We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

13

Time-Frequency Based Feature Extraction for Non-Stationary Signal Classification

Luis David Avendaño-Valencia, Carlos Daniel Acosta-Medina
and Germán Castellanos-Domínguez
Universidad Nacional de Colombia, Sede Manizales
Colombia

1. Introduction

Biosignal recordings are useful for extracting information about the functional state of an organism. For this reason, such recordings are widely used as tools for supporting medical decision. Nevertheless, reaching a diagnostic decision based on biosignal recordings normally requires analysis of long data records by specialized medical personnel. In several cases, specialized medical attention is unavailable, due to the high quantity of patients and data to analyze. Besides, the access to this kind of service may be difficult in remote places. As a result, the quality of medical service is deteriorated. In this sense, automated decision support systems are an important aid for improving pathology diagnosis and treatment, specially when long data records are involved. The successful performance of automatic decision systems strongly depends on the adequate choice of features. Therefore, non-stationarity is one of the most important problems to take into account. Non-stationarity is an inherent property of biosignals, as the underlying biological system has a time dependant response to environmental excitations. Specially, changes in physiological conditions and pathologies may produce significant variations.

It has been found that non-stationary conditions give rise to changes in the spectral content of the biosignal (Hassanpour et al., 2004; Quiceno-Manrique et al., 2010; Sepúlveda-Cano et al., 2011; Subasi, 2007; Tarvainen et al., 2009; Tzallas et al., 2008). Therefore, time-frequency (*t–f*) features have been previously proposed for examining the dynamic properties of the spectral parameters during transient physiological or pathological episodes. It is expected that *t–f* features reveal the correlation between the *t–f* characteristics of abnormal non-stationary behavior (Debbal & Bereksi-Reguig, 2007). For this reason, *t–f* methods should outperform conventional methods of frequency analysis (Tzallas et al., 2008). Besides, *t–f* features are expected to behave slowly enough along the time axis, so the usual stationary restrictions imposed on short-term intervals should work out better than when straightforwardly analyzing over the raw input biosignal.

Among *t–f* features, *time-frequency representations* (TFR) are one of the most common and widely used features. TFRs are the most complete characterization method for non-stationary biosignals, as they display the energy distribution of a signal in time and frequency domains. Several TFR estimators have been proposed, which can be classified as non-parametric (linear and quadratic) and parametric (Marchant, 2003; Poulimenos & Fassois, 2006). The selection of

a particular TFR estimator should be led by its estimation properties, the expected precision and the computational resources. *Time–frequency dynamic features* (TFDF) based on spectral sub-band methods, summarize *t–f* information in a compact fashion. TFDF set has also demonstrated its capability for discriminating between normal and pathological patterns (Quiceno-Manrique et al., 2010; Sepúlveda-Cano et al., 2011). In this sense, the TFDF set can be suitably estimated by filter bank methods, such as sub-band spectral centroid, sub-band spectral centroid energy, linear cepstral coefficients, or discrete wavelet transform. These features have a lower dimensionality than the original TFR, but they still contain a stochastic dependance along time axis that has to be considered in the analysis.

In order to make pattern recognition problems solvable, it is necessary to convert *t-f* patterns into feature vectors, which become condensed representations, ideally containing only relevant information. From the viewpoint of managing large quantities of data, it is even more useful if irrelevant or redundant attributes could be segregated from relevant and important ones (Cvetkovic et al., 2008). A direct approach is to use linear transform methods, such as principal component analysis (PCA) or partial least squares (PLS), to reduce the feature space dimensionality resulting from the *t-f* plane. A reduced feature set obtained by PCA can be effectively accommodated to loosely structured TFR planes, when quantitative and decision based analysis is required (Bernat et al., 2007; Englehart et al., 1999). However, for classification purposes, the obtained components are not always related to the most discriminative information, as their transformation is based only on feature variability while class labels are neglected. Supervised methods, like PLS, can be used to improve performance of linear transform methods taking into account label variability as well as feature variability (Avendano-Valencia et al., 2010).

Nonetheless, computing transformation matrices of linear transform methods becomes more expensive as the dataset dimension increases. This is the case of *t–f* patterns, which may contain thousands of data points. Thus, it might be necessary to select *a priori* a confined portion of relevant data from the *t–f* feature to achieve computational stability during the feature extraction process. Classification based on local regions of the *t–f* plane has achieved higher success rates than those based on the entire *t–f* plane (Tzallas et al., 2008), but there is a significant unsolved issue associated with local-based analysis, which is the selection of the size and location of relevant regions. As a result, the choice of the feature extractor in the *t–f* domain is highly dependent on the final application (Sejdic et al., 2009). Relevance analysis is a tool that may serve to select the most informative *t–f* features from a discriminative viewpoint. In relevance analysis, a relevance measure is defined to determine the dependence of the features with regard to the each class label. The application of relevance measures has demonstrated a significant improvement compared to when no relevance measure is used, as proven in the case of TFR based classification (Avendaño-Valencia et al., 2010), and in frequency sub-band feature based classification (Sepúlveda-Cano et al., 2011).

Despite the large efforts in designing appropriate *t-f* based feature extraction for biosignal classification, there are still some issues to be solved. Particularly, as time-frequency data has a natural order, feature extraction methods can benefit if the data is treated as stochastically dependent, thus capturing the evolving information of the structure. A former approach was to use multidimensional linear transform methods, separately taking into account time and frequency dependence. When PCA transformations are used, these methods are known as two dimensional PCA (2D-PCA) (Yang et al., 2004). Two dimensional methods can be extended for different kinds of linear transforms, as in the case of PLS

(Avendano-Valencia et al., 2010), with improved classification performance. Nonetheless, these methods do not impose a priori a dynamic structure for the *t–f* feature ordering.
In summary, both TFR and TFDF are features that describe the non–stationary behavior of a biosignal and allow the analysis of different kinds of dynamic behavior. These features should contain all the available information for discriminating among different classes. Nevertheless, from the pattern recognition point of view, this class of features is troublesome to use, due to two important aspects: *multiple dimensions* and *high dimensionality*. In contrast to conventional characterization methods, TFR and TFDF are multidimensional features. Thus, while a conventional feature vector lies on $\mathbb{R}^N$ vector space, TFR and TFDF lie on $\mathbb{R}^{N\times T}$ matrix space. A multidimensional nature is important for describing temporal dynamics and relationships between spectral bands and should be exploited by the pattern recognition framework. Moreover, the intrinsic dimensionality of TFR and TFDF features is very high, and can normally lead to thousands of feature points. In that case, the performance of the classifier is compromised, as stated by *the curse of dimensionality*. So, in order to decrease the computing and storage requirements, as well as improving the generalization capabilities of the classifier, it is mandatory to reduce the dimensionality of TFR and TFDF features. Therefore, the dimensionality reduction approach should encompass with the following properties:

– Should be able to exploit the multidimensional nature of the data, taking into account both temporal dynamics and relationships between spectral bands.
– Should be able to extract the most informative data and fully describe the multidimensional features into a small set of features, as well as avoiding all irrelevant information.

This chapter is devoted to the review and comparison different time-frequency based feature extraction methods for classifying non-stationary biosignals. Subsequently, based on previous work in (Avendano-Valencia et al., 2010), a new methodology is proposed, oriented to reducing the dimensionality of *t–f* based features. The proposed methodology consecutively performs both the selection of the most relevant features as well as a linear transform of the *t–f* planes. Initially, the most relevant features extracted from the TFR are selected using a relevance measure that selects the most discriminative *t–f* features. Therefore, both the irrelevant information and the computational burden are significantly decreased. Then, the data is projected into a lower dimensional subspace using linear transform methods. Special attention will be drawn to different forms of improving linear transform methods for the case of dynamic *t-f* features. For the sake of comparison, conventional state of the art approaches for *t–f* based classification, and proposed linear transform and relevance analysis methods are included in this study.
Comparison is made with reference to the problem of identifying normal, inter-ictal and ictal records from publicly available non-stationary electroencephalographic biosignal database for detection of epilepsy (Andrzejak et al., 2001). Most representative non-parametric and parametric TFR estimators are under comparison: short time Fourier transform (STFT), smoothed pseudo Wigner-Ville distribution (SPWVD), and parametric TFR based on smoothness priors time varying autoregressive moving average models (SP–TAR). Spectral sub-band and discrete wavelet transform are also considered in the analysis. The accuracy, computational effort and ease of use are analyzed to check the advantages and drawbacks of conventional and proposed *t–f* based classification approaches.

2. Materials and methods

2.1 Time-Frequency Representations and Time-Frequency Dynamic Features

Time-Frequency Representation is a joint representation of the energy distribution of a signal in both time and frequency domains. The TFR of a signal $x_i(t) \in \boldsymbol{L}_2(\mathbb{R})$ is mathematically represented as a function of time and frequency $X_i(t,\omega) \in \boldsymbol{L}_2(\mathbb{R}^2)$. TFR should be used when there is evidence of time-varying or non-stationary conditions on the signal. In such cases, the time or the frequency domain descriptions of the signal alone cannot provide comprehensive information for analysis and classification, thus *t–f* methods should outperform conventional analysis methods (Sejdic et al., 2009; Tzallas et al., 2008).

In the past, different forms of estimating TFR have been proposed. These estimation methodologies can be grouped into the following main approaches (Marchant, 2003; Poulimenos & Fassois, 2006): (*i*) non–parametric TFR (*linear TFR* and *quadratic -Cohen class-TFR*); and (*ii*) *parametric TFR*.

Linear TFR makes use of *t–f* functions derived from translating, modulating and scaling a basis function with a definite time and frequency localization (Sejdic et al., 2009). Thus, for a signal $x_i(t)$, the TFR is given by,

$$X_i(t,\omega) = \int_{-\infty}^{\infty} x_i(\tau)\phi^*_{t,\omega}(\tau)d\tau = \langle x_i(t), \phi_{t,\omega}\rangle \tag{1}$$

where $\phi_{t,\omega}$ represents the basis function which defines the specific transform method, and * represents the complex conjugate. The basis functions are assumed to be square integrable, this is $\langle \phi_{t,\omega}, \phi_{t,\omega}\rangle^{1/2} < \infty$. Short time Fourier transform – STFT, wavelets, and matching pursuit approaches are typical examples of this class of transforms.

Quadratic TFR are defined as the Fourier transform of the local autocovariance function, given by the product $x_i(t+1/2\tau)x_i^*(t-1/2\tau)$. Also, different *t–f* kernels might be used in order to attain the desired properties on the transform. Thus, a quadratic TFR is defined by,

$$X_i(t,\omega) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_i\left(u+\frac{1}{2}\tau\right)x_i^*\left(u-\frac{1}{2}\tau\right)\phi(\theta,\tau)e^{-j\theta t - j\tau\omega + j\tau u}dud\tau d\theta \tag{2}$$

where $\phi(\theta,\tau)$ is a two-dimensional kernel function, defining the specific representation and its properties. Wigner-Ville, Choi-Williams, and spectrogram, are some exemplary methods belonging to this category.

Parametric TFR are based on a parametric signal model, like a TARMA model. The model should be fitted to the analyzed signal, and then, the TFR can be derived from the time-varying parameters and residual variance (Poulimenos & Fassois, 2006). A TARMA(n_a, n_c) model is defined as,

$$x_i(t) = -\sum_{n=1}^{n_a} a_{i,n}(t)x_i(t-n) + \sum_{n=1}^{n_c} c_{i,n}(t)e_i(t-n) + e_i(t), \quad e_i(t) \sim \mathbf{N}(0, \sigma^2_{e_i}(t)) \tag{3}$$

where $a_{i,n}(t), n = 1,\ldots,n_a$ and $c_{i,n}(t), n = 1,\ldots,n_c$ are the time-varying AR/MA parameters, and $e_i(t)$ is the residual sequence assumed as zero mean white gaussian noise with time-dependant variance $\sigma^2_{e_i}(t)$, noted as $\mathbf{N}(0, \sigma^2_{e_i}(t))$. The parametric TFR derived from

TARMA model is defined as,

$$X_i(t,\omega) = \left| \frac{1 + \sum_{n=1}^{n_c} c_{i,n}(t) e^{-j\omega n}}{1 + \sum_{n=1}^{n_a} a_{i,n}(t) e^{-j\omega n}} \right|^2 \sigma_e^2(t) \tag{4}$$

which would be the power spectral density of the signal if the system were made stationary at time instant t. The effectiveness of this method mainly lies in the accurate selection of model orders n_a and n_c as well as the temporal change assumed form of the parameters $a_{i,n}(t)$ and $c_{i,n}(t)$, and residual variance $\sigma^2_{e_i}(t)$. If these conditions are met, the parametric TFR can improve in accuracy, resolution and tracking of time-varying dynamics compared with linear and quadratic TFR.

Further discussion on the properties and advantages/disadvantages of each TFR estimation approach is out of the scope of this study. For more information on this subject, the reader can refer to the following papers: linear, quadratic and parametric TFR (Marchant, 2003); parametric *t–f* analysis (M Tarvainen & Karjalainen, 2004; Poulimenos & Fassois, 2006); comparison of different TFR on EEG classification (Tzallas et al., 2008); parametric TFR based classification (Avendano-Valencia et al., 2010).

Time-frequency Dynamic Features – TFDF are a set of variables describing the temporal dependency of some spectral–related quantity. The TFDF set of the signal $x_i(t)$ can be mathematically represented as a set of $\dots N$ functions $\boldsymbol{x}_i(t) = [x_{i,1}(t), \dots, x_{i,N}(t)]^\top$ describing the dynamics on a particular frequency band. When sampled, these features can be arranged as a matrix $\boldsymbol{X}_i \in \mathbb{R}^{N \times T}$. The following dynamic variables are under consideration: *linear frequency cepstral coefficients – LFCC*, *subband spectral centroids – SSC*, *subband spectral centroid energy – SSCE*, and *discrete wavelet transform – DWT*. The TFDF set describes the quantity of energy that the analyzed signal has on each of the n–frequency bands defined by the specific approach. In the case of LFCC, the TFDF are extracted using the following expression:

$$LFCC_{i,n}(t) = \sum_{m=1}^{n_M} \cos\left(n \frac{\pi(m-1)}{2N} \right) \log s_m(t) \tag{5}$$

$$s_m(t) = \int_{-\infty}^{\infty} F_m(\omega) X_i(t,\omega) d\omega$$

where N is the number of desired LFCCs to be considered, $F_m(\omega), m = 1, \dots n_M$ is a set of n_M triangular log–filter banks, and $s_m(t), m = 1, \dots, n_M$ is the weighted sum of each frequency filter response set. On the other hand, SSCs are computed following the equation:

$$SSC_{i,n}(t) = \frac{\int_{-\infty}^{\infty} \omega F_n(\omega) X_i(t,\omega) d\omega}{\int_{-\infty}^{\infty} F_n(\omega) X_i(t,\omega) d\omega} \tag{6}$$

where the set of filters $F_n(\omega)$, $n = 1, \dots, N$ are linearly distributed along the spectrum. Additionally, the energy for each SSC (SSCE) can also be considered as a time–variant feature,

which for a fixed bandwidth $\Delta\omega$ is computed by means of:

$$SCCE_{i,n}(t) = \int_{SSC_{i,n}(t)-\Delta\omega}^{SSC_{i,n}(t)+\Delta\omega} X_i(t,\omega)d\omega \tag{7}$$

where $SSC_{i,n}(t)$ is the value of the SSC estimated using (6). Both LFCCs from Equation (5) and SSCE from Equation (7) measure the average energy on the frequency band defined by each filter bank $F_m(\omega)$, whereas SSCs from Equation (6) measure the average frequency for each one of the defined spectral subbands. On the estimation of the TFDF any of the TFRs defined in equations (1), (2) and (4) can be used.

The *discrete wavelet transform* is also a means for analyzing the signal at different frequency bands, with different resolutions by decomposing the signal into approximation and detail coefficients. The DWT describes the signal $x_i(t)$ using the wavelet orthogonal basis $\{\psi_{j,k}\}_{(j,k)\in\mathbb{Z}}$ as:

$$x_i(t) = \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\langle x_i(t),\psi_{j,k}\rangle\psi_{j,k}(t) \tag{8}$$

$$\psi_{j,k} = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t-2^jk}{2^j}\right)$$

So, each partial sum

$$d_{i,j} = \sum_{k=-\infty}^{\infty}\langle x_i(t),\psi_{j,k}(t)\rangle\psi_{j,k}(t)$$

can be interpreted as the detail variations at the scale 2^j. As a result, the signal is defined in terms of the coefficients, representing their energy content in a specified time-frequency region, determined by the mother wavelet $\psi_{j,k}$ at scale j and time shift k (Mallat, 2008).

Both TFR and TFDF can be arranged into matrices. In the case of the estimated TFR, any of the methods from equations (1), (2) or (4) when evaluated on specific discrete time and frequency points will yield a TFR matrix $\boldsymbol{X}_i \in \mathbb{R}^{N\times T}$, as

$$\boldsymbol{X}_i = \begin{bmatrix} X_i(t_1,\omega_1) & X_i(t_2,\omega_1) & \dots & X_i(t_T,\omega_1) \\ X_i(t_1,\omega_2) & X_i(t_2,\omega_2) & \dots & X_i(t_T,\omega_2) \\ \vdots & \vdots & \ddots & \vdots \\ X_i(t_1,\omega_N) & X_i(t_2,\omega_N) & \dots & X_i(t_T,\omega_N) \end{bmatrix}$$

where N and T are the number of time and frequency sampling points, respectively. Similarly, for the estimated TFDF obtained from (5), (6), (7) or (8), evaluation on different time points yields a TFDF matrix $\boldsymbol{X}_i \in \mathbb{R}^{N\times T}$, as

$$\boldsymbol{X}_i = \begin{bmatrix}\boldsymbol{x}_i(t_1) & \boldsymbol{x}_i(t_2) & \dots & \boldsymbol{x}_i(t_T)\end{bmatrix}$$

2.2 State of the art approaches for TFR based classification

Within the context of feature extraction in TFR and dynamic features, two main approaches are to be considered:

– **Distance based approaches:** (Aviyente, 2004; Sejdic et al., 2009; Sejdic & Jiang, 2007) For this approach, a geometrical or statistical distance function is computed between the analyzed TFR and a template TFR. Furthermore, considering the resemblance between TFR and probability density functions, distance measures between probability density functions can be extended to the *t-f* plane (Michel et al., 64-67). The most common distance measure used for probability functions is the Kullback-Leibler divergence measure, that can be adapted to TFR as follows:

$$d_{KL}(X_1, X_2) = \int\int X_1(t,\omega) \log \frac{X_1(t,\omega)}{X_2(t,\omega)} dt d\omega \tag{9}$$

where $d_{KL}(X_1, X_2)$ is the Kullback-Leibler divergence. The measure $d_{KL}(X_1, X_2)$ is non-symmetrical and therefore, non-metric. By applying $d_{KLS}(X_1, X_2) = d_{KL}(X_1, X_2) + d_{KL}(X_2, X_1)$ the Kullback-Leibler divergence is converted into a symmetric measure.

Once a distance measure is selected, the classifier decision is based on the shortest distance to any of the template TFRs. Choosing an effective template often requires familiarity with the problem, otherwise, the template can be obtained as a group average or by cluster analysis. Another solution is to compare the analyzed TFR with a properly labeled TFR set and use the nearest neighbor rule. As in every distance-based nearest neighbor approach, this approach suffers when the data dimensionality is very high, and as a result, poor performance can be obtained. For this reason, the choice of a feature extractor can have a strong influence on the obtained results.

– **Regional analysis approaches** (Bernat et al., 2007; Cvetkovic et al., 2008; Hassanpour et al., 2004; Subasi, 2007; Tzallas et al., 2008) In this approach, the TFR is decomposed in regions of interest and further analysis is carried out with respect to the behavior on each of such regions defined on the *t–f* plane (Tzallas et al., 2008). For example, a partition of the *t-f* plane can be regarded, defining a grid depending on some *a priori* information known about the information distribution over the plane. As it is recommended in Tzallas et al. (2008) for the case of EEG classification, a partition of the time domain into 8 s equal sized windows is defined, while the frequency domain is divided into five sub-bands corresponding to the EEG frequency bands defined by medical experts (0-2.5 Hz, 2.5-5.5 Hz, 5.5-10.5 Hz, 10.5-21.5 Hz, and 21.5-43.5 Hz). Then, the feature vector $\boldsymbol{z}_i$ is built as,

$$\boldsymbol{z}_i = \left[\bar{X}_{11} \ \ldots \ \bar{X}_{1K} \ \bar{X}_{21} \ldots \ \bar{X}_{2K} \ \ldots \ \bar{X}_{L1} \ldots \ \bar{X}_{LK} \ \bar{X}\right] \tag{10}$$

$$\bar{X}_{lk} = \int_{t_l}\int_{\omega_k} X_i(t,\omega) d\omega dt, \qquad \bar{X} = \int\int X_i(t,\omega) d\omega dt$$

where $t_l, l = 1, \ldots, L$ is the l-th time window, $\omega_j, k = 1, \ldots, K$ is the k-th frequency band, and L, K are the number of time windows and frequency bands respectively. Each feature represents the fractional energy of the signal in a specific frequency band and time window. Besides the total energy of the signal is included as an additional feature. Then, the feature vector $\boldsymbol{z}_i$ depicts the energy distribution of the signal. This method is less computationally expensive compared to distance based approaches, but an adequate selection of the

regions of interest should be attained in order to achieve accurate results. Other regional decomposition can be accomplished by transform methods as proposed in (Bernat et al., 2007; Hassanpour et al., 2004) or wavelet sub–band decompositions (Cvetkovic et al., 2008; Subasi, 2007).

2.3 Dimensionality reduction on TFR and TFDF using orthogonal transforms

The TFR or TFDF–based pattern recognition approach can be described within the following framework. Let ξ_i be an object from object space $\boldsymbol{\xi}$. The object ξ_i is associated with a class label $c_i = 1, \ldots, N_c$ from the class label space $\mathcal{C} \subset \mathbb{N}$. A set of N features are estimated on different times T from the object ξ_i, yielding the TFR/TFDF matrices $\boldsymbol{X}_i$ lying on a subspace $\mathcal{X}$ of $\mathbb{R}^{N\times T}$. A dimensionality reduction mapping is used to convert such matrices into a set of lower dimensionality features $\boldsymbol{z}_i$ in a reduced feature subspace $\mathcal{Z}$ of $\mathbb{R}^p$. Then, features $\boldsymbol{z}_i$ are used as inputs for training and classification. It is assumed that both subspaces $\mathcal{X}$ and $\mathcal{Z}$ contain the discriminant information of space $\boldsymbol{\xi}$ and thus the relationship between $\boldsymbol{\xi}$ and $\mathcal{C}$ can be recovered. A set of M objects are sampled from object space $\boldsymbol{\xi}$, yielding the subset $\{\xi_i, i = 1, \ldots, M\}$. The estimated features from this subset are written as $\{\boldsymbol{X}_i, i = 1, \ldots, M\}$, and the features in reduced feature space are written as $\{\boldsymbol{z}_i, i = 1, \ldots, M\}$. Figure 1 depicts the space definitions within the relationships between them.

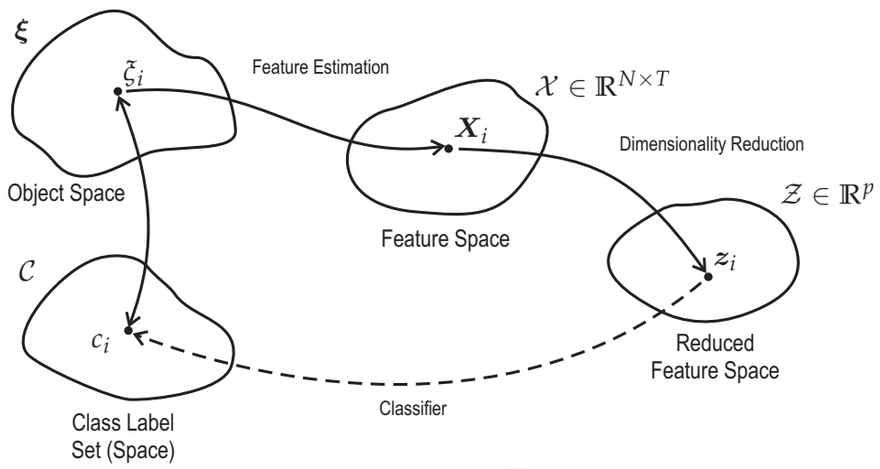


Fig. 1. Definition of object, class label, feature and reduced feature spaces, as well as mappings between spaces.

Within the explained framework, both TFR and TFDF are regarded as the *feature estimation* step, mapping from the object space (raw signals) to the feature space (space of feature matrices). Then, each one of the feature matrices, $\boldsymbol{X}_k \in \mathbb{R}^{N\times T}$ is described as follows:

$$\boldsymbol{X}_i = \left[\boldsymbol{x}_{c1}^{(i)}, \boldsymbol{x}_{c2}^{(i)}, \ldots, \boldsymbol{x}_{cT}^{(i)}\right] = \begin{bmatrix} \boldsymbol{x}_{r1}^{(i)} \\ \boldsymbol{x}_{r2}^{(i)} \\ \vdots \\ \boldsymbol{x}_{rN}^{(i)} \end{bmatrix} = \begin{bmatrix} x_{11}^{(i)} & x_{12}^{(i)} & \ldots & x_{1T}^{(i)} \\ x_{21}^{(i)} & x_{22}^{(i)} & \ldots & x_{2T}^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1}^{(i)} & x_{N2}^{(i)} & \ldots & x_{NT}^{(i)} \end{bmatrix},$$

where $\boldsymbol{x}_{ck}^{(i)}$ is the k–th column vector and $\boldsymbol{x}_{rl}^{(i)}$ is the l–th row vector of the feature matrix $\boldsymbol{X}_i$. Besides, $x_{lk}^{(i)}$ is the element in the l–th row and k–th column of the feature matrix $\boldsymbol{X}_i$.

Straightforward *dimensionality reduction* on matrix data by means of orthogonal transforms can be carried out, just by stacking matrix columns into a single vector, as follows:

$$\boldsymbol{x}_i = \left[(\boldsymbol{x}_{c1}^{(i)})^\top \ (\boldsymbol{x}_{c2}^{(i)})^\top \ \dots \ (\boldsymbol{x}_{cT}^{(i)})^\top \right] \tag{11}$$

where $\boldsymbol{x}_i \in \mathbb{R}^{NT}$ is the vectorized feature. Then, in order to reduce data dimensionality, a transform matrix $\boldsymbol{V} \in \mathbb{R}^{NT \times p}$, with $p \leq NT$ ($p \ll NT$) is defined to transform original feature space $\mathbb{R}^{NT}$ into a reduced feature space $\mathbb{R}^p$, by the linear operation $\boldsymbol{z}_i = \boldsymbol{x}_i \boldsymbol{V}$. The transformation matrix $\boldsymbol{V}$ can be obtained using a non–supervised approach, such as *principal component analysis – PCA*, or a supervised approach, such as *partial least squares – PLS* using data set matrix $\boldsymbol{X}$ and the class label matrix $\boldsymbol{c}$, defined as (Barker & Rayens, 2003),

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_M \end{bmatrix}; \qquad \boldsymbol{c} = \begin{bmatrix} \mathbf{1}_{M_1 \times 1} & \mathbf{0}_{M_2 \times 1} & \dots & \mathbf{0}_{M_{N_c} \times 1} \\ \mathbf{0}_{M_1 \times 1} & \mathbf{1}_{M_2 \times 1} & \dots & \mathbf{0}_{M_{N_c} \times 1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{M_1 \times 1} & \mathbf{0}_{M_2 \times 1} & \dots & \mathbf{1}_{M_{N_c} \times 1} \end{bmatrix}$$

where N_c is the number of class labels and $M_n, n = 1, \dots, N_c$ is the number of samples for each class in the training set. The vectorization approach in Equation (11) will be referred to as *vectorized PCA/PLS*, depending on the specific transform used. Vectorized PCA transform has been previously used in image recognition applications with the name of eigenfaces (Turk & Pentland, 1991). The vectorization approach was also used in (Bernat et al., 2007) to decompose and analyze TFR from P300 data, and in (Avendano-Valencia et al., 2010) for TFR–based classification. Nonetheless, this approach has several drawbacks, as the multidimensional structure of the data is not exploited nor analyzed; besides computational cost and memory requirements are increased.

When the feature matrix $\boldsymbol{X}_i$ is considered, a transform matrix $\boldsymbol{U} \in \mathbb{R}^{q \times N}$ can be used to reduce the number of rows in data matrix, as $\boldsymbol{z}_i^{(L)} = \boldsymbol{U}\boldsymbol{X}_i$, where $\boldsymbol{z}_i^{(L)} \in \mathbb{R}^{q \times T}$. Also, a transform matrix $\boldsymbol{V} \in \mathbb{R}^{T \times p}$ can be used to reduce the number of columns of data matrix, as $\boldsymbol{z}_i^{(R)} = \boldsymbol{X}_i \boldsymbol{V}$, where $\boldsymbol{z}_i^{(R)} \in \mathbb{R}^{N \times p}$. If both transforms are combined, further dimensionality reduction can be achieved, as

$$\boldsymbol{z}_i = \boldsymbol{U}\boldsymbol{X}_i\boldsymbol{V} \tag{12}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are the same as defined above, and $\boldsymbol{z}_i \in \mathbb{R}^{q \times p}$ is the reduced dimensionality feature. Estimation of the transformation matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ is carried out on the data matrices $\boldsymbol{X}^{(L)}$ and $\boldsymbol{X}^{(R)}$ respectively, which are defined as,

$$\boldsymbol{X}^{(L)} = \begin{bmatrix} \boldsymbol{X}_1^\top \\ \boldsymbol{X}_2^\top \\ \vdots \\ \boldsymbol{X}_M^\top \end{bmatrix}; \qquad \boldsymbol{X}^{(R)} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_M \end{bmatrix}$$

In image recognition literature this approach is known as 2D–PCA (Yang et al., 2004; Zhang & Zhou, 2005) when the transformation matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ in (12) are obtained

with PCA. 2D–PCA and 2D–PLS have also been used for TFR–based classification in (Avendano-Valencia et al., 2010).
Increased insight can be achieved if the functional structure of the data is taken into account. Let each column of the feature matrix $\boldsymbol{x}_{ck}^{(i)}$ be modeled as the weighted sum of basis functions ϕ_{jk},

$$\boldsymbol{x}_{ck}^{(i)} = \sum_{j=1}^{p} \boldsymbol{\alpha}_j^{(i)} \phi_{j,k} = \boldsymbol{\alpha}_1^{(i)} \phi_{1,k} + \boldsymbol{\alpha}_2^{(i)} \phi_{2,k} + \ldots + \boldsymbol{\alpha}_p^{(i)} \phi_{p,k} \tag{13}$$

where $\phi_{j,k}, j = 1, \ldots, p$ is a set of p basis functions evaluated in column k, and $\boldsymbol{\alpha}_j^{(i)} \in \mathbb{R}^N, j = 1, \ldots, p$ is the parameter vector of the model for the i-th object and the j-th base function. Then, the dynamic information is coded into the basis functions $\phi_{j,k}$ while the object information is coded in the set of parameter vectors $\boldsymbol{\alpha}_j^{(i)}$. Equation (13) can be rewritten as,

$$\boldsymbol{x}_{ck}^{(i)} = \begin{bmatrix} \alpha_{11}^{(i)} \\ \alpha_{21}^{(i)} \\ \vdots \\ \alpha_{N1}^{(i)} \end{bmatrix} \phi_{1,k} + \begin{bmatrix} \alpha_{12}^{(i)} \\ \alpha_{22}^{(i)} \\ \vdots \\ \alpha_{N2}^{(i)} \end{bmatrix} \phi_{2,k} + \ldots + \begin{bmatrix} \alpha_{1p}^{(i)} \\ \alpha_{2p}^{(i)} \\ \vdots \\ \alpha_{Np}^{(i)} \end{bmatrix} \phi_{p,k} = \begin{bmatrix} \alpha_{11}^{(i)} & \alpha_{12}^{(i)} & \ldots & \alpha_{1p}^{(i)} \\ \alpha_{21}^{(i)} & \alpha_{22}^{(i)} & \ldots & \alpha_{2p}^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{N1}^{(i)} & \alpha_{N2}^{(i)} & \ldots & \alpha_{Np}^{(i)} \end{bmatrix} \begin{bmatrix} \phi_{1,k} \\ \phi_{2,k} \\ \vdots \\ \phi_{p,k} \end{bmatrix}$$

$$\boldsymbol{x}_{ck}^{(i)} = \boldsymbol{A}_i \boldsymbol{\phi}_{ck} \tag{14}$$

then, the complete matrix can be written as,

$$\boldsymbol{X}_i = \left[\boldsymbol{x}_{c1}^{(i)} \; \boldsymbol{x}_{c2}^{(i)} \; \ldots \; \boldsymbol{x}_{cT}^{(i)} \right] = \boldsymbol{A}_i \left[\boldsymbol{\phi}_{c1} \; \boldsymbol{\phi}_{c2} \; \ldots \; \boldsymbol{\phi}_{cT} \right] = \boldsymbol{A}_i \boldsymbol{\Phi} \tag{15}$$

where the matrix $\boldsymbol{A}_i \in \mathbb{R}^{N \times p}$ accounts for the information in the object $\boldsymbol{\xi}_i$, while the matrix $\boldsymbol{\Phi} \in \mathbb{R}^{p \times T}$ accounts for the temporal information. The complete set of measurements from M objects $\boldsymbol{X}$ can be represented in the following form,

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_m \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1 \boldsymbol{\Phi} \\ \boldsymbol{A}_2 \boldsymbol{\Phi} \\ \vdots \\ \boldsymbol{A}_m \boldsymbol{\Phi} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \\ \vdots \\ \boldsymbol{A}_m \end{bmatrix} \boldsymbol{\Phi} = \boldsymbol{A} \boldsymbol{\Phi} \tag{16}$$

where $\boldsymbol{A} \in \mathbb{R}^{NM \times p}$ is a matrix containing the model parameters of the complete set of measurements from M objects. For the representation given in Equation (15), a procedure similar to 2D–PCA/PLS can be applied, using a transform matrix $\boldsymbol{U}$ to reduce the number of rows (measured variables) in data $\boldsymbol{X}_i$ by the linear operation,

$$\boldsymbol{z}_i^{(L)} = \boldsymbol{U} \boldsymbol{X}_i = \boldsymbol{U} \boldsymbol{A}_i \boldsymbol{\Phi} \tag{17}$$

where $\boldsymbol{U} \in \mathbb{R}^{q\times N}, q \leq N$ and $\boldsymbol{z}_i^{(L)} \in \mathbb{R}^{q\times T}$. Likewise, a transform matrix $\boldsymbol{V}$ can be used to reduce the number of columns (time points) in data $\boldsymbol{X}_i$ by the linear operation,

$$\boldsymbol{z}_i^{(R)} = \boldsymbol{X}_i\boldsymbol{V} = \boldsymbol{A}_i\boldsymbol{\Phi}\boldsymbol{V} = \boldsymbol{A}_i\boldsymbol{W} \tag{18}$$

with $\boldsymbol{V} \in \mathbb{R}^{T\times\tau}, \tau \leq T$, $\boldsymbol{W} = \boldsymbol{\Phi}\boldsymbol{V} \in \mathbb{R}^{p\times\tau}$ $\boldsymbol{z}_i^{(R)} \in \mathbb{R}^{N\times\tau}$. Combining both operations, temporal information and measurements can be compressed into a smaller representation, as follows

$$\boldsymbol{z}_i = \boldsymbol{U}\boldsymbol{X}_i\boldsymbol{V} = \boldsymbol{U}\boldsymbol{A}_i\boldsymbol{\Phi}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{A}_i\boldsymbol{W} \tag{19}$$

where $\boldsymbol{U}$, $\boldsymbol{V}$ and $\boldsymbol{W}$ are as explained above, and $\boldsymbol{z}_i \in \mathbb{R}^{q\times\tau}$. Notice also that if the matrix $\boldsymbol{\Phi}$ has orthogonal rows and also $\boldsymbol{V} = \boldsymbol{\Phi}^\top$, then $\boldsymbol{W} = \boldsymbol{\Phi}\boldsymbol{\Phi}^\top = \boldsymbol{I}_{p\times p}$, thus,

$$\boldsymbol{z}_i = \boldsymbol{U}\boldsymbol{A}_i\boldsymbol{\Phi}\boldsymbol{\Phi}^\top = \boldsymbol{U}\boldsymbol{A}_i \tag{20}$$

subsequently, the dimensionality reduction can only be carried out in the parameter matrix $\boldsymbol{A}_i$. Transform matrix $\boldsymbol{U}$ can be computed in the same way as in 2D–PCA/PLS approach, whereas $\boldsymbol{\Phi}$ can be any truncated orthogonal basis, for example, trigonometric basis, polynomial basis, etc. This approach will be referred to as *functional PCA/PLS* (*f*PCA/*f*PLS) in accordance with the specific transform method.

2.4 The concept of relevance in dynamic variables

Relevance analysis distinguishes variables that effectively represent the subjacent physiological phenomena according to an evaluation measure. Such representative variables are named *relevant features*, whereas the evaluation measure is known as *relevance measure*. Variable selection tries to reject those variables whose contribution to the representation target is none or negligible (*irrelevant features*), as well as those variables that have repeated information (*redundant features*). Therefore, the first issue concerning variable selection is selecting an appropriate relevance definition. Previous efforts have been made in this area by (Yu & Liu, 2004) and (Sepúlveda-Cano et al., 2011) for the case of static variables under non-supervised framework.

The notion of relevance can be cast into the supervised framework by considering the object set $\mathcal{X}_s = \{\boldsymbol{X}_i, i = 1,\ldots,M\} \subset \mathcal{X}$ including M observation samples from the feature subset $\mathcal{X}$. Each observation sample is associated with a class label $c_i \in \mathbb{N}$ constituting the sampled class label (sub)set $\mathcal{C}_s$ from class label set $\mathcal{C}$. Then, given $\mathcal{X}_s$ and $\mathcal{C}_s$, for each one of the $\boldsymbol{x}_{lk}$ features, the relevance function ρ is defined as follows:

$$\begin{aligned}\rho : \mathbb{R}^{N\times T} \times \mathbb{N} \times \mathbb{R} &\to \mathbb{R}\\ (\mathcal{X}_s, \mathcal{C}_s, \boldsymbol{x}_{lk}) &\mapsto \rho(\mathcal{X}_s, \mathcal{C}_s, \boldsymbol{x}_{lk}) \in \mathbb{R}\end{aligned} \tag{21}$$

where the feature relevance function ρ satisfies the following properties (Sepúlveda-Cano et al., 2011):

- *Non–negativity*, i.e. $\rho(\mathcal{X}_s, \mathcal{C}_s, \boldsymbol{x}_{lk}) \geq 0$, $\forall l, k$.
- *Nullity*, the function $\rho(\mathcal{X}_s, \mathcal{C}_s, \boldsymbol{x}_{lk})$ is null if and only if the feature $\boldsymbol{x}_{lk}$ has not relevance at all.

– *Non–redundancy,* if a feature $\boldsymbol{x}' = \alpha\boldsymbol{x} + \eta$, where the real–valued $\alpha \neq 0$ and η is some noise with mean zero and ε variance, then, the difference $|\rho(\mathcal{X}_s, \mathcal{C}_s, \boldsymbol{x}') - \rho(\mathcal{X}_s, \mathcal{C}_s, \boldsymbol{x})| \rightarrow 0$ as $\varepsilon \rightarrow 0$.

The value of the function $\rho_{lk} = \rho(\mathcal{X}_s, \mathcal{C}_s, \boldsymbol{x}_{lk})$ for the feature $\boldsymbol{x}_{lk}$ is called *relevance weight*. When all the features are considered, a relevance matrix can be built as $\boldsymbol{R} = [\rho_{lk}]$ for $l = 1, \ldots, N$ and $k = 1, \ldots, T$. Also, when relevance is concerned over any of the axes (rows or columns of the feature matrix), a relevance estimate can be obtained by averaging on rows or columns, thus obtaining

$$\boldsymbol{\rho}_r = \frac{1}{T}\sum_{k=1}^{T}\boldsymbol{\rho}_{ck} \qquad\qquad \boldsymbol{\rho}_c = \frac{1}{N}\sum_{l=1}^{N}\boldsymbol{\rho}_{rl} \tag{22}$$

where, within the specific TFR and TFDF framework, $\boldsymbol{\rho}_r$ accounts for the relevance of each time-varying feature, while $\boldsymbol{\rho}_c$ describes how relevance change through time. Then, the variable selection process is carried out by selecting those x_{lk} features whose relevance value ρ_{lk} is over a certain threshold δ. It is likely that most relevant time-varying features can be selected, extracting those time-varying features $\boldsymbol{x}_{rl}$ whose relevance value $\boldsymbol{\rho}_{rl}$ is over certain threshold ρ_{min}.

Within the literature several relevance measures have been considered, such as linear correlation, conditional entropy, symmetrical uncertainty and transformation–based measures (Avendano-Valencia et al., 2010; Yu & Liu, 2004). Among them, *symmetrical uncertainty* is capable of measuring both linear and complex non-linear relationships between variables. The symmetrical uncertainty is defined as follows

$$\rho_{su}(x_i, c) = 2\frac{H(x_i) - H(x_i|c)}{H(x_i) + H(c)}, \quad \rho_{su}(x_i, c) \in [0, 1], \forall i = 1, \ldots, m \tag{23}$$

where $H(x_i)$ is the entropy of the features and $H(x_i|c)$ is the conditional entropy of the features given the classes, both defined as

$$H(x_i) = -\int_{x_i} P(x_i)\log P(x_i)dx_i, \quad \forall i = 1, \ldots, m \tag{24}$$

$$H(x_i|c) = -\int_{c} P(c)\int_{x_i} P(x_i|c)\log P(x_i|c)dx_i dc, \quad \forall i = 1, \ldots, m \tag{25}$$

A value of $\rho_{su}(x_i, c) = 1$ indicates that the feature x_i completely predicts the values of the class labels c. Since the computation of Equation (23) requires the estimation of both $P(x_i)$ and $P(x_i|c)$, histogram–driven estimates might be used. Therefore, the integrals in Equations (24) and (25) become sums that are carried out along the histogram bins.

2.5 Selection of most informative areas from dynamic *t–f* variables

Once the relevance measure is properly determined, the selection of the feature vectors is carried out by choosing those variables with a relevance that exceeds a given threshold. Nonetheless, managing dynamic variables requires special handling since the considered features are no longer organized as vectors. Within this framework two approaches are proposed: a) *1D-Relevance:* Evaluate the relevance measure for each point of the TFR, and then select the most relevant time–frequency points to appraise a reduced feature vector that

will be later processed by conventional linear transform methods (PCA or PLS); this approach is described in Algorithm 1. b) *2D-Relevance:* Evaluate the relevance measure on rows or columns of the dynamic feature set as in Equation (22) and then select the most relevant time instants or frequency bands to appraise a TFR–based feature matrix, which will be further reduced using either the matricial approach: 2D–PCA, 2D–PLS; this approach is described in Algorithm 2. c) *Functional Relevance:* The relevance measure is evaluated as in b) to select the most relevant time instants of frequency bands and later, the most relevant data is further reduced using Ě*f*PCA or *f*PLS. The approach is described in Algorithm 3. The approach is described in Algorithm 3..

Algorithm 1 1D-Relevance: Selection of TFR–based features using relevance measures and dimensionality reduction

Input: TFR dataset $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_M\}$, relevance threshold $\rho_{\min}$.
Output: Reduced feature vector set $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_M\}$.

1. Convert TFR matrices into vectors
   for $i = 1$ to M **do**
   $\boldsymbol{x}_i = \mathrm{vec}(\boldsymbol{X}_k) = \left[(\boldsymbol{x}_{c1}^{(i)})^\top, (\boldsymbol{x}_{c2}^{(i)})^\top, \ldots, (\boldsymbol{x}_{cT}^{(i)})^\top\right]$
   end for
2. Compute the relevance measure vector $\rho(\boldsymbol{x})$ of the feature vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}$, using the relevance measure defined in Equation (23).
3. Select the most relevant variables from vectorized TFRs
   for $i = 1$ to M **do**
   $\hat{\boldsymbol{z}}_i = \left\{ x_{lk}^{(i)} \quad \forall k, l : \rho(x_{kl}) \geq \rho_{\min} \right\}$
   end for
4. Compute the transformation matrix $\boldsymbol{V}$ of PCA or PLS using the relevant feature vector set $\{\hat{\boldsymbol{z}}_1, \hat{\boldsymbol{z}}_2, \ldots, \hat{\boldsymbol{z}}_M\}$.
5. Transform the feature vectors $\hat{\boldsymbol{z}}_i$ into the reduced feature vector $\boldsymbol{z}_i$, as
   for $i = 1$ to M **do**
   $\boldsymbol{z}_i = \hat{\boldsymbol{z}}_i \boldsymbol{V}$
   end for

3. Experimental set–up

3.1 EEG database

The EEG signals correspond to 29 patients with medically intractable focal epilepsies. They were recorded by the Department of Epileptology of the University of Bonn as explained in (Andrzejak et al., 2001). The database comprises five sets (denoted as A-E) composed of 100 single channel EEG segments, which were selected and extracted after visual inspection from continuous multichannel EEG to avoid artifacts (e.g. muscular activity or eye movements). Datasets A and B consist of segments taken from scalp EEG records of five healthy subjects using the standard $10-20$ electrode placements. Volunteers were woken up, and relaxed with their eyes open (A) and eyes closed (B), respectively. Datasets C, D and E were selected

Algorithm 2 2D-Relevance: Frequency band selection from TFR using relevance measures and dimensionality reduction by matricial approach

Input: TFR matrix dataset $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_M\}$, relevance threshold $\rho_{\min}$.
Output: Reduced feature vector set $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_M\}$.

1. Follow steps 1. and 2. of Algorithm 1
2. Compute the average relevance value of row axis $\boldsymbol{\rho_r}$ as defined by Equation (22).
3. Select the most relevant frequency bands from TFR
   for $i = 1$ to M **do**
   $\hat{\boldsymbol{X}}_i = \left\{ \boldsymbol{x}_{rl}^{(i)} \quad \forall l : \rho_{rl} \geq \rho_{\min} \right\}$
   end for
4. Compute the transformation matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ of 2D–PCA (or 2D–PLS), using the reduced TFR matrices set $\{\hat{\boldsymbol{X}}_1, \hat{\boldsymbol{X}}_2, \ldots, \hat{\boldsymbol{X}}_M\}$.
5. Transform the reduced TFR matrices $\hat{\boldsymbol{X}}_i$ into the reduced feature vector $\boldsymbol{z}_i$, as
   for $i = 1$ to M **do**
   $\boldsymbol{Z}_i = \boldsymbol{U}\hat{\boldsymbol{X}}_i\boldsymbol{V}^\top$
   $\boldsymbol{z}_i = vec(\boldsymbol{Z}_i)$
   end for

from presurgical diagnosed EEG recordings. The signals were selected from five patients who achieved a complete control of the epileptic episodes after the dissection of one of the hippocampal formations, which was correctly diagnosed as the epileptogenic zone. Segments of set D were recorded in the epileptogenic zone, and segments of C in the hippocampal zone of the opposite side of the brain. While sets C and D only contain activity measured on inter–ictal intervals, set E only contains records with ictal activity. In this set, all segments were selected from every recording place exhibiting ictal activity. All EEG signals were recorded with an acquisition system of 128 channels, using average common reference. Data was digitized at 173.61 Hz with 12 bits of resolution.
In this study, the database described above is organized to create one classification task, where all the EEG segments are sorted into three different classes: A and B types of EEG segments were combined in a single class (Normal); C and D types were also combined in a single class (Interictal); and type E completes the third class (Ictal). This set is the one closest to real medical applications which include three categories: normal (i.e., types A and B) with 200 recordings, seizure free (i.e., types C and D) with 200 recordings, and seizure (i.e., type E) with 100 recordings. Validation of the classifier is carried out by 10 fold cross-validation, where the database is divided into 10 folds containing different records from each class. Nine of these folds are used for training and the remaining one for validation purposes. Training and validation folds are changed until the ten folds are used for validation.

3.2 Feature estimation

Time-Frequency Representations: Time-frequency analysis of EEG signals is carried out using STFT, SPWVD and parametric TFR based on SP-TARMA modeling. Analysis is performed in the frequency range from 0 to 43.4 Hz as recommended for EEG signals (Subasi, 2007). STFT is computed using a 512 point gaussian window with a 504 sample overlap. SPWV parameters

Algorithm 3 Functional Relevance: Frequency band selection from TFR using relevance measures and dimensionality reduction by functional approach

Input: TFR matrix dataset $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_M\}$, relevance threshold $\rho_{\min}$.
Output: Reduced feature vector set $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_M\}$.

1. Follow steps 1. and 2. of Algorithm 1
2. Compute the average relevance value of row axis $\boldsymbol{\rho_r}$ as defined by Equation (22).
3. Select the most relevant frequency bands from TFR
   for $i = 1$ to M **do**
   $\hat{\boldsymbol{X}}_i = \left\{ \boldsymbol{x}_{rl}^{(i)} \quad \forall l : \rho_{rl} \geq \rho_{\min} \right\}$
   end for
4. Compute the representation coefficients $\boldsymbol{A_i}$ of the basis $\boldsymbol{\Phi}$
   for $i = 1$ to M **do**
   $\hat{\boldsymbol{A}}_i = \hat{\boldsymbol{X}}_i \boldsymbol{\Phi}^{-1}$
   end for
5. Compute the transformation matrices $\boldsymbol{U}$ and of *f*PCA (or *f*PLS), using the representation coefficient matrix set $\{\hat{\boldsymbol{A}}_1, \hat{\boldsymbol{A}}_2, \ldots, \hat{\boldsymbol{A}}_M\}$.
6. Transform the reduced TFR matrices $\hat{\boldsymbol{A}}_i$ into the reduced feature vector $\boldsymbol{z}_i$, as
   for $i = 1$ to M **do**
   $\boldsymbol{Z}_i = \boldsymbol{U}\hat{\boldsymbol{A}}_i$
   $\boldsymbol{z}_i = vec(\boldsymbol{Z}_i)$
   end for

are adjusted as suggested in (Tzallas et al., 2008), using 64-point Hamming window for time and frequency smoothing window. For parametric SP-TAR, estimation, model order selection and validation is implemented following the recommendations in (Poulimenos & Fassois, 2006). The model order selection approach consists in finding the elbow of the *log likelihood function* computed for each record in the database using SP-TAR(n_a) models, with $n_a = 2, \ldots, 20$. The achieved selected orders n_a are between 10 to 14 for the SP–TAR model. The estimation parameters of each estimation methodology are summarized in Table 1. For all the methods considered, a set of TFRs with size $T = 512$ and $N = 256$ was obtained.

Time-Frequency Dynamic Features: Analysis by TFDF is carried out with SSC, SSCE, LFCC and DWT coefficients. Spectral centroids and cepstral coefficients are computed from STFT, using 8 frequency sub-bands. Filters banks are designed using gaussian functions with a 10% overlap between bands. DWT is computed using Daubechies 4 mother wavelet up to the fifth decomposition level as suggested in (Subasi, 2007). For each EEG recording, both wavelet coefficients and reconstructed time series for each decomposition level are considered in the analysis. The information regarding to estimation of each dynamic feature is summarized in Table 1.

Figure 2 shows exemplary TFRs obtained with the three different estimation approaches. Each column refers to a different estimation approach: STFT - short time Fourier transform, SPWVD - smoothed pseudo Wigner-Ville distribution, and SP-TAR - parametric TFR based

Estimation Method	Parameters
STFT	Window parameters: Gaussian window, length 512 samples, overlap 504 samples.
SPWV	Time and frequency smoothing window: Hamming 64 point window. Number of points in frequency axis 256.
SP-TARMA	Model order selection criteria: elbow of *log likelihood function*. Variance tradeoff parameter of Kalman filter/smoother $v = 0.01$. Smoothness priors order $\kappa = 2$. Number of points in frequency axis 256.
TFDF	**Parameters**
SSC & SSCE	Filter bank function: Gaussian function. Band overlap: 10%. Number of sub-bands: 10. Energy estimation bandwidth: $\Delta\omega$ 10.85 Hz.
LFCC	Number of sub-bands: 10. Number of LFCC: 8.
DWT	Mother wavelet: Daubechies 4. Decomposition level 5.

Table 1. Implementation details of TFR and TFDF estimation methods.

on SP-TAR model. Likewise, each row shows TFRs associated with some typical recordings from a database class (A to E).

3.3 Dimensionality reduction

The dimensionality reduction scheme considers a linear transform approach to the set of most relevant features extracted from the *t–f* feature set. Most relevant features are selected according to the symmetrical uncertainty relevance measure, described in Equation (23). The symmetrical uncertainty is evaluated on the TFR dataset yielding the relevance matrices and vectors shown in Figure 3. Plots in the left column show the average relevance measure over the frequency axis; plots in the central column show the TFR relevance average for each training fold set; plots in the right column show the sorted relevance measures vs. the percentage of features. In the case of the TFDF set, the relevance measure is evaluated and the relevance of each TFDF is obtained by averaging all of the relevance outcomes along time. The relevance measure obtained for each one of the analyzed TFDF sets is shown in Figure 4. After evaluating the relevance measure on each dataset, the most relevant features are further reduced using vectorized, 2D and functional linear transform approaches. The representation coefficients in functional linear transform method are computed by means of discrete cosine transform. The basis is truncated up to the first coefficient with 0.5% of the power of the largest coefficient. The number of features in the reduced feature space is selected as the minimum number that explains 95% of the complete dataset variance.

3.4 Results

The classification performance is measured the accuracy, sensitivity and specificity, defined by,

$$\text{Acc (\%)} = \frac{N_C}{N_T} \times 100; \quad \text{Sens (\%)} = \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100; \quad \text{Spec (\%)} = \frac{N_{TN}}{N_{TN} + N_{FP}} \times 100$$

where N_C is the number of correctly classified patterns, N_T is the total number of patterns used to feed the classifier, N_{TP} is the number of true positives (objective class accurately classified), N_{FN} is the number of false negatives (objective class classified as reference class), N_{TN} is the

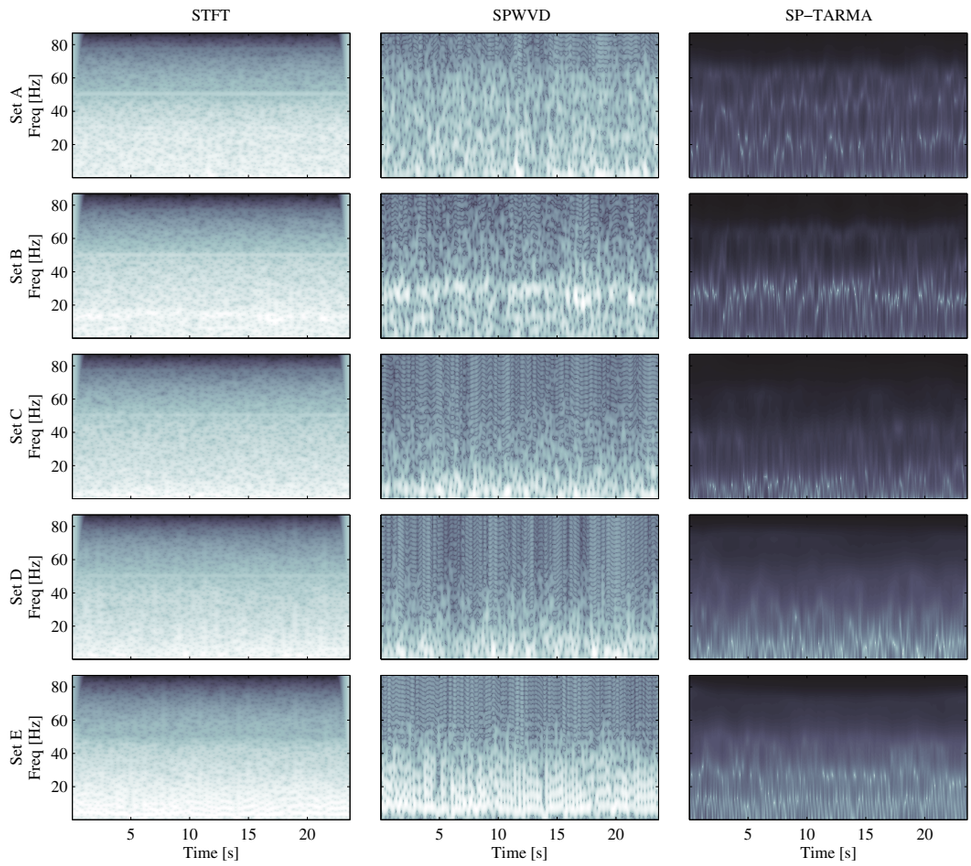


Fig. 2. Estimated TFR using the three different estimation approaches considered for exemplary signals from each class (A to E). STFT - short time Fourier transform, SPWVD - smoothed pseudo Wigner-Ville distribution, and SP-TARMA - parametric TFR based on SP-TARMA model.

number of true negatives (reference class accurately classified), and N_{FP} is the number of false positives (reference class classified as objective class). Mean and standard deviations of the accuracy, sensitivity and specificity are calculated for all the validation folds.
The following approaches are analyzed:

– Conventional approaches: *t–f* grid energy approach (`TF-grid`), TFR Kullback-Leibler distance based nearest neighbors approach (`TFR-KNN`) and statistical measures extracted from DWT coefficients (`DWT-averages`).

– Relevance analysis and linear transform based approaches: symmetrical uncertainty based vectorized PCA/PLS on TFR matrices and on TFDF (`1D- PCA TFR`, `1D- PLS TFR`, `1D- PCA TFDF`, and `1D-PLS TFDF`); symmetrical uncertainty based 2D–PCA/PLS on TFR matrices and on TFDF (`2D-PCA TFR`, `2D-PLS TFR`, `2D-PCA TFDF`, and `2D- PLS TFDF`); and symmetrical uncertainty based functional PCA/PLS on TFR matrices and on TFDF (*f*`PCA TFR`, *f*`PLS TFR`, *f*`PCA TFDF`, and *f*`PLS TFDF`).

Table 2 shows the performance outcomes of the conventional approaches. The feature vectors obtained with `TF-grid` and `DWT-averages` are further reduced by PCA, using a

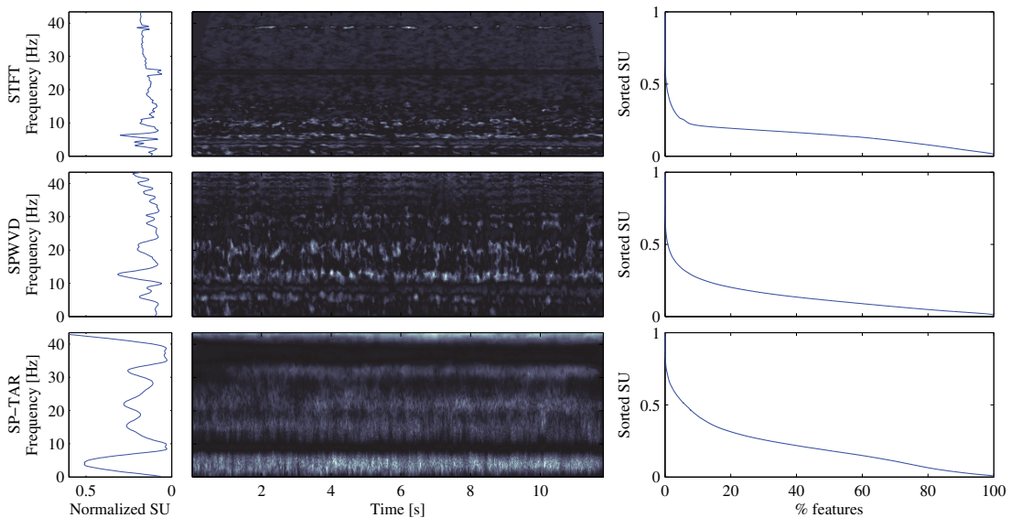


Fig. 3. Average symmetrical uncertainty values for TFR sets of EEG database. Left plots, relevance average on the frequency axis; center plots: fold average of the TFR relevance matrix; right plots: sorted relevance measure. Top–bottom: TFR estimator, STFT, SPWVD and parametric SP–TAR.

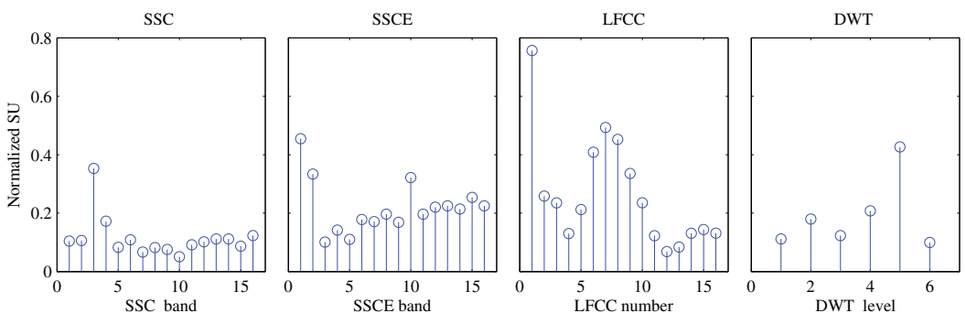
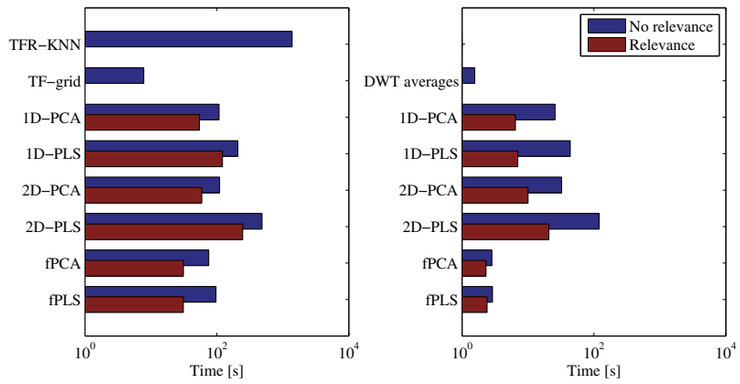


Fig. 4. Average symmetrical uncertainty values for TFDF sets of EEG database. Each plot shows the relevance value for each TFDF set, SSC, SSCE, LFCC and DWT level coefficients.

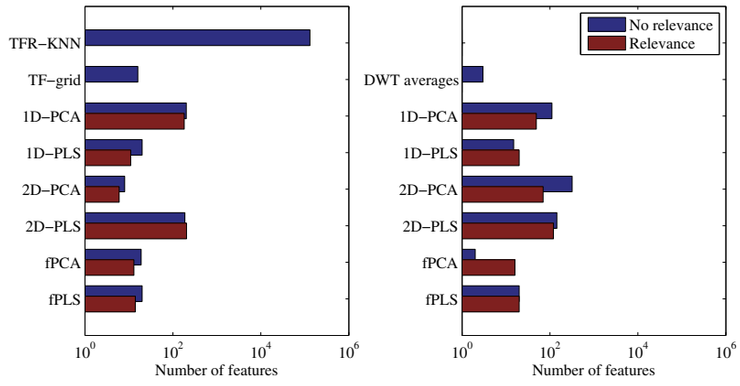
number of features accounting for 95% of the variability of the original data. All TFR–based approaches are tested with STFT, SPWVD and SP-TAR based TFR estimates. Table 3 shows the performance outcomes of the 1D-Relevance, 2D–Relevance and Functional Relevance methods described by Algorithms 1, 2, and 3, respectively. A comparison is made between linear transforms on the complete data and on the 50% of the most relevant variables. All TFR–based approaches are tested with STFT, SPWVD and SP-TAR based TFR estimates, whereas all TFDF–based approaches are tested with SSC, SSCE, LFCC and DWT subband coefficients. For all approaches, a 3 nearest neighbor classifier is used. In addition to classifier performance, the computing times and size of the reduced feature set are taken into account in the analysis. The average computing times during training for each method are shown in the Figure 5(a). Size of the reduced feature set for each method is shown in the Figure 5(b). Values for TFR–based approaches are shown on the left plot, and those for TFDF–based approaches in the plot on the right.

	Method	Accuracy	Sensitivity	Specificity
Conventional	STFT	91.20 ± 3.16	92.00 ± 2.58	94.33 ± 4.98
	`TF-grid` SPWVD	92.20 ± 3.94	94.50 ± 4.97	96.33 ± 3.67
	SP–TAR	91.40 ± 1.65	96.50 ± 2.42	93.67 ± 2.46
	STFT	88.60 ± 4.12	97.00 ± 3.50	97.33 ± 3.06
	`TFR-KNN` SPWVD	88.00 ± 4.90	96.00 ± 4.59	97.67 ± 2.74
	SP–TAR	85.20 ± 3.43	98.50 ± 2.42	92.00 ± 5.26
	`DWT-averages`	88.40 ± 3.86	93.00 ± 5.37	92.67 ± 4.39

Table 2. Performance outcomes of conventional *t–f* based classification approaches for epilepsy diagnosis on EEG database.



(a) Computing times.



(b) Size of the reduced feature set.

Fig. 5. Computing times and size of the reduced feature set of each one of the analyzed *t–f* based classification approaches. The left plot show the values for TFR–based approaches, the right plot show the values for TFDF–based approaches.

	Method		Complete feature set			50% most relevant variables		
			Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
1D–Relevance	1D-PCA TFR	STFT	92.20 ± 3.58	94.00 ± 5.16	98.00 ± 2.33	92.20 ± 4.26	93.50 ± 5.80	97.33 ± 3.44
		SPWVD	90.60 ± 4.99	91.50 ± 4.74	99.33 ± 1.41	92.60 ± 4.22	90.00 ± 4.08	98.33 ± 2.36
		SP–TAR	85.80 ± 3.94	93.00 ± 5.87	91.67 ± 4.23	91.00 ± 3.02	97.00 ± 3.50	93.67 ± 4.83
	1D-PLS TFR	STFT	96.60 ± 3.13	98.50 ± 2.42	97.67 ± 2.25	96.00 ± 3.27	97.00 ± 3.50	97.33 ± 2.63
		SPWVD	97.40 ± 2.50	99.50 ± 1.58	98.00 ± 2.33	95.40 ± 3.13	97.50 ± 2.64	98.00 ± 3.22
		SP–TAR	96.60 ± 2.99	100.00 ± 0.00	96.67 ± 3.51	96.80 ± 2.70	99.00 ± 2.11	97.33 ± 3.06
	1D-PCA TFDF	SSC	76.00 ± 7.72	74.50 ± 11.65	92.00 ± 5.92	78.40 ± 6.10	74.50 ± 8.96	91.33 ± 6.89
		SSCE	63.20 ± 10.38	63.50 ± 13.95	79.00 ± 12.18	60.40 ± 6.59	64.00 ± 14.49	73.67 ± 7.45
		LFCC	62.00 ± 7.42	57.00 ± 9.78	72.67 ± 6.44	64.00 ± 6.99	58.00 ± 8.56	76.67 ± 5.88
		DWT	85.20 ± 5.90	82.00 ± 11.35	90.00 ± 4.44	78.20 ± 5.20	71.50 ± 11.32	90.67 ± 3.06
	1D-PLS TFDF	SSC	86.20 ± 4.37	85.00 ± 7.45	96.00 ± 4.66	87.20 ± 3.91	85.50 ± 6.85	94.33 ± 3.53
		SSCE	93.40 ± 2.50	93.00 ± 4.83	98.33 ± 2.36	89.80 ± 3.58	88.00 ± 6.32	97.33 ± 2.63
		LFCC	91.00 ± 3.80	94.00 ± 7.38	93.33 ± 3.51	89.00 ± 4.55	91.50 ± 5.80	92.33 ± 4.17
		DWT	93.40 ± 3.53	97.00 ± 3.50	94.00 ± 3.44	91.20 ± 3.16	94.00 ± 3.16	93.67 ± 5.54
2D–Relevance	2D-PCA TFR	STFT	86.20 ± 3.94	69.00 ± 10.75	99.33 ± 2.11	85.40 ± 3.78	67.50 ± 10.87	99.33 ± 2.11
		SPWVD	90.60 ± 2.67	79.00 ± 9.66	99.33 ± 2.11	89.80 ± 3.82	76.50 ± 11.32	99.67 ± 1.05
		SP–TAR	90.00 ± 3.53	95.00 ± 2.36	94.33 ± 3.87	96.40 ± 2.80	98.50 ± 2.42	97.00 ± 3.31
	2D-PLS TFR	STFT	92.80 ± 3.16	88.50 ± 7.09	99.00 ± 2.25	92.00 ± 3.27	88.00 ± 7.15	98.67 ± 2.33
		SPWVD	94.20 ± 2.39	94.00 ± 5.16	98.00 ± 2.33	95.20 ± 2.53	95.00 ± 6.24	98.00 ± 2.81
		SP–TAR	92.80 ± 4.44	98.50 ± 2.42	94.33 ± 5.68	96.60 ± 2.50	98.50 ± 3.37	97.67 ± 2.74
	2D-PCA TFDF	SSC	78.00 ± 8.16	60.00 ± 15.09	99.33 ± 1.41	78.00 ± 8.27	61.50 ± 14.35	99.00 ± 1.61
		SSCE	93.00 ± 5.10	95.00 ± 6.24	96.67 ± 3.14	82.60 ± 4.62	87.00 ± 7.89	89.33 ± 5.40
		LFCC	79.40 ± 9.24	81.50 ± 16.84	87.00 ± 3.99	80.00 ± 10.02	83.00 ± 16.36	87.33 ± 4.66
		DWT	92.20 ± 3.71	96.50 ± 4.12	90.33 ± 4.57	94.60 ± 2.32	99.50 ± 1.58	92.67 ± 4.10
	2D-PLS TFDF	SSC	85.60 ± 3.63	81.50 ± 7.47	94.00 ± 3.44	87.00 ± 3.56	82.50 ± 7.91	94.33 ± 3.16
		SSCE	82.60 ± 6.11	75.00 ± 8.16	95.00 ± 2.36	83.60 ± 5.23	79.50 ± 7.98	92.67 ± 5.16
		LFCC	88.80 ± 4.02	91.50 ± 7.84	93.33 ± 3.51	89.80 ± 4.16	92.00 ± 8.23	94.33 ± 3.16
		DWT	92.80 ± 2.70	97.50 ± 3.54	92.33 ± 4.17	92.80 ± 4.13	97.50 ± 3.54	92.67 ± 6.05
Functional Relevance	*f*PCA TFR	STFT	95.20 ± 2.86	96.00 ± 3.16	98.00 ± 3.22	90.00 ± 5.16	92.50 ± 4.86	93.00 ± 4.83
		SPWVD	92.40 ± 2.95	95.00 ± 4.08	94.67 ± 3.58	91.00 ± 3.68	92.50 ± 4.86	94.67 ± 3.58
		SP–TAR	96.60 ± 3.41	99.50 ± 1.58	98.00 ± 2.81	92.80 ± 6.48	98.50 ± 3.37	94.67 ± 4.77
	*f*PLS TFR	STFT	94.80 ± 3.29	96.00 ± 3.16	97.67 ± 3.16	93.00 ± 2.87	95.50 ± 4.38	95.00 ± 4.23
		SPWVD	94.80 ± 2.35	97.00 ± 3.50	97.00 ± 2.92	91.80 ± 2.90	94.00 ± 6.15	95.33 ± 3.58
		SP–TAR	97.00 ± 2.87	99.50 ± 1.58	97.67 ± 2.74	97.40 ± 2.50	99.50 ± 1.58	98.00 ± 2.33
	*f*PCA TFDF	SSC	84.60 ± 4.53	86.50 ± 5.80	90.67 ± 4.66	85.80 ± 5.29	80.50 ± 8.32	92.00 ± 4.22
		SSCE	93.80 ± 2.20	94.00 ± 3.16	98.67 ± 1.72	85.20 ± 3.91	83.50 ± 5.30	93.00 ± 2.92
		LFCC	87.40 ± 4.72	89.00 ± 7.38	91.33 ± 3.22	79.20 ± 6.75	79.00 ± 15.06	87.33 ± 4.39
		DWT	86.80 ± 4.54	90.50 ± 5.50	89.67 ± 7.61	88.00 ± 3.89	94.50 ± 3.69	90.67 ± 6.99
	*f*PLS TFDF	SSC	78.40 ± 5.48	72.00 ± 10.06	90.67 ± 4.66	84.80 ± 5.18	82.00 ± 7.53	94.00 ± 4.66
		SSCE	72.60 ± 5.58	45.00 ± 13.94	95.67 ± 3.16	90.20 ± 6.00	92.50 ± 6.35	95.33 ± 4.50
		LFCC	55.60 ± 3.24	39.50 ± 8.64	73.00 ± 5.32	88.20 ± 4.05	90.50 ± 6.85	93.00 ± 2.46
		DWT	73.20 ± 4.34	55.50 ± 9.26	89.00 ± 4.73	89.20 ± 3.29	94.50 ± 3.69	92.00 ± 5.02

Table 3. Performance outcomes of *t–f* classification approaches based on linear transforms for epilepsy diagnosis on EEG database.

4. Discussion

4.1 Time–Frequency Representations and Time–Frequency Dynamic Features

Three different TFR estimators have been studied in the experimental framework, STFT, SPWVD and parametric TFR based on SP-TAR modeling. Accurate estimation of the TFR is very important to achieve adequate performance of the classifier. Nonetheless, after proper adjustment of TFR estimators, it is possible to obtain good results. The results in tables 2 and 3 show that the performance outcomes attained by STFT, SPWVD and parametric TFR are very similar. According to the dimensionality reduction approach, each TFR estimator shows

either better or worse results. The best results are found with parametric TFR and SPWVD, but, in general, there is no tendency of the best performance. Both STFT and SPWVD are easy to adjust and compute, whereas parametric TFR requires an increased set up time and user expertise to achieve proper performance. So, in terms of the problem addressed in this study, it may be enough to use simple TFR estimators. Nevertheless, more specialized approaches may be needed for other applications under different non–stationary conditions.
On the other hand, four TFDF sets have been considered in this study, SSC, SSCE, LFCC and DWT coefficients. Those features are shown to be easier to manage than TFR features and require less computing time, as they are more compact representations (see Figure 5). The performance of the TFDF based approach is lower than the performance of its TFR counterpart. The performance obtained using different dimensionality reduction approaches is random and the dispersion values are high. Only the SSCE and DWT coefficients show proper performance. The best performance is achieved using DWT coefficients.

4.2 Classification based on *t–f* features

Several approaches for TFR and TFDF have been analyzed. Such approaches can be grouped into conventional approaches, vectorization transform methods, bidimensional (2D) transform methods and functional transform methods. Distance based and local averaging techniques are grouped within the conventional approaches. Nearest neighbor classifier based on computation of Kullback-Leibler distance between TFR matrices is shown to be the most computationally expensive approach with the worst performance. Also, distance computation is carried out using the entire dataset, so the generalization capabilities of the classifier are poor, explaining the low performance on the EEG database. Local averaging techniques like the *t–f* grid approach and DWT averages are very easy to implement, fast to compute and yield acceptable results. These properties make them an appealing choice for making former tests on a database and for fixing TFR and wavelet parameters. Nevertheless. Nevertheless, as dynamical properties of the *t–f* features are lost, local averaging techniques cannot reach highest performance.
Vectorization transform methods are a straightforward application of the conventional linear transform methods to functional and matricial data. Advantages of this method are ease of use and accurate results. In fact, the best performance of the TFDF set is obtained using a vectorization approach, which can be explained by the unidimesional structure of these features that fits better the representation as vector. Difficulties lie in the management of huge data matrices which require large memory resources, for example, when covariance matrices should be computed in the 1D–PCA approach. Moreover, the performance outcomes are very different when supervised and non–supervised methods are used, specially when the TFDF set is used. Bidimensional transform reduces the memory requirements of data matrices, but increases the computational cost, as transform matrices should be computed for each dimension. These methods demonstrate stable performances for supervised and non–supervised methods, but seem to work inappropriately with TFDF features. Functional transform methods exhibit the best performance among the studied approaches. Computational requirements are low, the classification performance is high and stable, and lowest reduced representation space is achieved amongst all the considered dimensionality reduction methods. This approach is well suited to TFR–based classification, for both supervised and non–supervised linear transform methods. In general, all the linear transform methods can extract maximum information from the *t–f* feature set, thus achieving

adequate performance rates, but may require more attention the selection of an adequate reduced feature set size.
Relevance analysis is shown to be a highly beneficial tool for analyzing high dimensional data, as shown by the results on *t–f* feature sets from EEG biosignals. By means of the symmetrical uncertainty measure, it is possible to reduce the feature set by 50%, before linear transform methods are applied. Thus computational requirements and time are significantly reduced and numerical stability is improved. Besides, in most of the cases, the performance rates remain stable, and performance even improves in some cases. More stable behavior might be achieved, for example, if the relevance threshold were selected according to the accuracy rate of the classifier.

5. Conclusion

Throughout this chapter, several *t–f* based dimensionality reduction methods for classification of non–stationary biosignals were presented and discussed. A novel approach, based on feature selection by relevance analysis and linear transform of the most relevant features, was presented. Vectorized and bidimensional linear transform approaches were defined and reviewed, and the functional linear transform approach, specially designed to manage matricial and functional data, was proposed. The recommended *functional relevance* approach was tested on the detection of epileptic events on a public non–stationary electroencephalographic database. Comparison was made with other conventional state of the art approaches. Results show the benefits of the proposed functional linear transform approach and relevance analysis. Thus, it is demonstrated that the proposed *functional relevance* approach is able to exploit the multidimensional nature of data, while the relevance analysis facilitates the extraction of the most informative data.
Several questions are still open for further research. To begin with, the influence of the selected number of variables on the final classification performance is still to be analyzed. In this study, the explained variability of each eigenvector was used to select the size of the reduced feature set, but an automatic discriminative criterion could improve the classifier performance. On the other hand, the automatic selection of the relevance threshold is more difficult. A criterion, based on the classification rates could be used, but this increases the computational effort in the training stage. If the redundance of the feature set is also regarded, the most informative variables would be chosen as those that have the highest relevance value among similar redundant variables. In regard to the proposed functional linear transform, it is also important to compare different basis functions, as the final performance is compromised when coupling the basis function to the analyzed data. Finally, the importance of the relevance measure and the large influence that it has on the proposed methodology should be highlighted. For this reason, other approaches can also benefit if a relevance measure is considered in the design. Moreover, it is still to be analyzed the influence of different relevance measures on the performance of the methodology.

6. Acknowledgements

This research is carried out under grant "*Servicio de monitoreo remoto de actividad cardiaca para el tamizaje clinico en la red de telemedicina del departamento de Caldas*", funded by Universidad Nacional de Colombia and Universidad de Caldas.

7. References

Andrzejak, R., Lehnertz, K., Rieke, C., Mormann, F., David, P. & Elger, C. (2001). 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* 64: 71–86.

Avendano-Valencia, L. D., Godino-Llorente, J., Blanco-Velasco, M. & Castellanos-Dominguez, G. (2010). Feature extraction from parametric timeŰfrequency representations for heart murmur detection, *Annals of Biomedical Engineering* .

Avendaño-Valencia, L., Martínez-Vargas, J., Giraldo, E. & Castellanos-Domínguez, G. (2010). Reduction of irrelevant and redundant data from tfrs for eeg signal classification, *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*.

Aviyente, S. (2004). Information processing on the time-frequency plane, *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, pp. 617–620.

Barker, M. & Rayens, W. (2003). Partial least squares for discrimination, *Journal of chemometrics* 17(3): 166–173.

Bernat, E. M., Malone, S. M., Williams, W. J., Patrick, C. J. & Iacono, W. G. (2007). Decomposing delta, theta, and alpha time-frequency ERP activity from a visual oddball task using PCA., *International journal of psychophysiology : official journal of the International Organization of Psychophysiology* 64(1): 62–74.

Cvetkovic, D., Ubeyli, E. & Cosic, I. (2008). Wavelet transform feature extraction from human ppg, ecg, and eeg signal responses to elf pemf exposures: A pilot study, *Digital Signal Processing* 18(5): 861–874.

Debbal, S. & Bereksi-Reguig, F. (2007). Time-frequency analysis of the first and the second heartbeat sounds, *Applied Mathematics and Computation* 184(2): 1041 – 1052.

Englehart, K., Hudgins, B., Parker, P. A. & Stevenson, M. (1999). Classification of the myoelectric signal using time-frequency based representations, *Medical Engineering and Physics* 21(6-7): 431 – 438.

Hassanpour, H., Mesbah, M. & Boashash, B. (2004). Time-frequency based newborn EEG seizure detection using low and high frequency signatures, *Physiological Measurement* 25(4): 935–944.

M Tarvainen, J. Hiltunen, P. R.-a. & Karjalainen, P. (2004). Estimation of nonstationary eeg with Kalman smoother approach: An application to event-related synchronization, *IEEE Transactions On Biomedical Engineering* 51(3): 516–524.

Mallat, S. (2008). *A Wavelet Tour of Signal Processing*, 3rd edn, Academic Press.

Marchant, B. P. (2003). Time-frequency analysis for biosystems engineering, *Biosystems Engineering* 85(3): 261 – 281.

Michel, O., Baraniuk, R. & Flandrin, P. (64-67). Time-frequency based distance and divergence measures, *Proceedings of IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*.

Poulimenos, A. & Fassois, S. (2006). Parametric time-domain methods for non-stationary random vibration modelling and analysis – A critical survey and comparison, *Mechanical Systems and Signal Processing* 20(4): 763–816.

Quiceno-Manrique, A., Godino-Llorente, J., Blanco-Velasco, M. & Castellanos-Dominguez, G. (2010). Selection of dynamic features based on timeâĂŞfrequency representations

for heart murmur detection from phonocardiographic signals, *Annals of Biomedical Engineering* 38: 118–137. 10.1007/s10439-009-9838-3.

Sejdic, E., Djurovic, I. & Jiang, J. (2009). Time-frequency feature representation using energy concentration: An overview of recent advances, *Digital Signal Processing* 19(1): 153–183.

Sejdic, E. & Jiang, J. (2007). Selective Regional Correlation for Pattern Recognition, *IEEE Transactions on Systems, Man and Cybernetics* 37(1): 82–93.

Sepúlveda-Cano, L. M., Acosta-Medina, C. D. & Castellanos-Dominguez, G. (2011). Relevance analysis of stochastic biosignals for identification of pathologies, *EURASIP Journal on Advances on Signal Processing* 2011: 10.

Subasi, A. (2007). EEG signal classification using wavelet feature extraction and a mixture of expert model, *Expert Systems with Applications* 32(4): 1084–1093.

Tarvainen, M., Georgiadis, S., Lipponen, J., Hakkarainen, M. & Karjalainen, P. (2009). Time-varying spectrum estimation of heart rate variability signals with kalman smoother algorithm, *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pp. 1 –4.

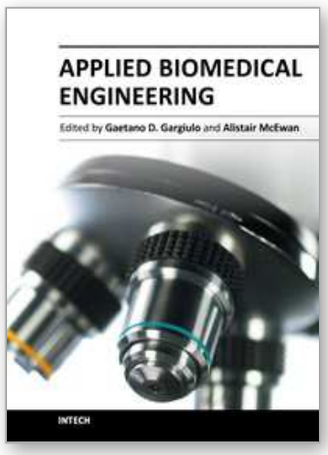
Turk, M. & Pentland, A. (1991). Eigenfaces for recognition, *Journal of Cognitive Neurosicence* 3(1): 71–86.

Tzallas, A. T., Tsipouras, M. G., Fotiadis, D. I. & Member, S. (2008). Epileptic seizure detection in electroencephalograms using time-frequency analysis, *IEEE Transactions on Information Technology in Biomedicine* 13(5): 703—-710.

Yang, J., Zhang, D., Frangi, A. & Yang, J. (2004). Two-dimensional PCA: A new approach to appearance-based face representation and recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(1): 131–137.

Yu, L. & Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy, *Journal of Machine Learning Research* 5: 1205–1224.

Zhang, D. & Zhou, Z.-H. (2005). (2D)2PCA: Two-directional two-dimensional PCA for efficient face representation and recognition, *Neurocomputing* 69(1-3): 224 – 231. Neural Networks in Signal Processing.



Applied Biomedical Engineering

Edited by Dr. Gaetano Gargiulo

ISBN 978-953-307-256-2
Hard cover, 500 pages
Publisher InTech
Published online 23, August, 2011
Published in print edition August, 2011

This book presents a collection of recent and extended academic works in selected topics of biomedical technology, biomedical instrumentations, biomedical signal processing and bio-imaging. This wide range of topics provide a valuable update to researchers in the multidisciplinary area of biomedical engineering and an interesting introduction for engineers new to the area. The techniques covered include modelling, experimentation and discussion with the application areas ranging from bio-sensors development to neurophysiology, telemedicine and biomedical signal classification.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Luis David Avendaño-Valencia, Carlos Daniel Acosta-Medina and Germán Castellanos-Domînguez (2011). Time-Frequency Based Feature Extraction for Non-Stationary Signal Classification, Applied Biomedical Engineering, Dr. Gaetano Gargiulo (Ed.), ISBN: 978-953-307-256-2, InTech, Available from: http://www.intechopen.com/books/applied-biomedical-engineering/time-frequency-based-feature-extraction-for-non-stationary-signal-classification

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen