



RENIECYT - LATINDEX - Research Gate - DULCINEA - CLASE - Sudoc - HISPANA - SHERPA UNIVERSIA - E-Revistas - Google Scholar
DOI - REDIB - Mendeley - DIALNET - ROAD - ORCID

Title: Revisión del estado del arte de la extracción de características de señales electroencefalográficas
A review of the state of the art of feature extraction of electroencephalographic signals

Authors: ESQUEDA-ELIZONDO, José Jaime, JIMÉNEZ-BERISTÁIN, Laura, CHÁVEZ-GUZMÁN, Carlos Alberto, GUERRA-FRAUSTRO, Ricardo Jesús Renato

Editorial label ECORFAN: 607-8695
BCIERMMI Control Number: 2019-325
BCIERMMI Classification (2019): 241019-325

Pages: 13
RNA: 03-2010-032610115700-14

ECORFAN-México, S.C.
143 – 50 Itzopan Street
La Florida, Ecatepec Municipality
Mexico State, 55120 Zipcode
Phone: +52 1 55 6159 2296
Skype: ecorfan-mexico.s.c.
E-mail: contacto@ecorfan.org
Facebook: ECORFAN-México S. C.
Twitter: @EcorfanC

www.ecorfan.org

Holdings		
Mexico	Colombia	Guatemala
Bolivia	Cameroon	Democratic
Spain	El Salvador	Republic
Ecuador	Taiwan	of Congo
Peru	Paraguay	Nicaragua

Introduction

Methodology

Pre-Processing

Coherence

Features

EEG time Variations Considerations

Feature Extraction

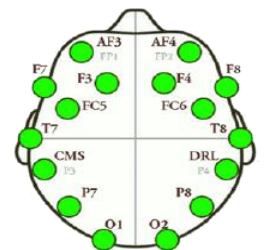
Experiment in User B

Conclusions

Acknowledgements

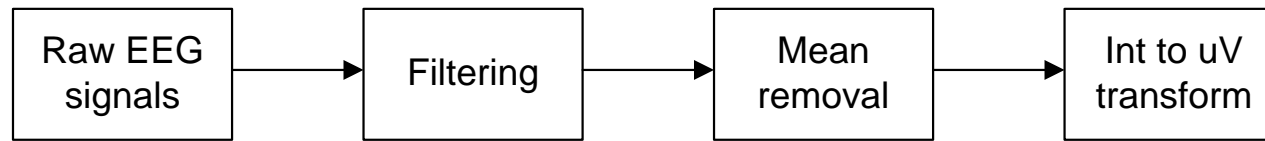
Introduction

- With the appearance of a variety of commercial Brain Computer Interfaces (BCI) systems, there has been a need to carry out research in the EEG pattern detection.
- BCI systems can be used in applications for medical, smart environments, detection, rehabilitation, gaming, education, authentication and other areas.
- This paper has the intention to help other people that are beginning in EEG signal processing.
- BCI systems use electroencephalographic (EEG) sensors (electrodes placed on the scalp) to acquire these signals and send them to an external device where it can be transformed into instructions or commands.
- Generally, a BCI system is divided in the following processes: signal acquisition, pre-processing, feature extraction and classification.
- The key factors that define the effectiveness of a BCI system are the feature extraction and the classification stages.
- BCIS are made with one to 256 electrodes and there are wired or wireless.
- There are open source BCIs like OpenBci or others like Epoc Emotiv, that require a license.



Pre-Processing

- Preprocessing stage generally includes eliminating the isoelectric lines of each channel, filtering and converting data from integer to a floating point that represents the real voltage level measured.



- Using digital signal processing techniques, it is possible to reduce noise, interference and the artifacts present in the signal. This must be done before the feature extraction stage.

Features

The features are numeric parameters used to describe signals, in this case, the EEG signals. Some common features are: amplitude of the EEG signals, band power, power spectrum density, autoregressive parameters, time-frequency features and features based in inverse models. Feature extraction allows us to obtain useful or descriptive information hidden in a signal by decreasing the unnecessary or redundant information.

Feature considerations:

- Noise and atypical values: EEG signals have noise and they have atypical values because of the low signal to noise ratio.
- High dimensionality: feature vectors normally have high dimensionality because they come from different channels and from several time segments before they can be concatenated in only one feature vector.
- Time information: the extracted features have time information like the patterns related to some variations in a specific time window of the EEG.
- Non stationarity: EEG signals are non-stationary because they change rapidly with respect to time.
- Small training sets: datasets are relatively small because the training process requires time and data.

EEG Time Variations Considerations

Most of the brain activity patterns used in a BCI system are related to particular time variations of the signal, possibly in specific frequency bands. For this reason, it is necessary to consider the timeline during the feature extraction stage:

- Concatenation of different time segments features: this consists of feature extractions from different time segments and concatenating them into one single features vector.
- Combination of classificatory algorithms in different time segments: this consists of making the feature extraction and the classification in different time segments and then combining the results of the different classifiers
- Dynamic classification: this consists of extracting features from different time segments in order to construct a temporal sequence of feature vectors.

The feature extracting techniques can be grouped in:

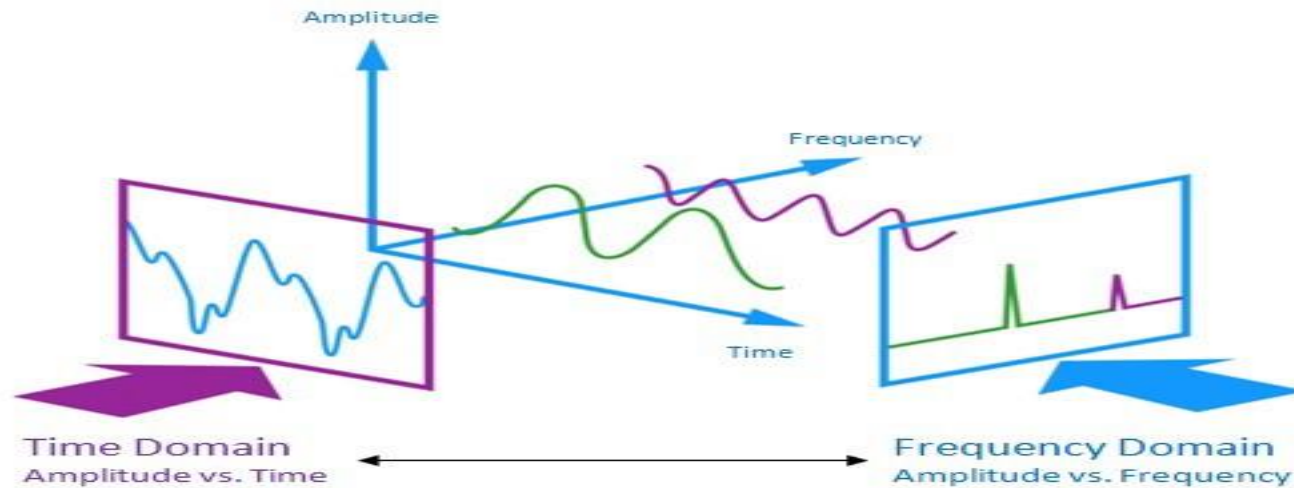
- Time domain (1960s)
- Frequency domain (1960s-1980s)
- Joint time – frequency domain (1980s-2000s)
- Signal decomposition and sparse domains (2000- today)

Time Domain Feature Extraction Techniques

- The basic features that can be extracted are their statistical properties, such as mean, variance, standard deviation, quadratic mean value, etc.
- There are also some other feature extraction time domain techniques, such as cross correlation, Autoregressive modelling (AR), Linear Predictive Coding (LPC), Cepstrum analysis and Kernel based modelling.
- AR modeling and LPC calculate the future values as a function of the current and past values.
- The Cepstrum analysis is based on the rate of change of different frequency spectrum bands of the analyzed signal. It is similar to the homomorphic filtering, where the signals are transformed by joint additions and multiplication operations.
- One of the advantages of the AR modelling is that it makes an enhanced data compaction and for spectral peaks detection, and also reduces signal noise and it provides a very good resolution. The order of the AR model cannot be determined a priori.
- Cepstrum does not need to calculate the Fourier Transform of the signal, so it is considered a time domain processing. Cepstrum requires that the analyzed signal must be stationary over the time interval. This technique has been used in seismic, voice and geo-physical signals.

Time Domain Feature Extraction Techniques

- The power Cepstrum feature vector contains the Mel-frequency Cepstral Coefficients (MFCC) that are calculated by transforming the spectrum into Mel scale.
- These coefficients capture the spectral energy measurements over short time windows. On the other hand, Cepstrum tends to lose the non-stationary values of the spectrum, like the time variant or transient patterns in the analyzed signals.



Frequency Domain Feature Extraction Techniques

- These techniques transform the signals from time domain to frequency domain in order to extract low level features of a windowed signal.
- But sometimes these characteristics may not represent the real stationarity of the signal and only obtain global information that approximately classifies the signals.

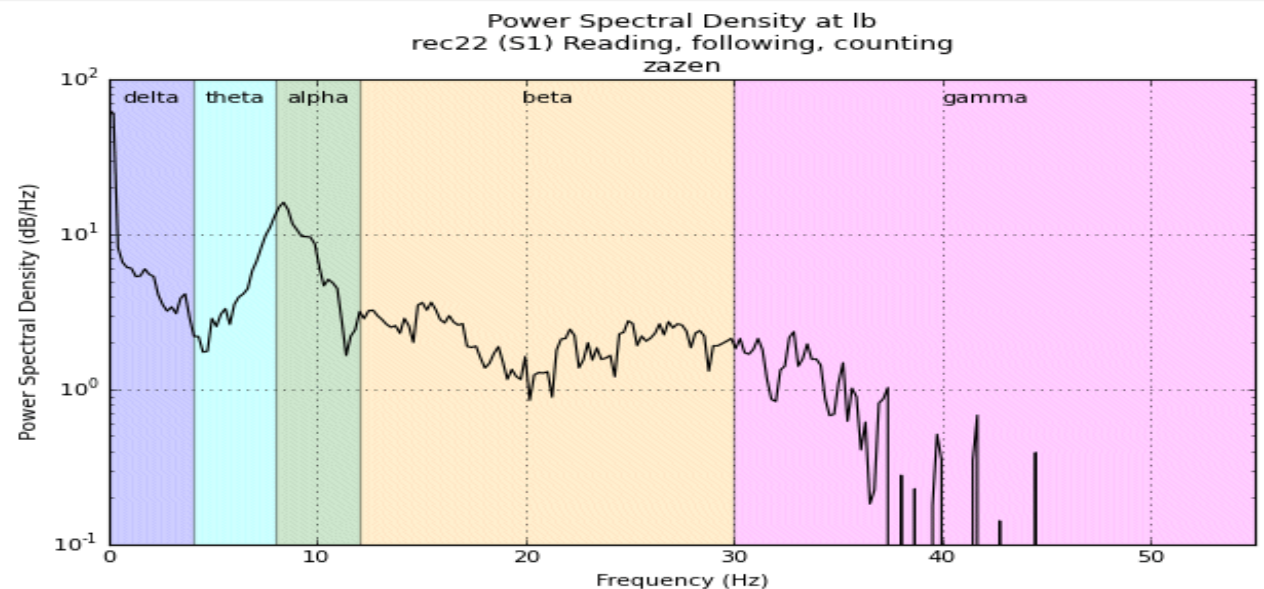
Power Spectrum Density

- The power spectrum density (PSD) function shows how the power of a signal is distributed in the frequency.
- This can be obtained with the Wiener-Khinchine Theorem or with the Short Time Fourier Transform. The Wiener-Khinchine Theorem express that the PSD and the Correlation are Fourier Transform pairs, as it is shown in:

$$S_{xx}(f) = \mathfrak{F}[R_{xx}(\tau)]$$

$$R_{xx}(\tau) = \mathfrak{F}^{-1}[S_{xx}(f)]$$

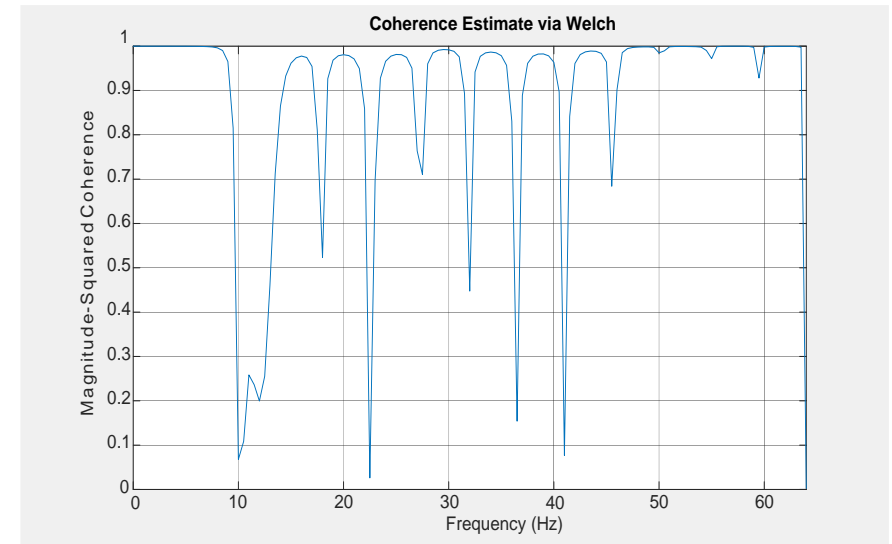
- Where $S_{xx}(f)$ is the self-spectrum or PSD of the signal $x(t)$ and $R_{xx}(\tau)$ is the autocorrelation of the signal $x(t)$.



Coherence

- Coherence is a frequency function given in normalized units between 0 and 1, that shows how much one signal $x(t)$ corresponds in power to another signal $y(t)$ for each frequency. Is a quadratic correlation coefficient that estimates the amplitude and phase consistence between two signals for every frequency.
- When the signal $x(t)$ entirely corresponds to the signal $y(t)$, that means that are the same signal. And is given by:

$$\Gamma^2(\mathbf{f}) = \frac{|S_{xy}(\mathbf{f})|^2}{S_{xx}(\mathbf{f})S_{yy}(\mathbf{f})}; \quad 0 \leq \Gamma(\mathbf{f}) \leq 1.$$



- Any pair of signals can be coherent in some frequency bands and have no coherence in another bands.
- Contrasting with amplitude measurements, coherence is the measurement of the synchronization between two signals principally based in the phase consistence.
- For each frequency, the coherence measures when the signals are related by a linear time invariant transformation, which means that they have a constant amplitude and phase displacement rate

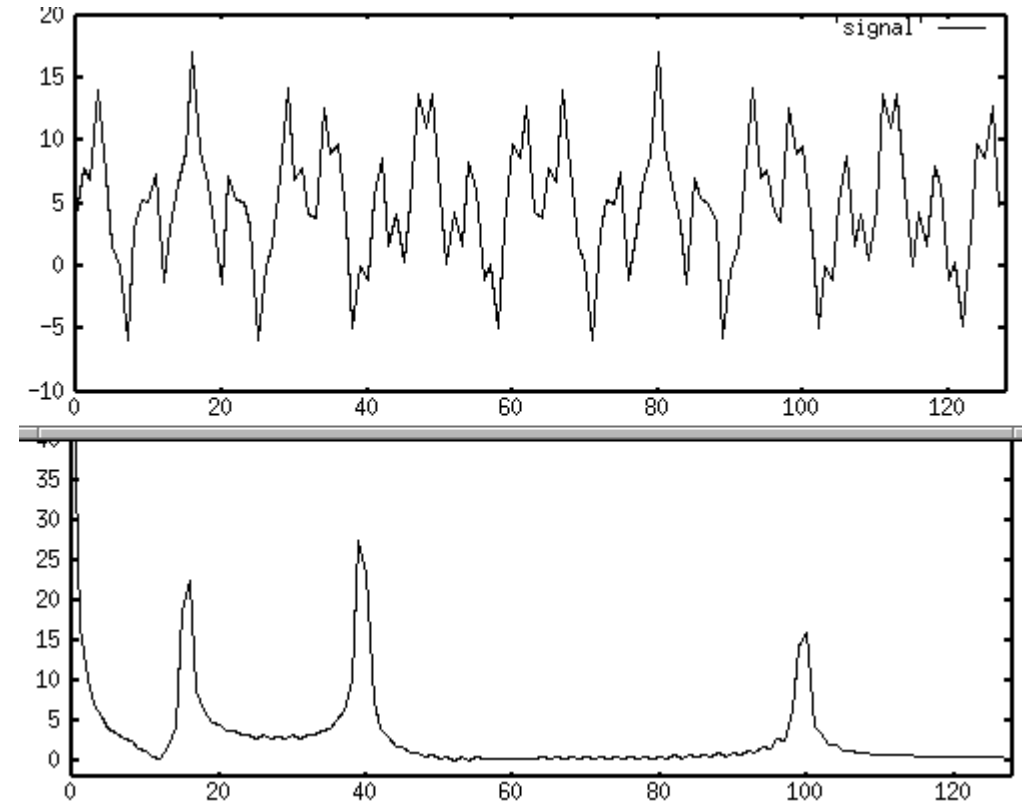
Discrete Cosine Transform

The Discrete Cosine Transform (DCT) is a variant of the Discrete Fourier Transform (DFT), in which the signal is decomposed in a sum of cosine functions (not in sine and cosine functions as in the FDT) and these cosine coefficients are the interested features that describe the signals. The Discrete Cosine Transform is defined as:

$$C(u) = \alpha(u) \sum_{x=0}^{N-1} f(x) \cos\left(\frac{(2x+1)u\pi}{2N}\right)$$

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} & \text{para } u = 0 \\ \sqrt{\frac{2}{N}} & \text{para } u = 1, 2, \dots, N-1 \end{cases}$$

The DCT is usually used to compress information because of its energy compaction property



Wavelet Transform

The Wavelet Transform is a mathematical tool that can carry out a Time-Frequency analysis of a signal. This is an advantage compared with Discrete Fourier Transform based techniques because with Fourier we can know which frequential components are present in the signal, but we do not know in which moment they appeared. With Wavelet Transform provides that information.

The Wavelet Transform is defined by:

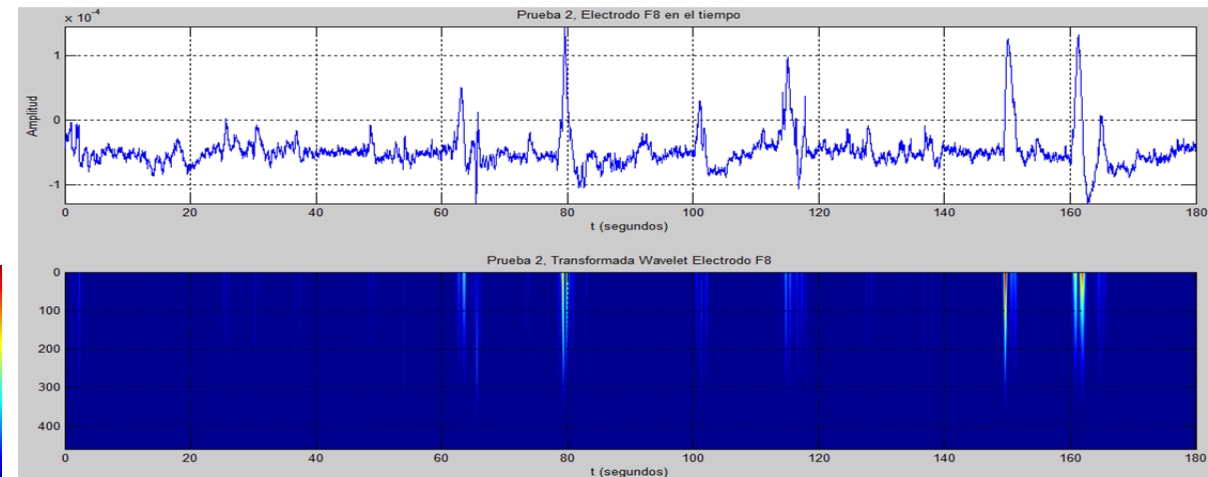
$$W_f(s, \tau) = \int f(t) \psi_{s,\tau}^*(t) dt \quad \text{where:} \quad \psi_{s,\tau}^*(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \text{ is the mother Wavelet.}$$

The Continuous Wavelet Transform (CWT) is used in EEG signals to reduce noise in order to obtain a clearer and more precise signal. It can also be used to observe the Time-Frequency relationship of a signal. This Transform shows the moments with higher power in function of both time and frequency, representing them graphically in scalograms.

$$\iint |CWT(\tau, \alpha)|^2 \frac{d\tau, d\alpha}{\alpha^2} = E_x \quad , \text{ where: } E_x = \int |x(t)|^2 dt$$

For a signal $x(t)$, if the FDT is given by the relation $W(t, \omega)$, so the total energy of the signal will be:

$$\iint W(t, \omega) = \int |x(t)|^2 dt = \int |X(\omega)|^2 d\omega \quad)$$



Entropy

The entropy is a parameter that measures the randomness of a chaotic system or signal. Due to the complexity and non-linearity of the EEG signals, entropy can be used to analyze them. There are different types of entropy. For example, Shannon or log-energy entropies are used to measure the irregularities of EEG signals.

Shannon entropy (HSE) measures the average information contained in the probability distribution function (PDF).

$$H_{SE} = - \sum_{k=1}^N p_k^2 \times \log_2(p_k^2)$$

Log-energy entropy (HLE) describe the electro-physiological behavior of different kinds of EEG signals.

$$H_{LE} = \sum_{k=1}^N \log_2(p_k^2)$$

The Renyi entropy (HRE) has been used to derive the continuous family of information measures.

$$H_{RE} = \frac{1}{1-\alpha} \log \sum_{k=1}^N p_k^\alpha$$

The Tsallis Wavelet entropy (HTE) can extract some improved features by reducing the negative effects of the wavelet aliasing.

$$H_{TE} = \frac{1}{\alpha - 1} \left(1 - \sum_{k=1}^N p_k^\alpha \right)$$

The Renyi entropy and the Tsallis entropy are of order α , where $\alpha > 0$ and $\alpha \neq 1$. If $\alpha = 2$, the estimation equally emphasizes the sub-Gaussian and the super-Gaussian components and is known as quadratic Renyi entropy and quadratic Tsallis entropy.

Machine Learning

- Machine learning, also known as automatic learning is defined as the use of algorithms for data organization, pattern recognition in order for the computer to learn with these models and solve tasks without pre-programming.
- Machine learning is the computer's science area that defines Artificial Intelligence.
- Machine learning algorithms learn from the input data, and with these data, the machines are trained in order to execute different tasks autonomously.
- Deep learning is the part of the automatic learning that using high level algorithms, imitates the neural network of the brain.
- These are complex algorithms made from several layers of neurons fed by a huge volume of data and are capable of image and speech recognition and to execute advanced tasks without human intervention.

Experiences using Epoc BCI

During this research we worked with the Epoc+ made by the Emotiv Co. One of the lessons we learned with this project is that when we work with kids with Autistic Spectrum Disorder (ASD), first we have to spend some time trying to make the ASD user accepts and tolerates the headset. We recommend that a visual inspection always be done first in order to eliminate and always record a video of the EEG recording process to have a timeline that can be used to locate movements that can generate disturbances.



ECORFAN®

© ECORFAN-Mexico, S.C.

No part of this document covered by the Federal Copyright Law may be reproduced, transmitted or used in any form or medium, whether graphic, electronic or mechanical, including but not limited to the following: Citations in articles and comments Bibliographical, compilation of radio or electronic journalistic data. For the effects of articles 13, 162,163 fraction I, 164 fraction I, 168, 169,209 fraction III and other relative of the Federal Law of Copyright. Violations: Be forced to prosecute under Mexican copyright law. The use of general descriptive names, registered names, trademarks, in this publication do not imply, uniformly in the absence of a specific statement, that such names are exempt from the relevant protector in laws and regulations of Mexico and therefore free for General use of the international scientific community. BCIERMMI is part of the media of ECORFAN-Mexico, S.C., E: 94-443.F: 008- (www.ecorfan.org/ booklets)