33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

EEG Seizure Identification by using Optimized Wavelet Decomposition

RD. Pinzon-Morales[1], *Member, IEEE*, A. Orozco-Gutierrez[1], G. Castellanos-Dominguez[2]
[1] Universidad Tecnologica de Pereira, Colombia, [2] Universidad Nacional de Colombia, Manizales, Colombia

Abstract— **A methodology for wavelet synthesis based on lifting scheme and genetic algorithms is presented. Often, the wavelet synthesis is addressed to solve the problem of choosing properly a wavelet function from an existing library, but which may be not specially designed to the application in hand. The task under consideration is the identification of epileptic seizures over electroencephalogram recordings. Although basic classifiers are employed, results rendered that the proposed methodology is successful in the considered study achieving similar classification rates that had been reported in literature.**

I. Introduction

Wavelet analysis has been widely used in bioelectric signals including electroencephalogram (EEG), evoke–related potentials, microelectrode recordings, electrocardiogram, among others [1]–[3]. Although there are plenty of wavelet prototypes in the literature, there is not an established rule that states which wavelet should be used for each application. Instead, it is usual to test, more or less arbitrarily, different wavelet shapes to find the one that suits better.

In the present work, the application under consideration is the identification of EEG–related seizures, which are transient signs and/or symptoms of abnormal, excessive or synchronous neuronal activity in the brain in Epilepsy disorders. Conventionally, a routine EEG 20-minute recording of the patient's brain waves is analyzed to evaluate suspected seizures. Diagnostic difficulties arise when a patient has a suspected seizure (or a neurological event of unclear etiology) that is not obvious in the routine EEG. The current gold standard is the continuous EEG recording along with video monitoring of the patient, which usually requires patient admission. Besides, it is costly and not always available. So, an automatic system for EEG seizure identification would be of outstanding help in diagnosis.

To generate a wavelet function, suitable for the EEG–related seizure identification, several works, which have been proposed for wavelet synthesis including lifting schemes with deterministic and evolutionary, may be employed. On one hand, analytical methods of adaptive synthesis have shown to be burdensome, owing to the complexity of mathematical conditions on orthogonality, symmetry, compactness, and smoothness [2], [4]. On the other hand, evolutionary approaches, such as genetic algorithms (GAs), cultural algorithms, and ant systems had presented newsworthy features for wavelet design. For example, a completely new wavelet shape from the analyzed signal is designed in [5] using an ensemble of GA and Lagrange optimization. In this line of analysis, additional speed improvements to the GAs with the ant systems are introduced in [4], reporting better performance than existing wavelets for signal denoising.

In this work, the capabilities of the lifting schemes, which are devoted for both wavelet designing and performing the discrete wavelet transform, are presented. In addition, GAs are extended for wavelet synthesis. The designed wavelet function is assumed to exhibit feature spaces with maximum class separability due to the use of clustering validation measures as fitness functions within the GA-based optimization framework.

II. Materials and Methods

A. Wavelet Decomposition by Lifting Scheme

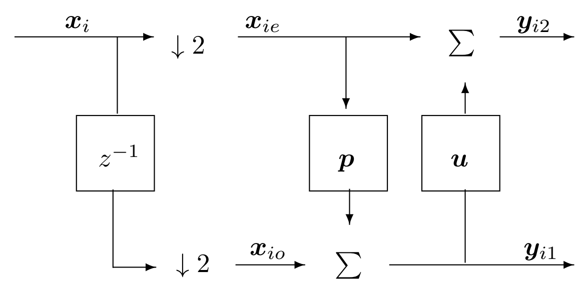

Fig. 1. One step decomposition with the Lifting Schemes

Given a set of samples from a discretized measured EEG recordings, $\{\boldsymbol{x}_m \in \mathbb{R}(T) : m = 1, \ldots, M\}$, each one of length T, then, the input observation matrix $\boldsymbol{Y} = [\boldsymbol{y}_1 \cdots \boldsymbol{y}_M]^\top$, $\boldsymbol{Y} \in \mathbb{R}^{M \times n_W}$ is described by both a $n_\xi-$dimensional feature set, $\boldsymbol{\Xi} = \{\boldsymbol{\xi}_i : i = 1, \ldots, n_\xi\}, \boldsymbol{\Xi} \in \mathbb{R}^{M \times n_\xi}$ and a label set $\boldsymbol{c} = \{c = 1, \ldots, n_C\}$. Each row of the input object matrix is accomplished by using wavelet–based decomposition, that is, $\boldsymbol{y}_m = \mathcal{W}\{\boldsymbol{x}_m\}$, $\boldsymbol{y}_m \in \mathbb{R}^{1 \times n_W}$. In this work, the lifting–scheme (LS) wavelet structure (shown in Fig. 1) is selected because of its flexibility and fast implementation of the wavelet transform, which not only allows the analysis of signals but also the design of any bi–orthogonal wavelet [6].

Implementation of LS involves the following three steps: division, prediction, and update. In the division step, every input observation signal $\boldsymbol{y}_i^{(l)} = \{y_i(n) : n = 1, \ldots, n_W\}$ is split into even samples $\boldsymbol{y}_{ie}^{(l)} = \{y_i(2n) : n = 1, \ldots, n_W/2\}$, as well as into odd samples $\boldsymbol{y}_{io}^{(l)} = \{y_i(2n-1) : n = 1, \ldots, n_W/2\}$ at scale l. This procedure is also referred as *lazy wavelet*. Then in the prediction step, the vector $\boldsymbol{p}$ is convolved with $\boldsymbol{y}_{ie}^{(l)}$ to predict $\boldsymbol{y}_{io}^{(l)}$ and for eliminating low order polynomials from $\boldsymbol{y}_i$, and thus, obtaining the detail symbols at scale $l+1$, $\boldsymbol{y}_{i2}^{(l+1)} = \{y_{i2}^{(l+1)}(n) : n = 1, \ldots, n_W/(2^{(l+1)})\}$, which are described as follows:

$$y_{i2}^{(l+1)}(n) = y_{io}(n) - \sum_{r=-n_p/2+1}^{n_p/2} p(r) y_{ie}^{(l)}(n+r),$$

where $\boldsymbol{p} = \{p(r) : r = 1, \ldots, n_p\}$ are coefficients of prediction. The super–index (l) indicates the decomposition level, where $l = 0$ is the input EEG pattern, that is, $\boldsymbol{y}_i^{(0)} = \boldsymbol{x}_i$, and n_p is the order of $\boldsymbol{p}$. Next, in the update stage, an update on the even samples $\boldsymbol{y}_{ie}^{(l)}$ is accomplished by using the update vector $\boldsymbol{u}$ on the previous symbols $\boldsymbol{y}_{i2}^{(l+1)}$, and adding the result to $\boldsymbol{y}_{ei}^{(l)}$.

The update sequence $\boldsymbol{y}_{i1}^{(l+1)} = \{y_{i1}^{(l+1)}(n) : n = 1, \ldots, n_W/(2^{(l+1)})\}$ can be seen as rough view of $\boldsymbol{y}_i$,

$$y_{i1}^{(l)}(n) = y_{ie}^{(l)}(n) - \sum_{j=-n_u/2}^{n_u/2-1} u(j) y_{i2}^{(l)}(n+j-1),$$

where $\boldsymbol{u} = \{u(j) : j = 1, \ldots, n_u\}$ are update coefficients, and n_u is the order of $\boldsymbol{u}$.

Grounded on the LS, it is possible to implement the multi-resolution analysis by applying iteratively the above described LS steps up to scale l over both detail and approximation coefficients. In particular, a singular basis that can be obtained is the dyadic, commonly used in the discrete wavelet transform yielding only $n_d = l+1$ wavelet nodes with fixed logarithmic *t-f* resolution.

B. Optimization of Wavelet Function

The optimization of the wavelet function relies on the customization of both LS vectors, $\boldsymbol{u}$ and $\boldsymbol{p}$, under conditions of compact support, symmetry, linear phase, and bi–orthogonality, yielding wavelet functions with unique *t-f* features. Furthermore, by doing so the resulting object matrix $\boldsymbol{Y}$ is optimized to exhibit the desire characteristics such as maximum class separability [1] or auto–similarity preservation [3]. Taking into account that the present application is a classification–oriented task, the proposed optimization procedure depicted in Fig. 2, which is based on GA, incorporates clustering validation measures to generate wavelets that permit the construction of signal representations with maximum class separability.

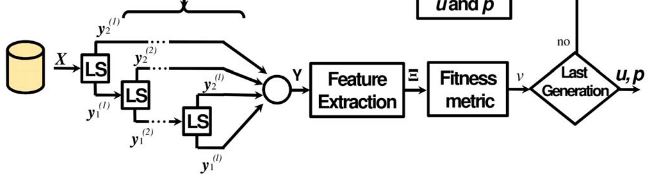

Fig. 2. Evolutionary–based optimization procedure for wavelet synthesis. Notation LS stands for an elemental circuit implementing lifting scheme.

As suggested in [7], to secure the compact support, symmetry, linear phase, bi-orthogonality of the resulting wavelet functions, both the symmetrical linear phase and normalization constraints are introduced, which are also embedded into the lifting formalism [6]:

$$p(r) = p(r+1), \quad \sum\nolimits_{r=1}^{n_p/2} p(r) = 1/2, \tag{1a}$$

$$u(j) = u(-j+1), \quad \sum\nolimits_{j=1}^{n_u/2} u(j) = 1/4 \tag{1b}$$

Once the LS vector orders, n_u and n_p, are fixed, the GA must evolve only $(n_u/2-1)+(n_p/2-1)$ values.

Besides, the proposed methodology involves clustering validation metrics into the GA–procedure. Thus, incorporating those measures into the LS intends that the resulting discriminant representation maximizes the separability among patterns to be analyzed. In particular, the *The Dunn's Index* is considered here that is based on geometrical considerations and defined as follows [8]:

$$v_{DI}(\boldsymbol{c}, \boldsymbol{\xi}) = \min_{1 \le i \le c} \left\{ \min_{1 \le j \le c, i \ne j} \left\{ \frac{\delta(\boldsymbol{\xi}(i,l), \boldsymbol{\xi}(j,l))}{\max_{\forall c}\{\Delta(\boldsymbol{\xi}(c,l))\}} \right\} \right\} \tag{2}$$

where $\Delta(\boldsymbol{\xi}(c,p)) = \max_{j,i \in l}\{\|\xi(c,i) - \xi(c,j)\|_f\}$ is the diameter of ith class, and $\delta(\boldsymbol{\xi}(c,i), \boldsymbol{\xi}(c,j)) = \min_{\forall l}\{\|\boldsymbol{\xi}(c,i) - \boldsymbol{\xi}(c,j)\|_f\}$ is the distance in terms of $f-$norm between ith and jth classes.

C. Construction of the Discriminant Feature Set

Once the wavelet–decomposition is assessed, the feature matrix Ξ can be constructed by a library of measures taken form the wavelet nodes. The selection of such measures should be carried out accordingly to the properties that are to be highlighted from the biological phenomena. The present study considers the following set of morphological metrics that had been already tested in EEG seizure identification rendering acceptable performance [9]: i) the mean of the absolute value $\xi_{i1}(\boldsymbol{y}_{i1}) = \boldsymbol{E}\{|y_{i1}| : \forall i\}$, ii) the average power $\xi_{i2}(\boldsymbol{y}_{i1}) = \boldsymbol{E}\{y_{i1}^2 : \forall i\}$, iii) the standard deviation $\xi_{i3}(\boldsymbol{y}_{i1}) = \boldsymbol{E}\{\|(y_{i1} - \bar{y_1})\|^2 : \forall i\}$, and iv) the ratio of the absolute mean values of adjacent bands $\xi_{i4}(\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2}) = \boldsymbol{E}\{|y_{i1}|\}/\boldsymbol{E}\{|y_{i2}|\} : \forall i$, where $\boldsymbol{y}_{i1}$ and $\boldsymbol{y}_{i2}$ are two wavelet nodes with adjacent frequency bands. Notation $\boldsymbol{E}\{\cdot\}$ stands for expectation operator. So, the feature vector is constructed by taking the aforementioned features from each wavelet node in $\boldsymbol{Y}$, thus, $\Xi = \{\boldsymbol{\xi}_{i1}, \boldsymbol{\xi}_{i2}, \boldsymbol{\xi}_{i3}, \boldsymbol{\xi}_{i4} : i = 1, \ldots, 4\} \in \mathbb{R}^{M \times n_\xi}$, where $n_\xi = 4n_d$.

D. Algorithm for Signal Analysis

The algorithm enclosing the above procedure for signal analysis holds the following steps:

- **A**: Initialize $(n_u/2-1)+(n_p/2-1)$ values of the LS vectors to be optimized, i.e. $\boldsymbol{p}$ and $\boldsymbol{u}$, with random values within the interval $[-1, 1]$. The remain values, $(n_u/2+1)+(n_p/2+1)$, are calculated using the linear and normalization constrains: Eq. (1a) and Eq. (1b), respectively.
- **B**: Extract a subset $\boldsymbol{x}_t \in \mathbb{R}(T), t = 1, \ldots, M_t$, of input signals $\boldsymbol{x}$, from the database from each class of $\boldsymbol{c}$, $M_t < M$ (subindex t stands for training set).

C: Decompose each pattern with the LS up to level l, using the dyadic decomposition yielding n_d wavelet nodes that conform $\boldsymbol{Y}_t$,

D1: Perform M_t feature extractions to device the feature vector Ξ_t,

D2: Compute the clustering validation measure $v_{DI}(\Xi_t, \boldsymbol{c}_t)$ and integrate it into the GA optimization as fitness function.

F: Evolve the GA until the fixed number of iterations and, finally, provide optimized LS vectors $\boldsymbol{p}$ and $\boldsymbol{u}$.

III. EXPERIMENTS AND RESULTS

A. Database

A public EEG dataset, which is available and described in [10] and includes recordings for both healthy and epileptic subjects, is used. The complete data set embraces five sets (denoted A–E) each containing $100E$ single-channel EEG segments, each one having 23.6 s duration. The data were digitized at 173.61 samples per second using 12 bit resolution and they have the spectral bandwidth of the acquisition system, which varies from 0.5 Hz to 85 Hz. Sets A and B consisted of segments taken from five healthy volunteers with eyes open and closed, respectively. Sets C, D, and E are originated from EEG archive of presurgical diagnosis. Sets C and D contain only activity measured during seizure free intervals, whereas the set E only contain seizure activity.

In this paper, we study the following four different medical classification cases created from the above described dataset:

I: Normal and seizures recordings are classified, namely, A-type EEG segments and the seizure class including the E-type, respectively.

II: All the EEG segments are used and classified into two different classes: A, B, C, and D types are included in the first class and type E in the second.

III: A three-class task, that is, normal, seizure-free and seizure. The normal class includes only the A-type EEG segments, the seizure-free class the D-type EEG segments, and the seizure class the E-type.

IV: All the EEG segments are used classified into three different classes: A and B types of EEG segments are combined to a single class, C and D types are also combined to a single class, and type E is the third class. Case IV is the one closest to real medical applications including three categories; normal (i.e., types A and B), seizure-free (i.e., types C and D) and seizure (i.e., type E).

B. Off-line Optimization of Wavelet Function

The purpose of the off-line optimization, involving the design of $\boldsymbol{u}$ and $\boldsymbol{p}$, is the synthesis of a completely new application-oriented mother wavelet, based on the GA procedure, as described in Sec. §II-D. Furthermore, the new mother functions are expected to generate wavelets with maximum class separability in the v_{DI} sense. In the concerning case, $n_u = n_p = 6$, so that the high and low frequencies are weighted similarly, $c = 5$, and the fraction value M_t is fixed to 30, as studied in a previous work [1].

Regarding the GA, the following five parameters must be selected: (i) the arithmetic operator, (ii) the mutation operator, (iii) the population scale, (iv) the number of generations, and (v) the bounds of iteration. Parameters (i) to (iv) can be set directly from previous literature. Besides, the arithmetic crossover and no uniform mutation operators are employed, as recommended in [5]. And for the sake of simplicity, the population scale is set to be 30, whereas the number of generations is set equal to 20. Accordingly, the working iteration parameters of the GA are selected to range within the interval $[-5, 5]$, which is regarded to the possible values for the predictor and update coefficients that meet the normalization constraints, i.e, Eq. (1a) and Eq. (1b).

At the end, the chromosome length, during the GA procedure, is $2(N/2-1) = 4$. To avoid local minimum in the GA convergence, above steps are repeated ten times with random initializations. Finally, the best mother wavelet is selected as the one with the higher v_{DI} value. The temporal response of the wavelets functions along with the filter frequency response associated to operators $\boldsymbol{u}$ and $\boldsymbol{p}$, are shown in Fig. 3, for the considered clustering validation measure. As seen, results show explicitly the customized temporal and frequency dynamic that can be achieved through the optimization methodology. It is important to emphasizes that achieved wavelet dynamics is barely reached by classical wavelets since those are not suppose to be designed to the current application.

C. Validations Strategy and Classifier

Considering that the core of the proposed study is not the classification stage, but the characterization, basic classifiers are employed during the EEG seizure identification stage, i.e., linear Bayes classifier (LDC) and *k*-nn classifier, because of its simplicity of implementation. The *k*-fold cross–validation approach is used to manage the database. Therefore, data is randomly divided into two subsets: the training and validation. The former set comprises 30% patterns, while the remaining 70% patterns are related to the latter set. There is no overlapping between sets. Ten folds are randomly generated for each experiment. Classification performance is given in terms of the rate of observations correctly classified.

Results presented in Table I shown that the proposed methodology is successful on the identification of EEG seizures regarding the four cases under consideration. It is important to notice that mentioned results are similar to those presented in literature without using elaborated classifiers, what is more, employing such classifier along with the proposed methodology may increase the classification rate above reported results [11]. In general, it is also appreciable that the *k*-nn classifier yielded better performance that the linear classifier, although, the deviation is just about 1%.

Globally, the post-processing with PCA after the wavelet decomposition reduces dramatically the dimensionally of the wavelet–based feature space. Indeed, PCA can be employed due to the orthogonal representation yielded by the wavelet decomposition with a minor loss in classification performance. Fig. III-C illustrates the fact that PCA convergence

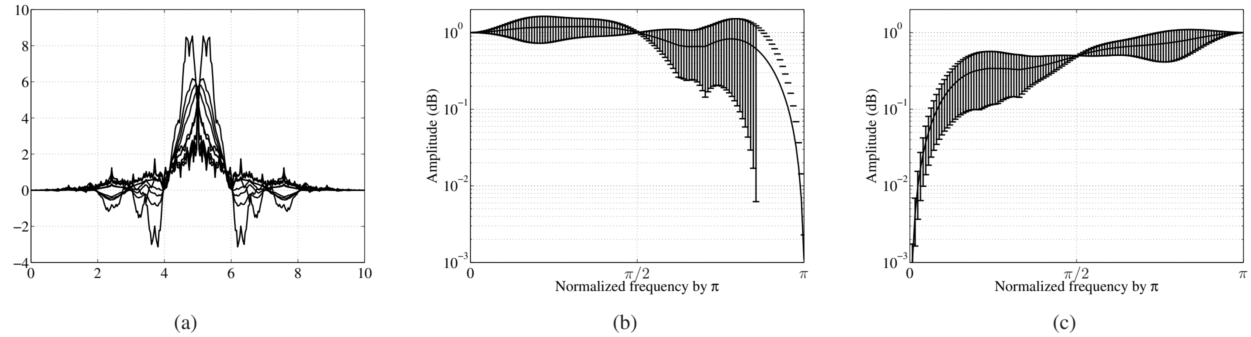

Fig. 3. Temporal response of the wavelet functions associated to the optimized filters $\boldsymbol{u}$ and $\boldsymbol{p}$, along with their frequency response. Results are given considering ten runs of the GA-based optimization with random initializations.

TABLE I

CLASSIFICATION PERFORMANCE. BOLD VALUES ARE RELATED WITH THE HIGHEST SCORE OBTAINED FOR THE CURRENT EXPERIMENT

	Without post–processing		With post–processing	
	k-nn	LDC	*k*-nn	LDC
Case I	99,83±0,5	97,83±1,7	(1)99,83±0,5	(16)97,83±1,7
Case II	97,47±0,7	96,00±2,1	(5)**97,80±0,7**	(10)**96,47±1,9**
Case III	95,00±1,5	86,33±3,4	(2)**95,22±1,9**	(16)**87,00±3,4**
Case IV	95,33±1,0	90,87±2,5	(11)**95,87±0,8**	(15)**94,07±1,6**

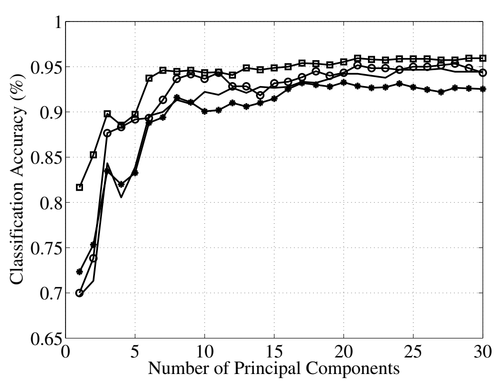

Fig. 4. (Right) Number of PCA components against classification performance for the dyadic decomposition. (Left) Number of PCA components against classification performance for the binary decomposition.

is achieved with few principal components (less than five). The reduction achieved is about 90% (from $\mathbb{R}^{20} \rightarrow \mathbb{R}^{5}$).

IV. CONCLUSIONS

This paper has presented a methodology for synthesis of wavelet functions for EEG seizure identification. It introduces several improvements. Firstly, construction of unique mother wavelets from the signal's information itself with adaptable spectral characteristics by mean of genetic algorithms and lifting scheme. Secondly, incorporation of clustering validation measures that allow the creation of feature spaces with maximum class separability form the wavelet coefficients. Thirdly, low dimensional representation by employing principal components analysis. Finally, high classification accuracy is achieved in four different clinical scenarios using basic classifiers.

ACKNOWLEDGMENT

This research is carried out under grant "*Identificación Automática del Trastorno por Déficit de Atención y-o Hiperactividad sobre registros de Potenciales Evocados Cognitivos*", founded by COLCIENCIAS.

REFERENCES

[1] R.-D. Pinzon-Morales, A. Orozco, and G. Castellanos-Dominguez, "Novel signal-dependent filter bank method for identification of multiple basal ganglia nuclei in parkinsonian patients," *Neural Engineering*, vol. (In press), 2011.

[2] R. C. Guido, S. Barbon, L. S. Vieira, F. Lopes, C. Dias, J. Pereira, P. Scalassara, and E. S. Fonseca, "Introduction to the discrete shapelet transform and a new paradigm: Joint time-frequency-shape analysis," in *Circuits and Systems (ISCAS 2008) IEEE International Symposium on*, 2008, pp. 2893 –2896.

[3] R.-D. Pinzon-Morales, A. Orozco, G. Castellanos-Dominguez, and R. Capobianco, in *Neural Engineering, 2011. NER '11. 5th International IEEE/EMBS Conference on*, 2011.

[4] P. Arpaia, C. Romanucci, and A. Zanesco, "Optimal design of wavelet filters based on lifting scheme by means of cultural algorithms," in *Instrumentation and Measurement Technology Conference (IMTC 2006) Proceedings of the IEEE*, 2006, pp. 1245 –1250.

[5] L. Zhen, H. Zhengjia, Z. Yanyang, and W. Yanxue, "Customized wavelet denoising using intra- and inter-scale dependency for bearing fault detection," *Journal of Sound and Vibration*, vol. 313, no. 1-2, pp. 342 – 359, 2008.

[6] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *Applied and Computational Harmonic Analysis*, vol. 3, no. 2, pp. 186 – 200, 1996.

[7] A. Gouze, M. Antonini, M. Barlaud, and B. Macq, "Design of signal-adapted multidimensional lifting scheme for lossy coding," *Image Processing, IEEE Transactions on*, vol. 13, no. 12, pp. 1589–1603, Dec. 2004.

[8] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 28, no. 3, pp. 301–315, 1998.

[9] A. Subasi, "Eeg signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, 2007.

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

[11] A. Tzallas, M. Tsipouras, and D. Fotiadis, "Automatic seizure detection based on time-frequency analysis and artificial neural networks," *Intell. Neuroscience*, vol. 2007, pp. 18:1–18:13, August 2007.