extraction of relevant information from an EEG signal is related to the information obtained with the MQ sinusoidal model. We propose to use three features based on energy, frequency and the length of the principal track. Fig. 2 (bottom) shows the existence of a principal track in the seizure corresponding to non-normal activity. Similarly in other EEG record with a duration of 75 secs. (see Fig. 3), we can observe a longer track F clearly visible during the seizure. These appreciations make it possible to introduce a new feature based on the duration of the principal track and use it in the detection task.

Apart from having the duration of the principal track, it also becomes necessary to measure other characteristics as energy and frequency to bring better information about this principal track. We have an EEG segmented into K segments, each EEG segment gives us the values E_k, F_k and L_k and we subsequently construct a three dimensional feature vector for each segment. The procedure to be applied to each segment is explained below.
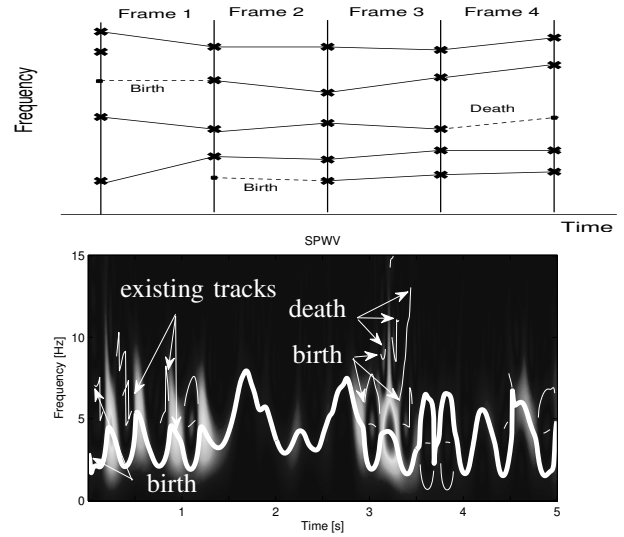

Fig. 2. Upper: Frequency matching process for determining frequency tracking in a TFD window. Each path in the graph is called a track. The birth of a track occurs when there is no partial in the previous frame to connect a peak in the current frame. Conversely, death occurs when a partial does not exist in the next frame to connect a peak in the current frame. Bottom: Peak-matching on the SPWV from a real EEG segment in a seizure. There is a principal track (largest length), marked with a thick line, and other minors tracks, marked with a thinner line. These tracks serve to summarize the spectral content on the time-frequency plane calculated by SPWV.
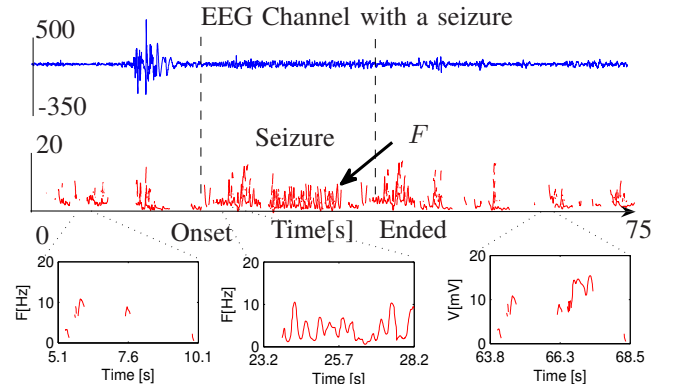

Fig. 3. Track extraction using a record with a seizure. The figure shows the EEG in time domain (upper) and time frequency plane using track extraction (middle). The length of the register is 75 secs. Taking zoom in a window (5 secs.) on three different EEG parts, we can observe how a dominant and sustained frequency F appears when there is a seizure, while tracks appear discontinued in the non-seizure periods.

1) Energy: Denoting $\vartheta_k(t, f)$ the time-frequency plane using the MQ sinusoidal model for segment k, where $\vartheta_k(t, f)$ only takes nonzero values on points corresponding to the identify tracks, it is possible to calculate energy as:

$$E_k = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \vartheta_k(t, f) dt df \qquad (8)$$

2) Frequency: Similarly, we calculate the frequency of segment k using the frequency marginal

$$F_k = \int_{-\infty}^{\infty} \vartheta_k(t, f) dt \qquad (9)$$

3) Length: L_k is the length of the largest track in the k-th EEG segment.

III. RESULTS

Table II shows the results of N1 problem detection and the 6 EEG records from patients that suffer the same type of focal temporal lobe epilepsy. We use EEG registers with the same size. Note that features change in different registers, suggesting that factors specific to individual patients play an important role in determining how accurate a seizure may be detected.

TABLE I

ANALYSIS OF DIFFERENT SEIZURES IN DIFFERENT PATIENTS.

EEG	L [secs.]	F [Hz]	E [%]
1	2	1.7	26
2	3	5.9	37
3	2.5	4.4	28
4	3.2	2.5	32
5	2.6	6.5	1
6	3	3.8	27
N1	0.6	4.3	4

Our analysis have shown that there are features related to epileptic activity. Taking the median values from the feature vector as threshold decision, we present the results of sensitivity and specifity, wich are defined as follows:

Sensitivity: Percentage of EEG segments containing abnormal activity correctly classified.

Specificity: Percentage of EEG segments not containing abnormal activity correctly classified.

The results based on these two measures are presented in Table II and Table III. Note the good performance of our method when we use different EEG data from the same patient (Table II) and how this performance is also good when we try to detect epileptic activity in different patients (Table III).

TABLE II

SENSITIVITIES AND SPECIFITIES OF EEG'S IN SAME PATIENT.

EEG	Sensitivity [%]	Specifity [%]
1	89	97
2	90	99
3	80	89
4	30	100
5	77	94

TABLE III

SENSITIVITIES AND SPECIFITIES OF EEG'S IN DIFFERENT PATIENTS.

Patient	Sensitivity [%]	Specifity [%]
1	72	99
2	88	93
3	56	97
4	90	99
5	66	100
6	30	100
N1	97	85

IV. CONCLUSIONS AND FUTURE WORKS

A new approach to identify abnormal discharges in epileptic EEG signals using track extraction is presented. Our results suggest that the track extraction method is a powerful tool for detecting in EEGs signals. The feature vector based on duration, frequency and energy $< E_k, F_k, L_k >$ obtained from track extraction in every frame is simple and useful for the detection task.

The initial results show that the track extraction is a suitable method for automatic seizure detection based on features extraction from TFDs, and opens the possibility of studying new criteria for detecting abnormal EEG. In our studies with patients that suffer from focal epileptic seizures, we found similar characteristics present in the seizures and we obtained a clear isolation of tracks in epileptic EEGs. This results present high-quality resolution and facilitates the extraction of features.

Future works implies the study of a wide range of machine learning methods to better exploit the features proposed here to finally obtain improved seizure detections.

REFERENCES

[1] N. Acir, I. Oztura, M. Kuntalp, B. Baklan, C. Guzelis, "Automatic detection of epileptiform events in EEG by three-stage procedure based on artificial neural networks,"*IEEE Trans. Biomed. Engineering*, vol. 52, 2005, pp. 30–40

[2] K. Lehnertz, C.E. Elger, "Spatio-temporal dynamics of the primary epileptogenic area in temporal lobe epilepsy characterized by neuronal complexity loss,"*Electroencephalography and Clinical Neurophysiol.*, vol. 95, 1995, pp. 108–117

[3] J. Iriarte, E. Urrestarazu, M. Valencia, M. Alegre, A. Malanda, C. Viteri, and J. Artieda, "Independent Component Analysis as a tool to eliminate artifacts in EEG: A quantitative study," *Journal of Clin. Neurophysiol.*, vol. 20, 2003, pp. 249–257

[4] S. Grewal, J. Gotman, "An Automatic warning system for epileptic seizures recorded on intracerebral EEGs,"*Clin. Neurophysiol.*, vol. 116, 2005, pp. 2460–2472

[5] B. Gonzalez, S. Sanei, J.A. Chambers, "Support vector machines for seizure detection,"*Proceedings of the IEEE ISSPIT*, 2003, pp. 126–129

[6] B. Boashash, H. Carson, M. Mesbah, "Detection of seizures in newborns using time-frequency of EEG signals,"*Proceedings of the Tenth IEEE workshop on Statistical Signal and Array Processing*, 2000, pp. 564–568.

[7] A.T. Tzallas, M.G. Tsipouras and D.I. Fotiadis, "The use of Time-frequency distributions for epileptic seizure detection in EEG recordings," *Proceedings of the IEEE EMBS*, 2007, pp. 3–6

[8] L. Cohen, *Time-Frequency Analysis*, Prentice Hall. Upper Saddle River, NJ, 1995.

[9] A. Papandreu-Suppappola, *Applications in Time-Frequency signal processing*, CRC Press. Tempe, Arizona, 2003.

[10] Piotr Jerzy Durka, *Time-Frequency analysis of EEG*, Thesis, Institute of Experimental Physics, Warsaw University, 1996

[11] Carlos Guerrero-Mosquera and Angel Navia Vazquez, "Automatic removal of ocular artifacts from EEG data using adaptive filtering and Independent Component Analysis," accepted for the *17th European Signal Processing Conference (EUSIPCO)*, 2009.

[12] R.J. McAulay, and T.F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoustic, Speech, and Signal Processing*, vol. 34, 1986, pp. 744 –754

[13] W.J. Williams, H.P. Zavery, J.C. Sackellares, "Time-Frequency Analysis in electrophysiology signals in epilepsy," *IEEE Engineering in Medicine and Biology*, vol. 14, 1995, pp.133–143

[14] L. Senhadji, F. Wendling,"Epileptic transient detection: wavelets and time-frequency approaches,"*Neurophysiol Clin.*, vol. 32, 2002, 175–192