Technological Educational Institute of Crete | Department of Informatics Engineering

Master of Science

Machine learning data preparation for epileptic seizures prediction

Andriana Theodorakopoulou

Approved by:
Dr. Giorgos Papadourakis

June 2017

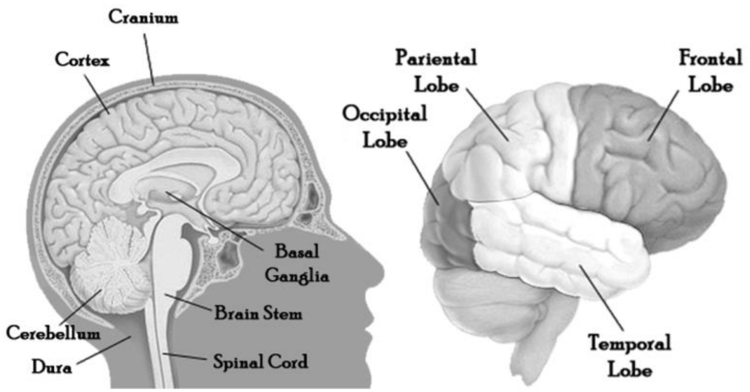
Outline

- Introduction
- Seizures Prediction
- Machine Learning
- Data Preparation
- Feature Engineering
- Results
- Future Work

Objective of the thesis

- This study aims to investigate and extract features for epilepsy prediction seizures emphasizing on data preparation of the appropriate EEG signals.
- The presented algorithms were applied on the CHB-MIT Scalp EEG database which was collected at the Children's Hospital Boston (CHB).
- The database is consisted of scalp EEG recordings from pediatric subjects with intractable seizures.

What is Epilepsy



- Epilepsy is the second most common brain disorder and affects approximately 1% of the world's population.

- It is characterized by the occurrence of ***unforeseeable*** and ***uncontrollable*** seizures.

- When a seizure takes place, a disruption of the electrical impulses in the brain is observed.

States of Epilepsy Seizures and Symptoms

A seizure often has five separate phases:

- The ***prodromal stage,*** includes mostly emotional signs. This prodromal group of symptoms befalls days or hours before a seizure follows.
- In an ***aura***, changes in activity, emotions, hearing, smell, taste, visual perception are involved. Auras are a small partial seizure that is often followed by a larger event. They usually come a few seconds to a few minutes before the actual seizure.

States of Epilepsy Seizures and Symptoms

- ***Pre-ictal*** period, refers to the state immediately before the actual seizure, stroke, or headache, though it has recently come to light that some characteristics of this stage (such as visual auras) are actually the beginnings of the ictal state.
- ***Ictal*** period, refers to a physiologic state or event such as a seizure, stroke, or headache. In electroencephalography (EEG), the recording during a seizure is said to be ”ictal”.
- The ***Post-ictal****,* this is the healing period after the seizure. Some people recover immediately, while others may take minutes to hours to feel like their usual self.

Diagnosis And Treatment

- Epilepsy is usually difficult to diagnose quickly. In most cases, it cannot be confirmed until you have had more than one seizure.

- It can be difficult to diagnose because many other conditions, such as migraines and panic attacks, can cause similar symptoms.

- Some of the most important pieces of information needed to diagnose epilepsy are the details about the seizure or seizures.

Diagnosis And Treatment

The most common ways to verify and determine the type of the epilepsy are the following:

- Electroencephalogram (EEG): Can detect unusual brain activity associated with epilepsy by measuring the electrical activity of your brain through electrodes placed on your scalp.

- Magnetic resonance imaging (MRI) scan: It can often detect possible causes of the condition, such as defects in the structure of your brain or the presence of a brain tumour.

Diagnosis And Treatment

Epilepsy treatment is based on the type and the grade of the seizures that a person has:

- Anti-Epileptic Drugs: Approximately 30% of the patients remain unresponsive.
- Epilepsy surgery: Requires long-term electrophysiological evaluation.
  - only 50% of the candidates undergo surgery
  - only 60% of surgery cases result in seizure free
- Electrical stimulation: Vagus Nerve Stimulation (VNS)

Open Problems

- Is the seizure occurrence random or there are patterns?
- Can we create a mechanism that will be able to predict seizures? How precise will be? What about false positive feedbacks?
- Are there any seizure pre-cursors? What measurements can be used to indicate them?

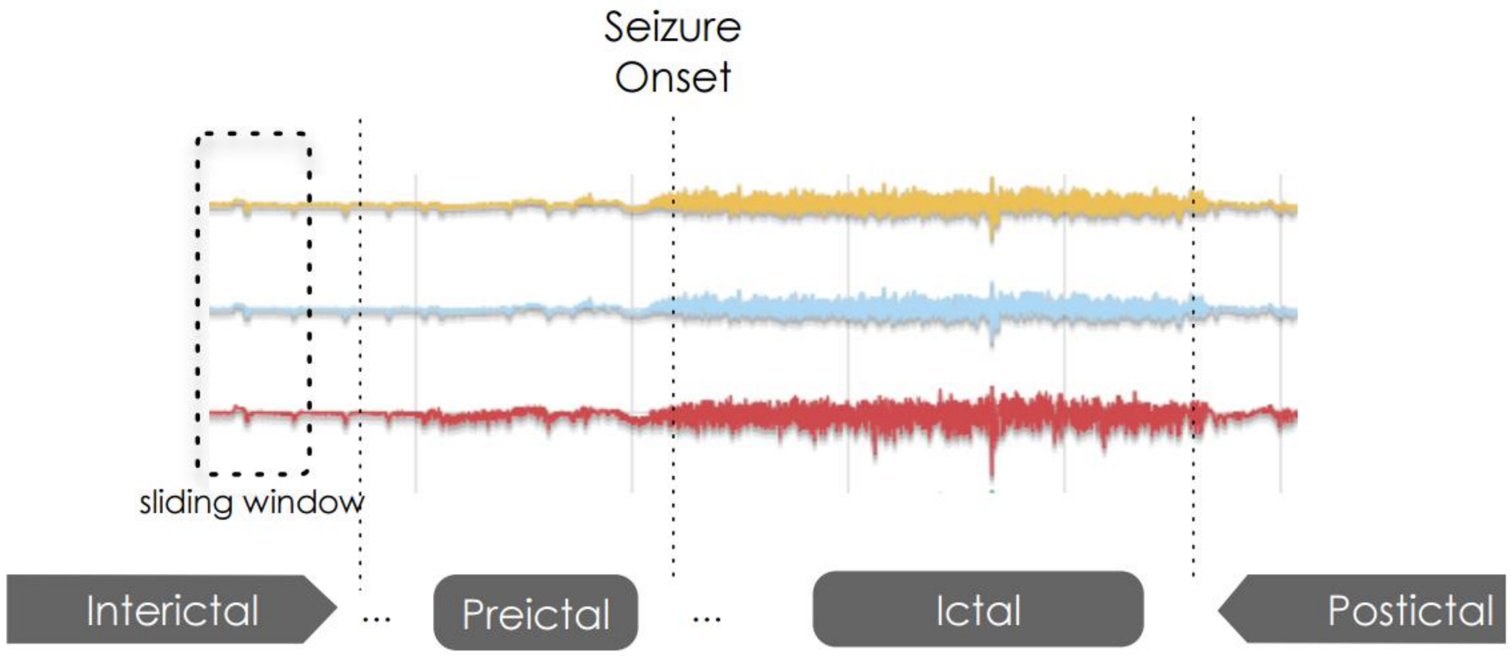
Seizures Prediction

Why seizures prediction is important?

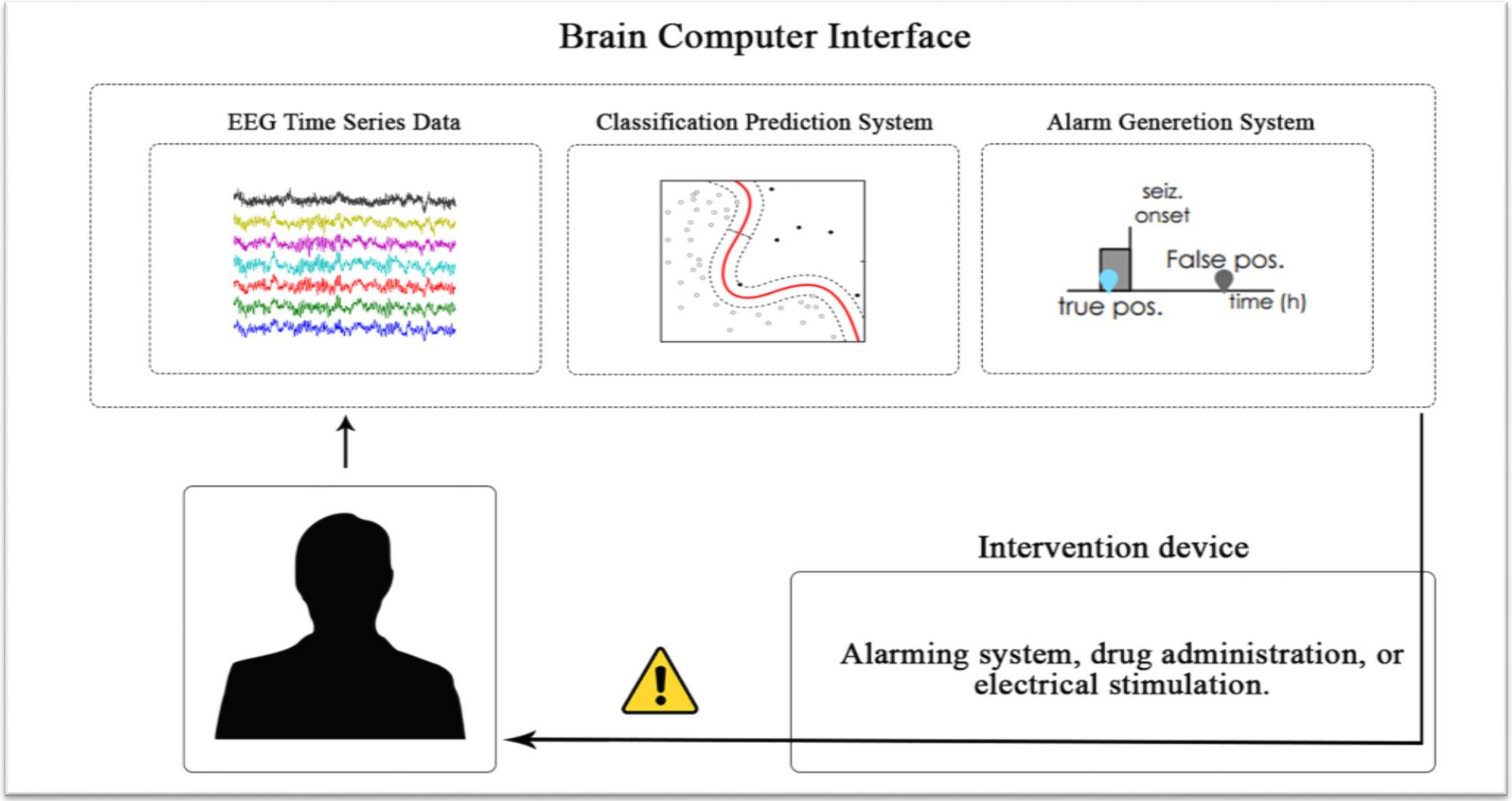
- It can improve the patients quality of life.
- Good prediction mechanism would be beneficial for closed-loop therapeutic systems.

Background: EEG Time Series

- EEG - evaluation of the physiological state of the brain
  - spatial and temporal resolution
  - characterization the rapidly changing electrical activity of the Brain
  - Electrode placement: scalp and/or intracranial



Seizures Prediction



Brief History

More than 50 years ago

- Seizures were believed to be abruptly
- Some patients report auras and prodromes symptoms

1970

- Viglione were the first to propose seizure prediction from the EEG 1970 - 1975

Brief History

80’s – 90’s

- Improvement of technology
- Implementation of non linear methods
- First method suggesting that seizure prediction is possible

Last 15 years

- Many methods have been proposed, most of them retrospective

Now

- Need of long term continues recordings
- Prospective analysis
- Improving technical and methological analysis

Related Work

Year	Author	Methods	Sensitivity
1998	Lehnertz and Elger	Effective correlation dimension	94 %
1999	Le Van Quyen et al.	Similarity index	83 %
2000	Le Van Quyen et al.	Similarity index	94 %
2001	Iasemidis et al.	Lyapunov exponent	91 %
2002	Schindler et al.	LIFU (surface EEG)	100 %
2003	Mormann et al.	Phase coherence, cross correlation	86 %
2004	VP Nigam, D Graupe	Nonlinear filtering	92,2 %
2005	Srinivasan et al.	Time and frequency features, ANN	99,6 %
2006	Guler and Ubeyli	PSD features, modified mixture of experts	98,6 %
2007	Tzallas et al.	Time frequency analysis, ANN	96,3 %
2008	Polat and Gunes	PCA-FFT, AIRS classifier	100 %
2009	Ocak	Wavelet transform, ApEn	94,5 %

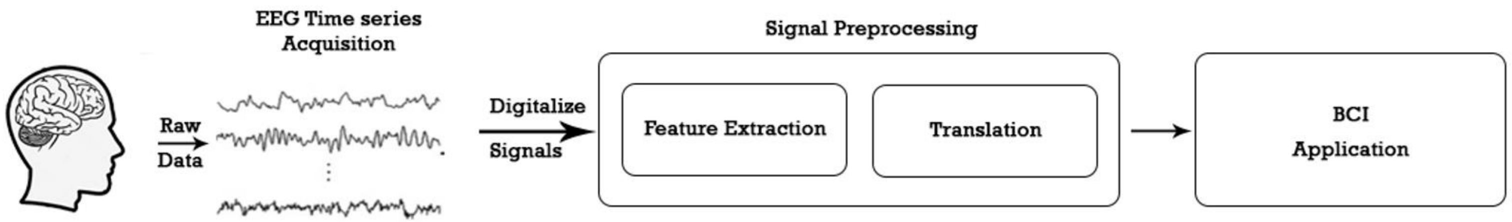
Related Work

Year	Author	Methods	Sensitivity
2010	Kumar et al.	Entropy measures, Recurrent Elman network (REN)	99,75 %
2011	Fathima et al.	Discrete wavelet transform	99,4 %
2012	Martis et al.	EMD features	95,33 %
2013	Alam et al.	EMD statistics, ANN	100 %
2014	Schindler et al.	Hilbert Huang transform, Bayesian Classiffiers	96,55 %
2015	Martis et al.	wavelet leader based scaling and cumulant estimation	80,5 %

Machine Learning

- Machine learning is a subfield of computer science that, according to Arthur Samuel in 1959, gives *"computers the ability to learn without being explicitly programmed.*

- A machine learning project may not be linear, but it has a number of well known steps:
  - Define Problem.
  - **Prepare Data.**
  - Evaluate Algorithms.
  - Improve Results.
  - Present Results.



Why data preparation is important?

- Data preparation is crucial for any data analysis. If your data is messy, there's no way you can make sense of it.

- Computers are great at handling large, even enormous data sets, speedy computing and recognizing patterns. But they fail miserably if you give them the wrong input.

Why data preparation is important?

- Data Preparation is the heart of data science. It includes data cleansing and feature engineering.

- Domain knowledge is also very important to achieve good results.

Data Preparation = Data Cleansing + Feature Engineering

Why data preparation is important?

Data Cleansing puts data into the right shape and quality for analysis. It includes many different functions, for example the following:

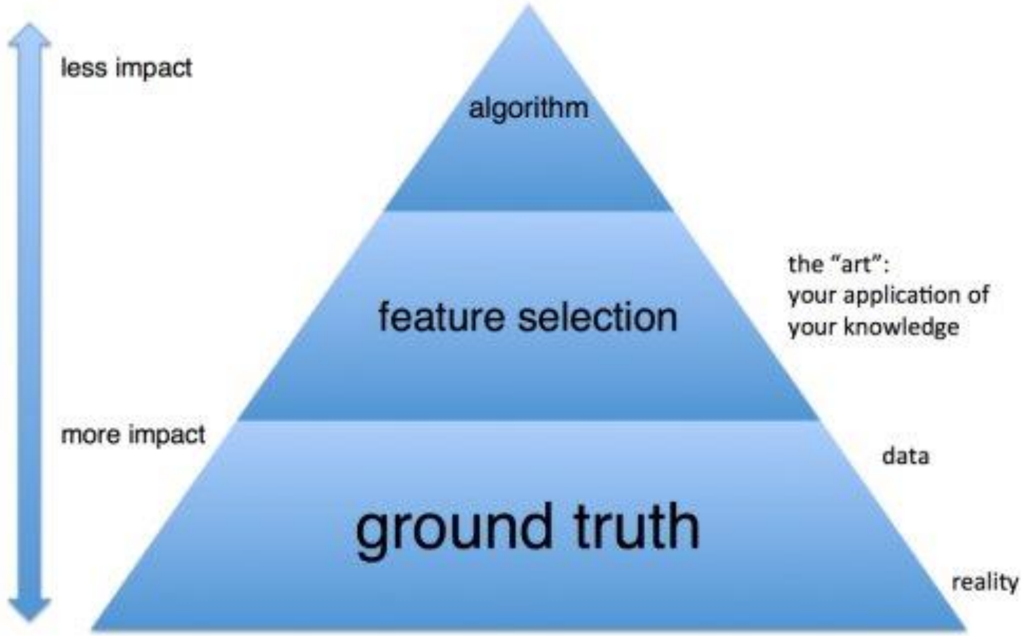
- Basics (select, filter, removal of duplicates, …)
- Sampling (balanced, stratified, ...)
- Data Partitioning (create training + validation + test data set, ...)
- Transformations (normalisation, standardisation, scaling, pivoting, ...)
- Binning (count-based, handling of missing values as its own group, …)
- Data Replacement (cutting, splitting, merging, ...)
- Weighting and Selection (attribute weighting, automatic optimization, ...)
- Attribute Generation (ID generation, ...)
- Imputation (replacement of missing observations by using statistical algorithms)

Why data preparation is important?

Feature Engineering selects the right attributes to analyze. You use domain knowledge of the data to select or create attributes that make machine learning algorithms work. It includes:

- Brainstorming or testing of features
- Feature selection
- Validation of how the features work with your model
- Improvement of features if needed
- Return to brainstorming / creation of more features until the work is done

Data Preparation is Key for Success in Machine Learning Projects



Machine learning algorithms

Step 1: Data Selection

- The ***CHB-MIT Scalp EEG Database*** is one of the most cited resources used in prediction and detection experiments. It is also one of the few publicly available invasive EEG datasets.

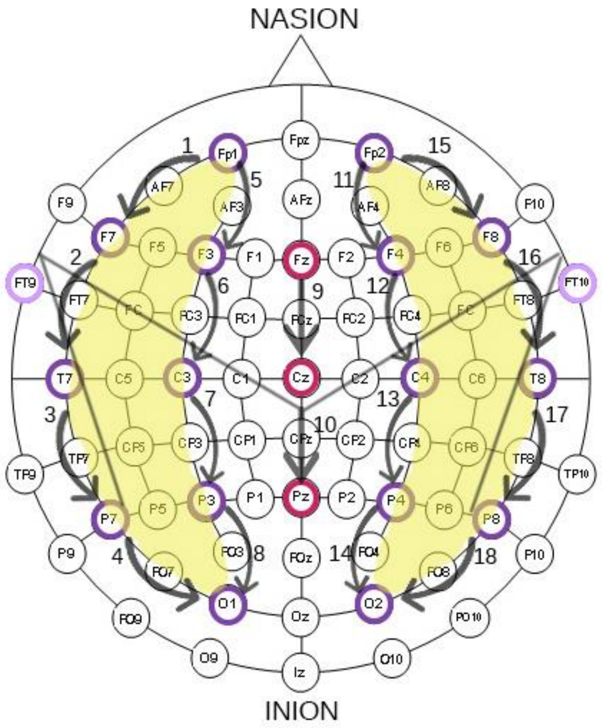
- The database contains 24 hour-long continuous pre-surgical invasive EEG recordings of ***22 patients*** *(5 males, 17 females)* suffering from epilepsy.

- All the subjects were monitored for up to several days following withdrawal of anti-seizure medication.

Step 1: Data Labels

In order to perform measurements for feature extraction, it is needed to localize the channels from the dataset recordings, which contain the raw brain waves with some technical specification of each signal.

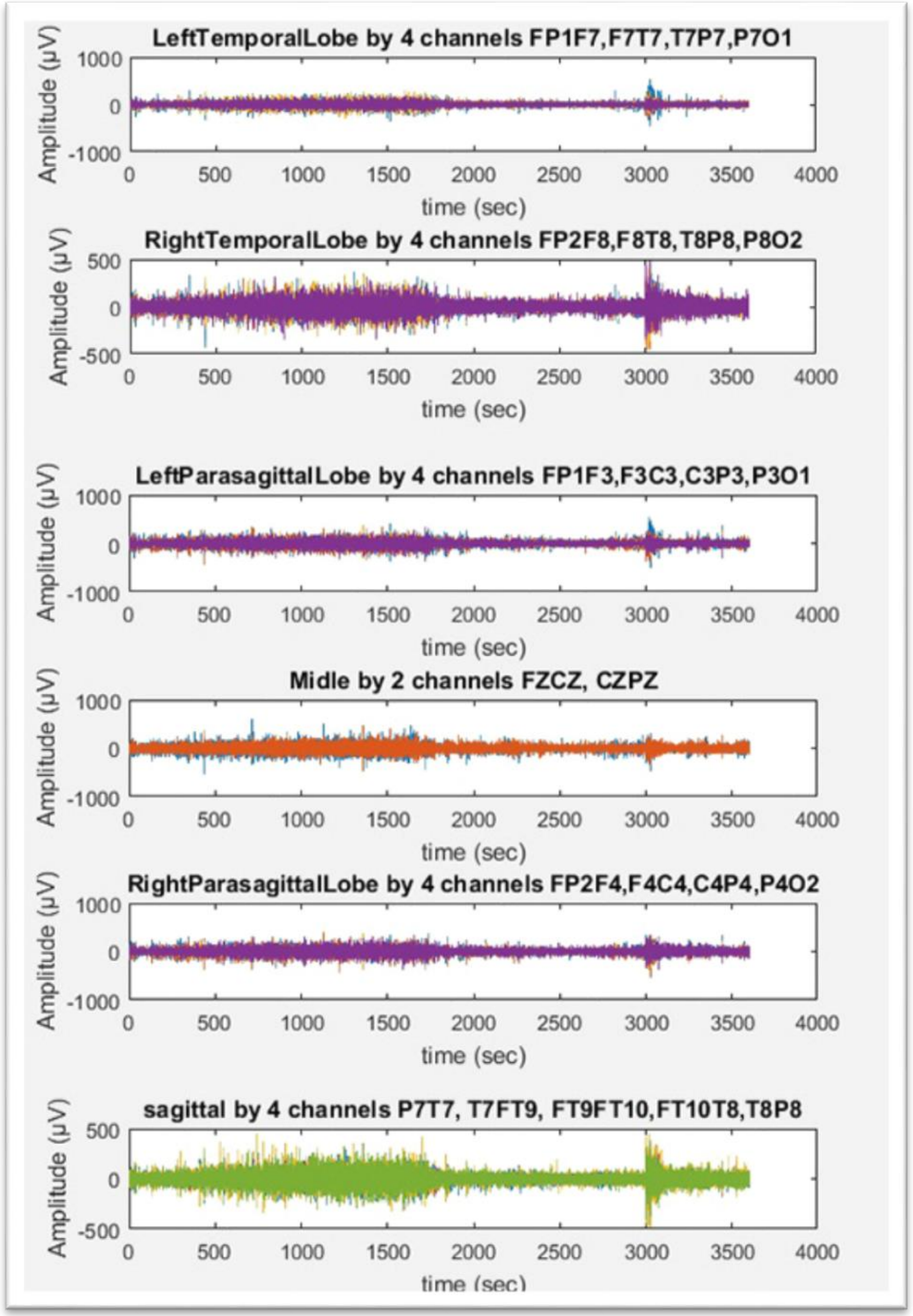
Bipolar montage “double banana”



Channel	Plane
1 - 4	Left Temporal
5 - 8	Left Parasagittal
9 - 10	Sagittal or Midline
11 - 14	Right Parasagittal
15 - 18	Right Temporal

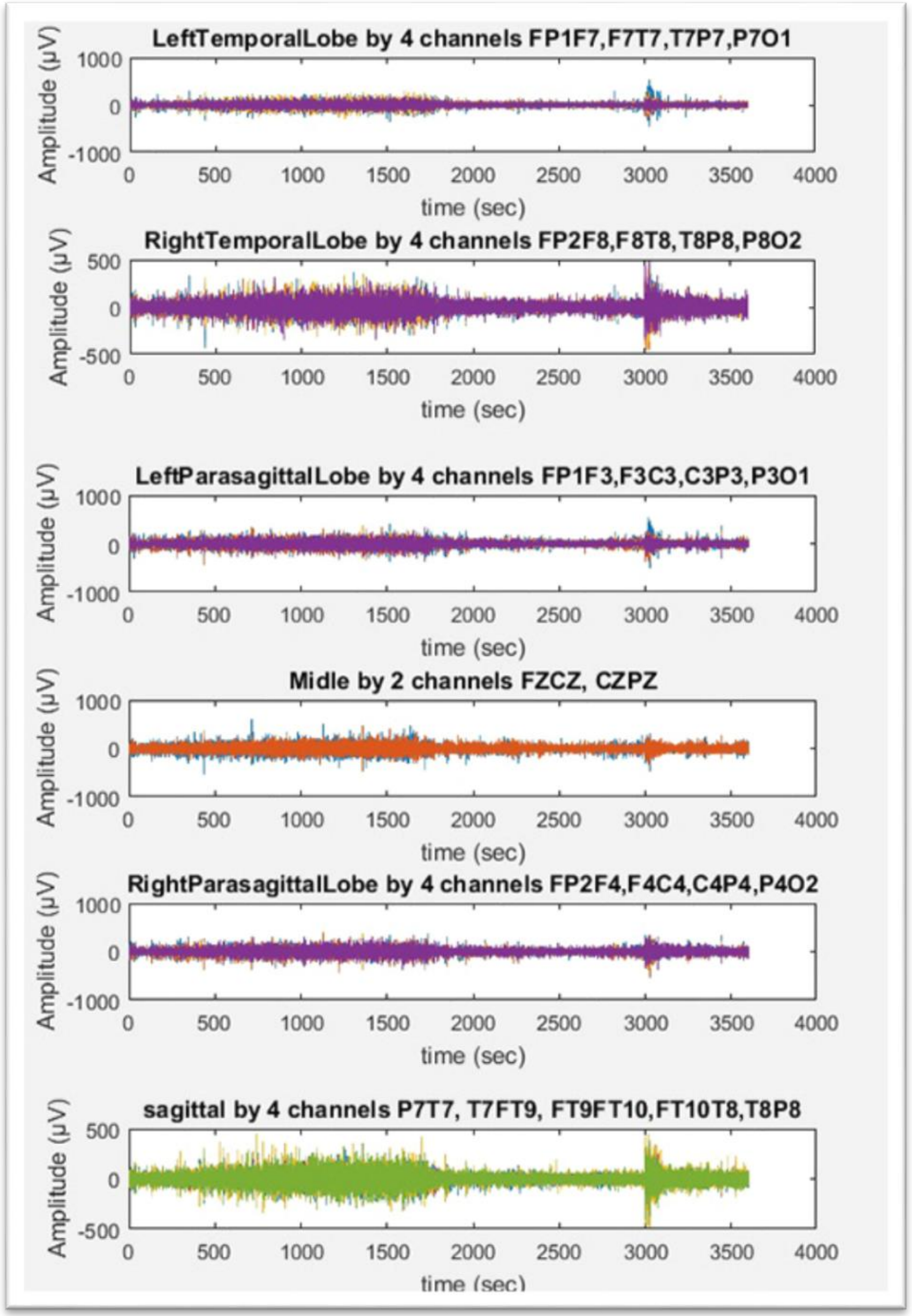
Step 1: Data Labels

Channel Number	Differential Electrodes	Longitudinal bipolar montage double banana
1	FP1-F7	
2	F7-T7	
3	T7-P7	Left Temporal
4	P7-O1	
5	FP1-F3	
6	F3-C3	
7	C3-P3	Left Parasagittal
8	P3-O1	
9	FP2-F4	
10	F4-C4	
11	C4-P4	Right Parasagittal
12	P4-O2	



Step 1: Data Labels

Channel Number	Differential Electrodes	Longitudinal bipolar montage double banana
13	FP2-F8	
14	F8-T8	
15	T8-P8	Right Temporal
16	P8-O2	
17	FZ-CZ	
18	CZ-PZ	Sagittal or Midline
19	P7-T7	
20	T7-FT9	
21	FT9-FT10	Mid-Sagittal
22	FT10-T8	
23	T8-P8	



Step 2: Sampling and Filtering

- Resampling is mandatory for the EEG recordings. Using a custom made algorithm we produce several datasets of different sample rates.
  - 32Hz, 64Hz, 128 Hz, 256Hz, 320Hz and of course the original 512 Hz

  This action saves memory, computational requirements and removes some noise. However this action removes signal information.

Step 2: Sampling and Filtering

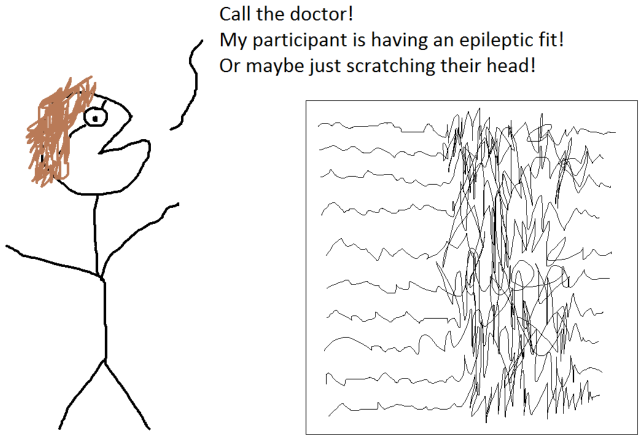
Using Digital Signal Processing methods:

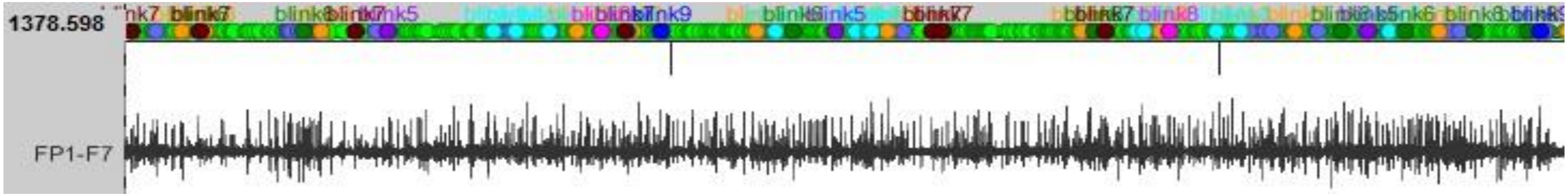
- Remove the frequencies 0 - 4 Hz owing to noise in most channels. This action is based on several bibliographic sources.
- And the frequencies above 45 Hz using a low-pass filter.

Step 2: Remove Artifacts

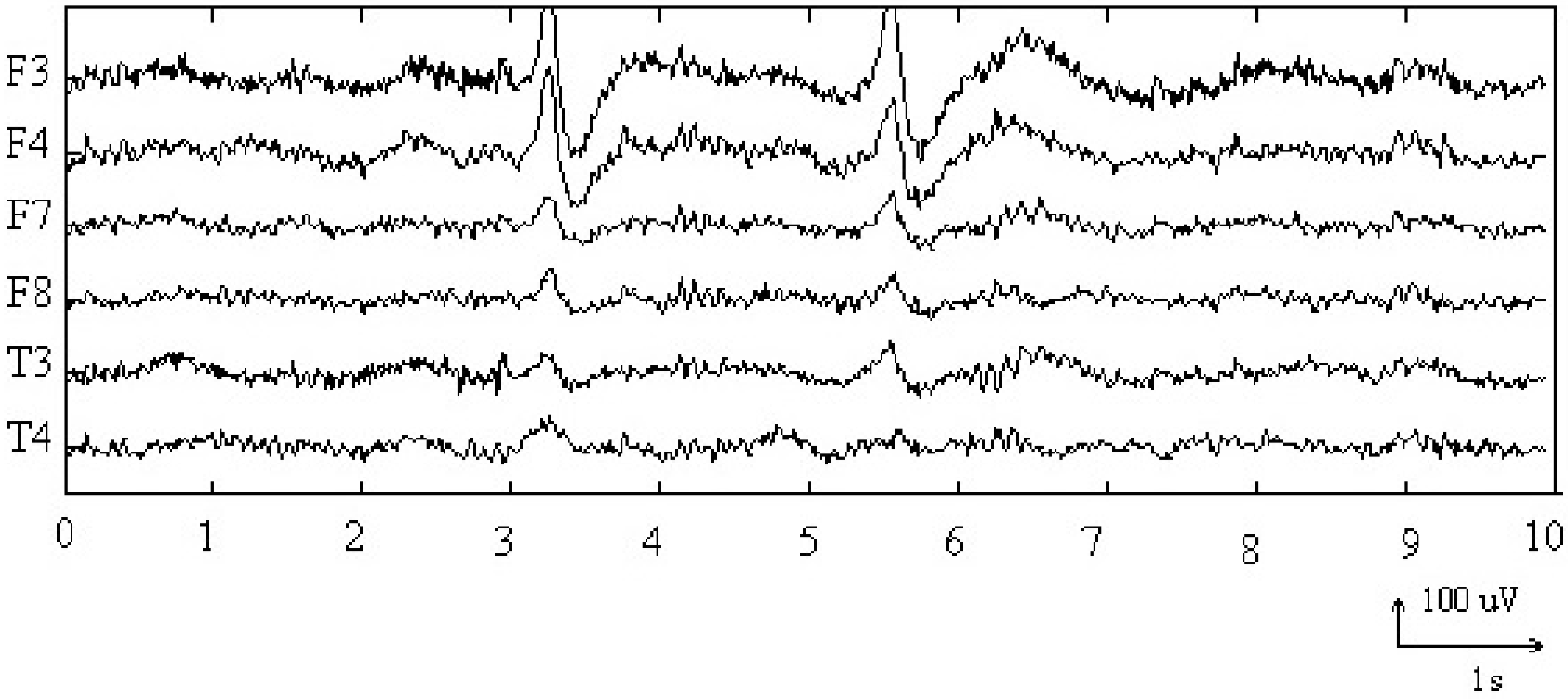
The noises of artifacts affect the EEG signals and the structure of the waveform.

- Eye Blinks
- Eye movement
- Muscular artifacts
- Electrode artifacts
- Detect and remove eye blinks using Brainstorm toolbox



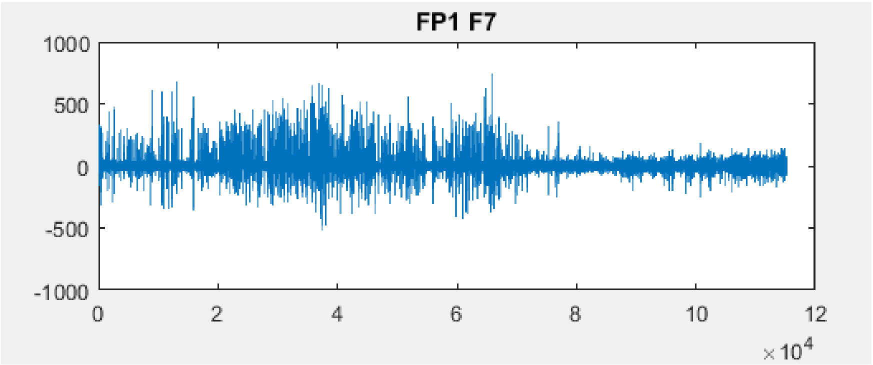


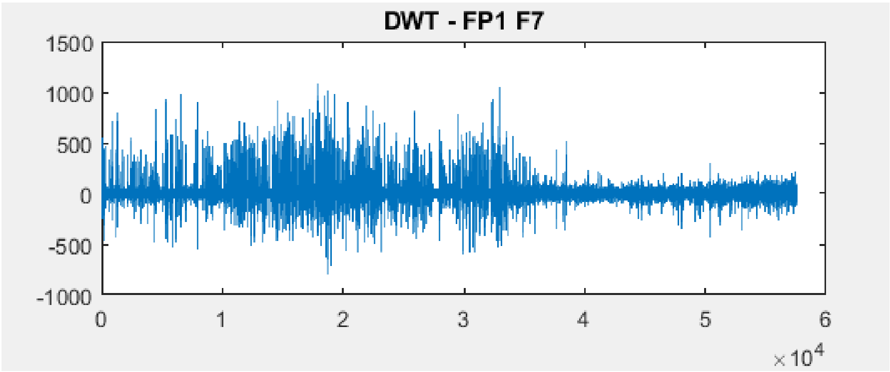
Step 2: Remove Artifacts



Step 3: Transform Data

- Discrete Wavelet transformation is a linear time-frequency transform and suitable for analysis of transient and non-stationary phenomena as well as noise reduction.
- Decompose a given signal $x(t)$ into increasingly finer detail based on two sets of basis functions, the wavelets and the scaling functions
- Localize information in both time and frequency





Feature Engineering

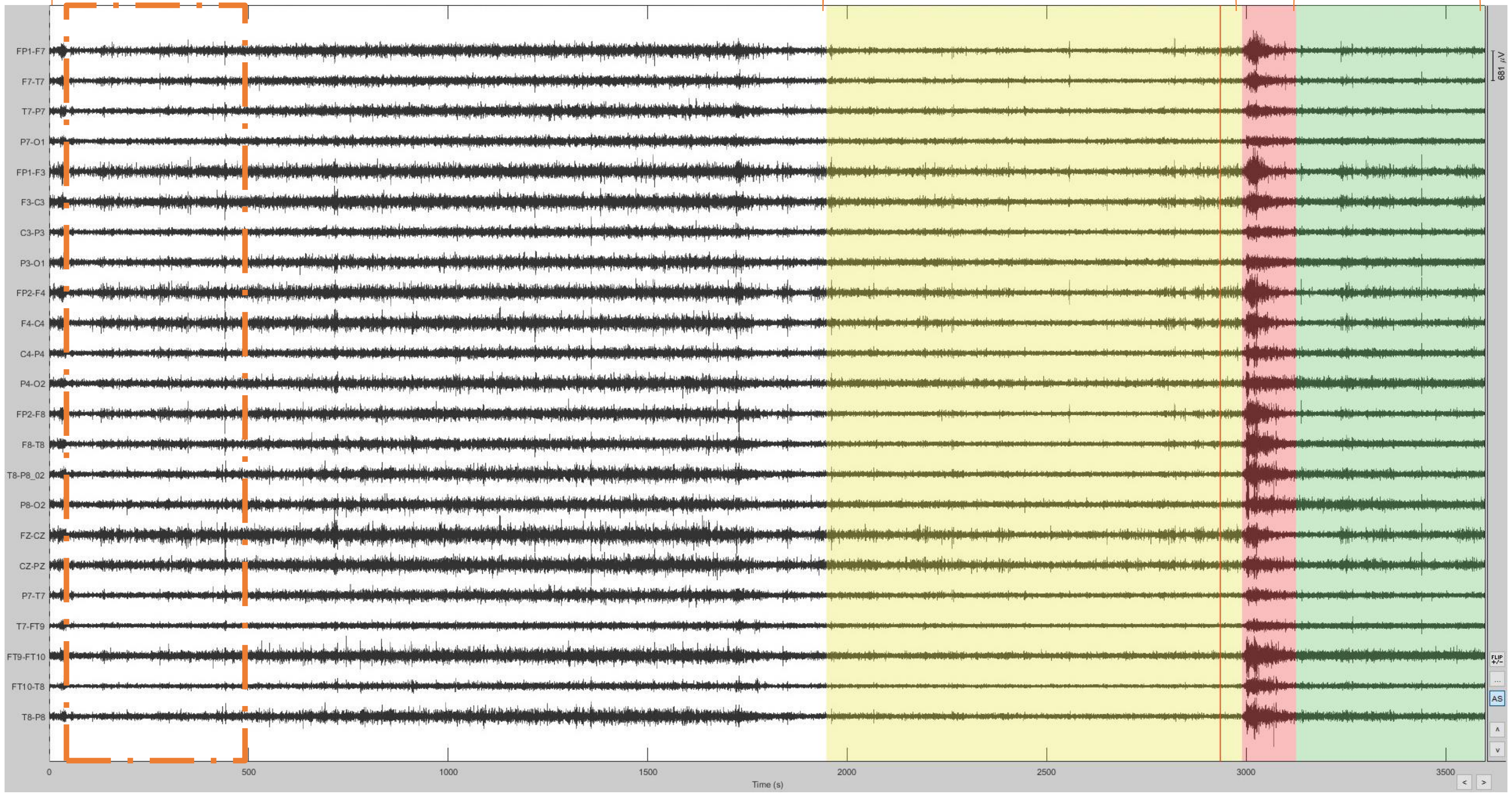
EEG time series epochs & Moving Window Analysis

Inter Ictal

Pre Ictal

Ictal

Post Ictal



Sliding Window 5’’

Seizure Onset

Feature Extraction methods

The following list presents some of the most known features extraction methods:

- Statistical Measures
- Hjorth Parameters
- Accumulated Energy
- Autoregressive Models
- Coherence
- Lyapunov Exponent
- Dynamical Similarity Index
- Entropies
- Correlation Dimension
- Brain Wave Patterns

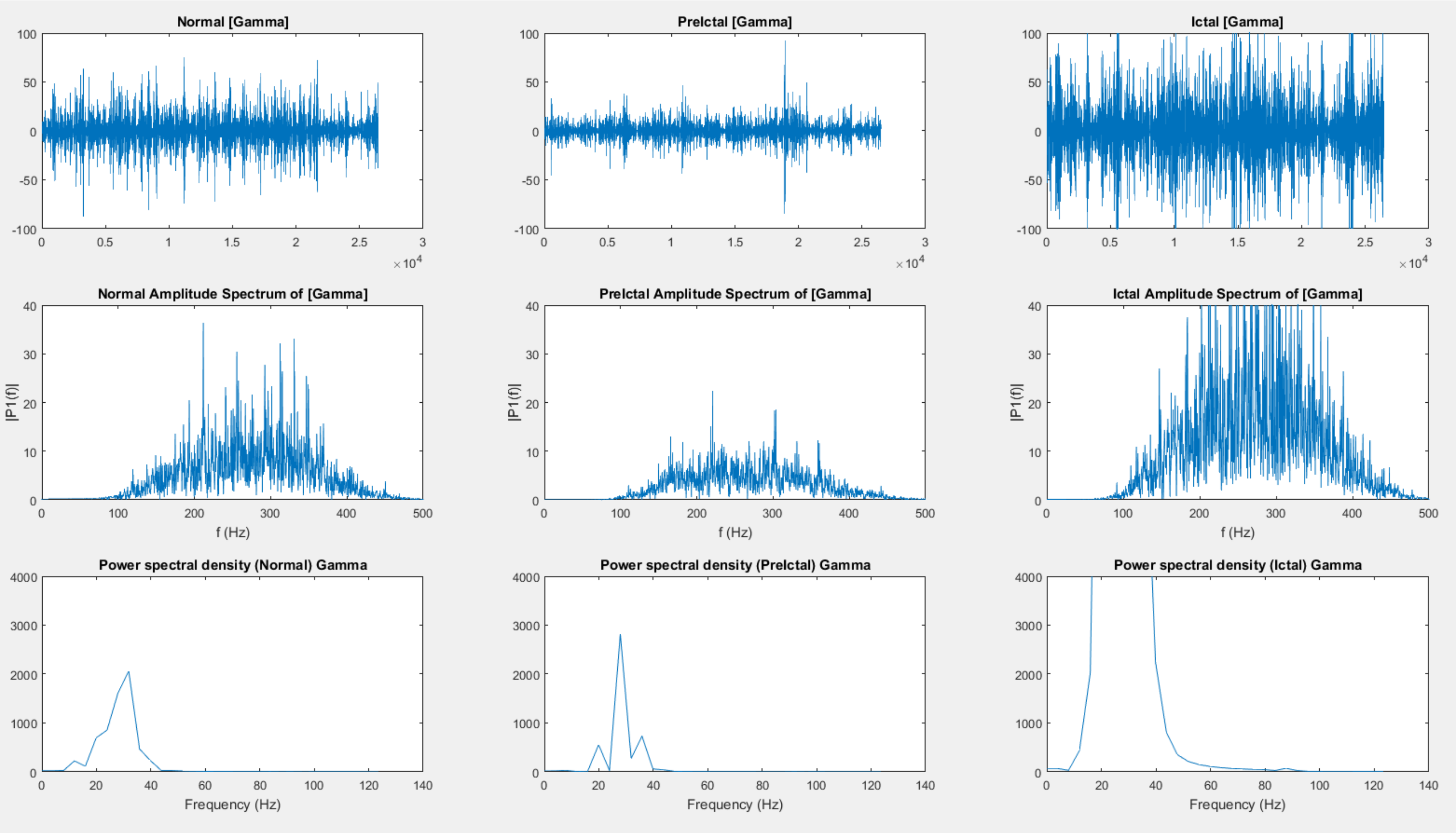
Brain Waves

- Each brain wave has a purpose and helps serve us in optimal mental functioning.
- If one of the five types of brain waves is either overproduced or underproduced in our brain, it can cause problems.
- Often, the onset of a clinical seizure is characterized by a sudden change of frequency in the EEG measurement.
- It is normally within the alpha wave frequency band with slow reduction in frequency but increase in amplitude during the seizure period.

Brain Waves

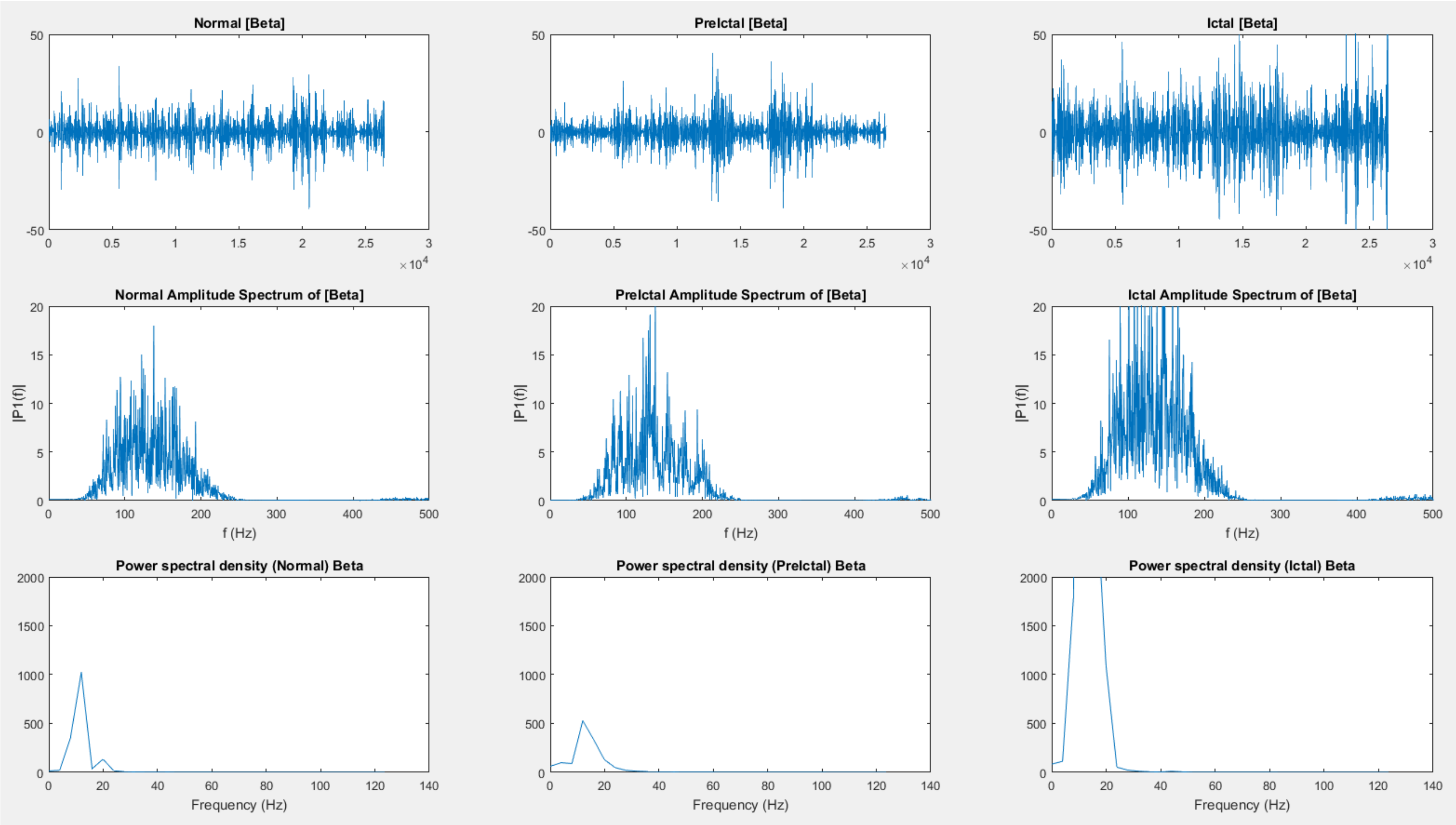
	Gamma	Beta	Alpha	Theta	Delta
Frequency Range Hz	40 - 100	12 - 40	8 - 12	4 - 8	0 - 4
Too much	Anxiety, high arousal, stress	Adrenaline, anxiety, high arousal, inability to relax, stress	Daydreaming, inability to focus, too relaxed	Depression, hyperactivity, impulsivity, inattentiveness	Brain injuries, learning problems, Inability to think
Too little	Depression, learning disabilities	Daydreaming, depression, poor cognition	Anxiety, high stress, insomnia	Anxiety, poor emotional awareness, stress	Inability to rejuvenate body, inability to revitalize the brain, poor sleep

Brain Waves - Gamma waves



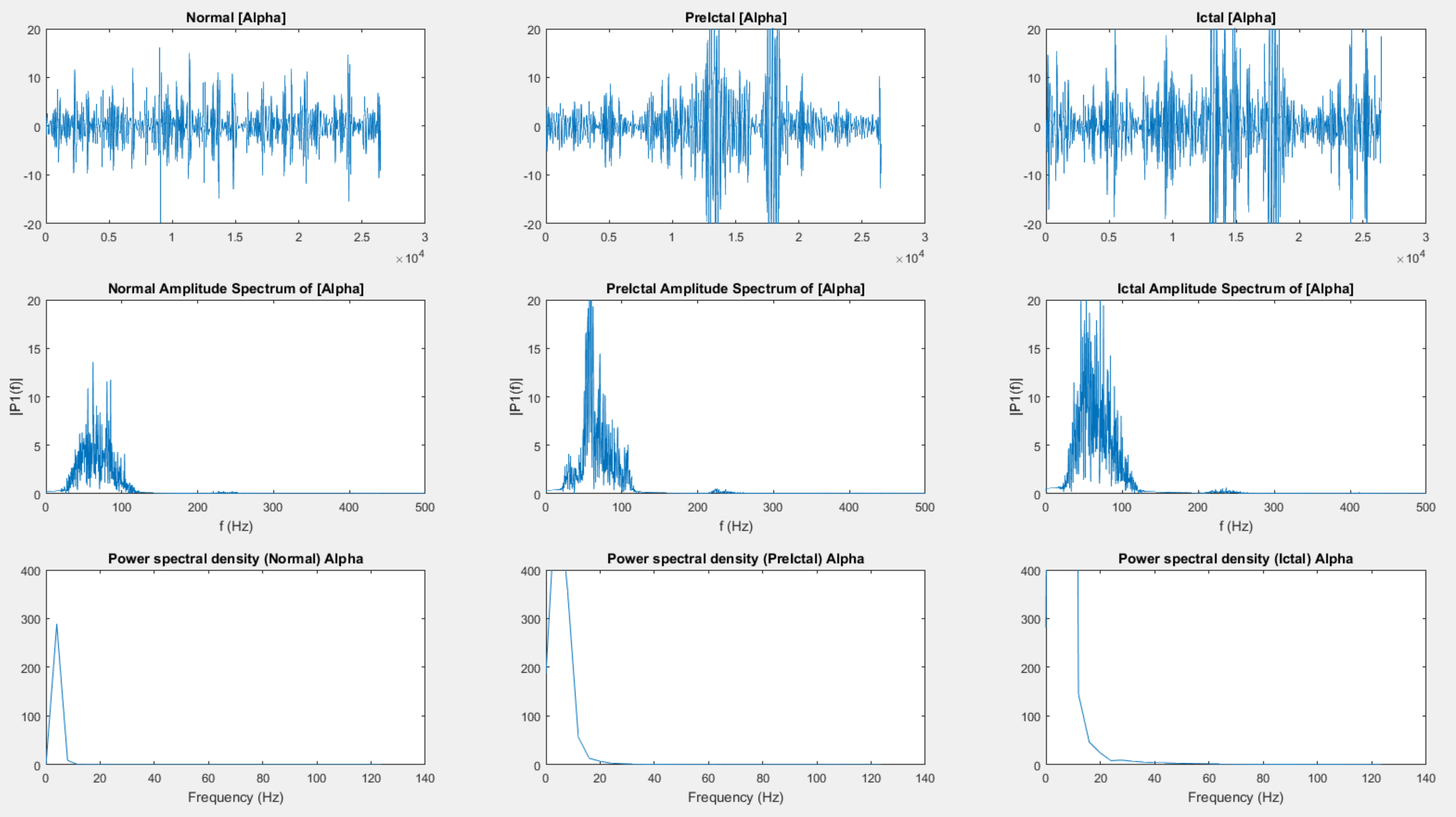
- Too Little: Depression
- Too Much: Anxiety, high arousal, stress

Brain Waves - Beta waves



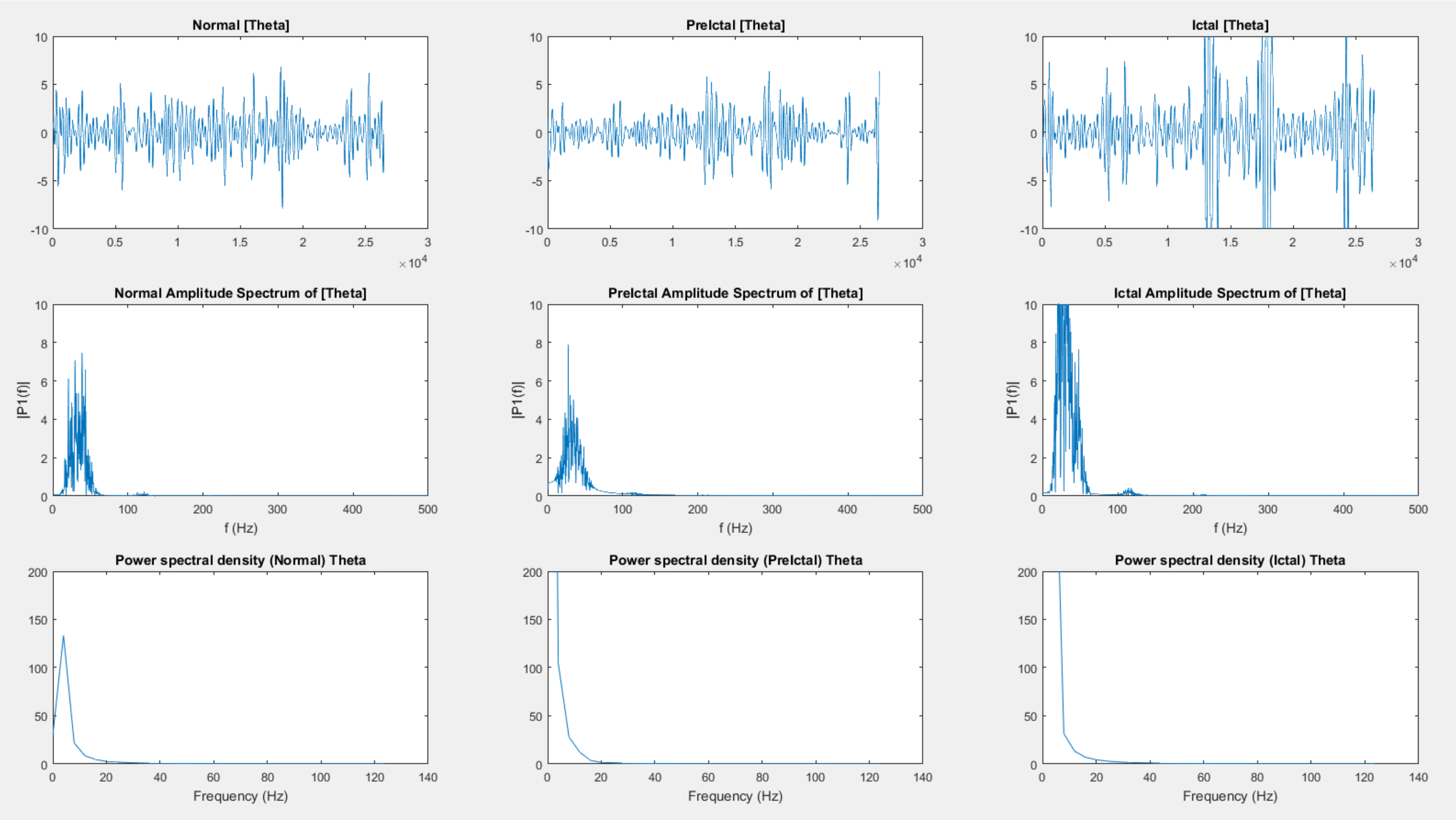
- Too Little: Daydreaming, depression, poor cognition
- Too Much: Adrenaline, anxiety, high arousal, inability to relax, stress

Brain Waves - Alpha waves



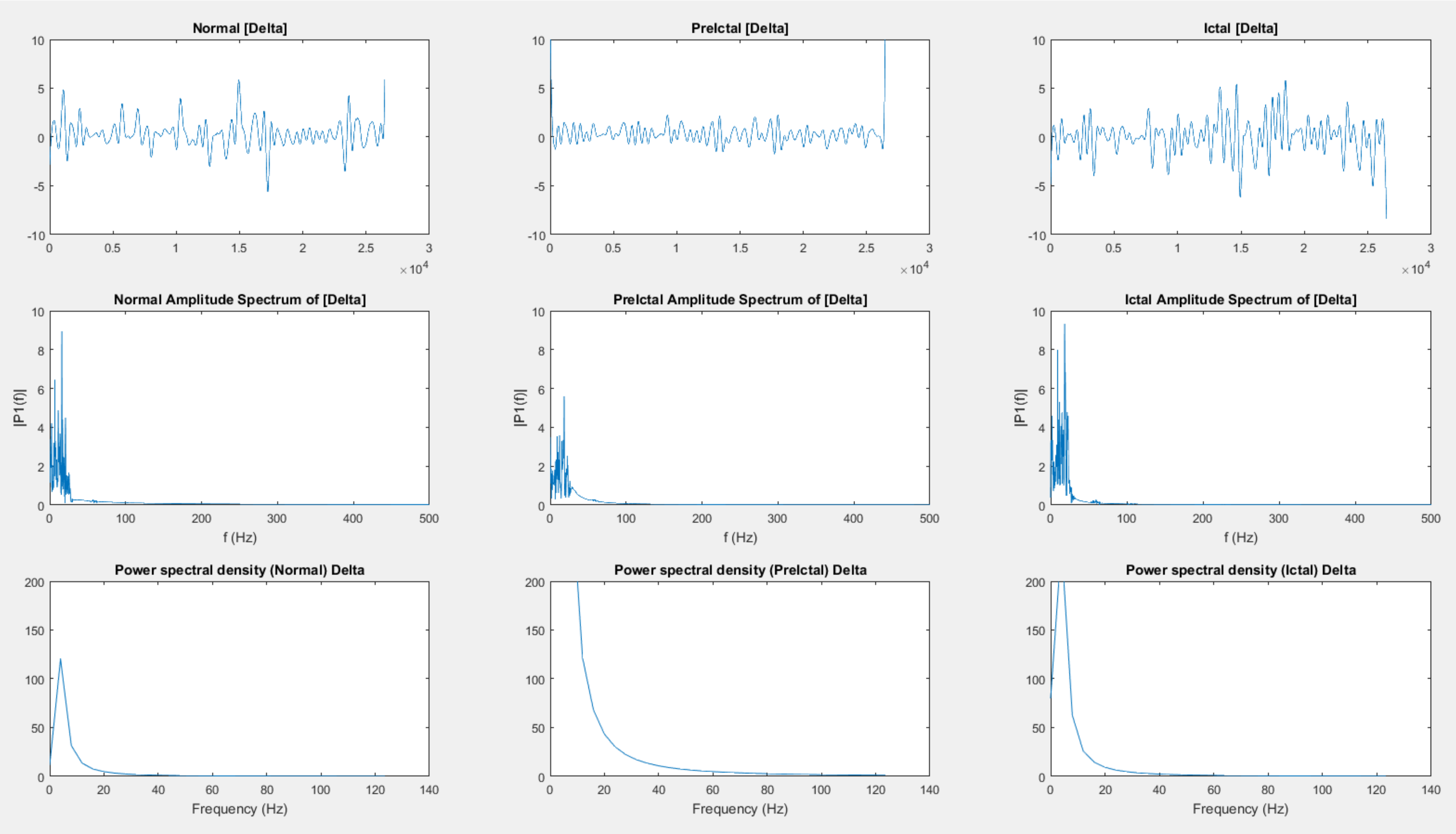
- Too Little: Daydreaming, inability to focus, too relaxed
- Too Much: Adrenaline, anxiety, high arousal, inability to relax, stress

Brain Waves - Theta waves



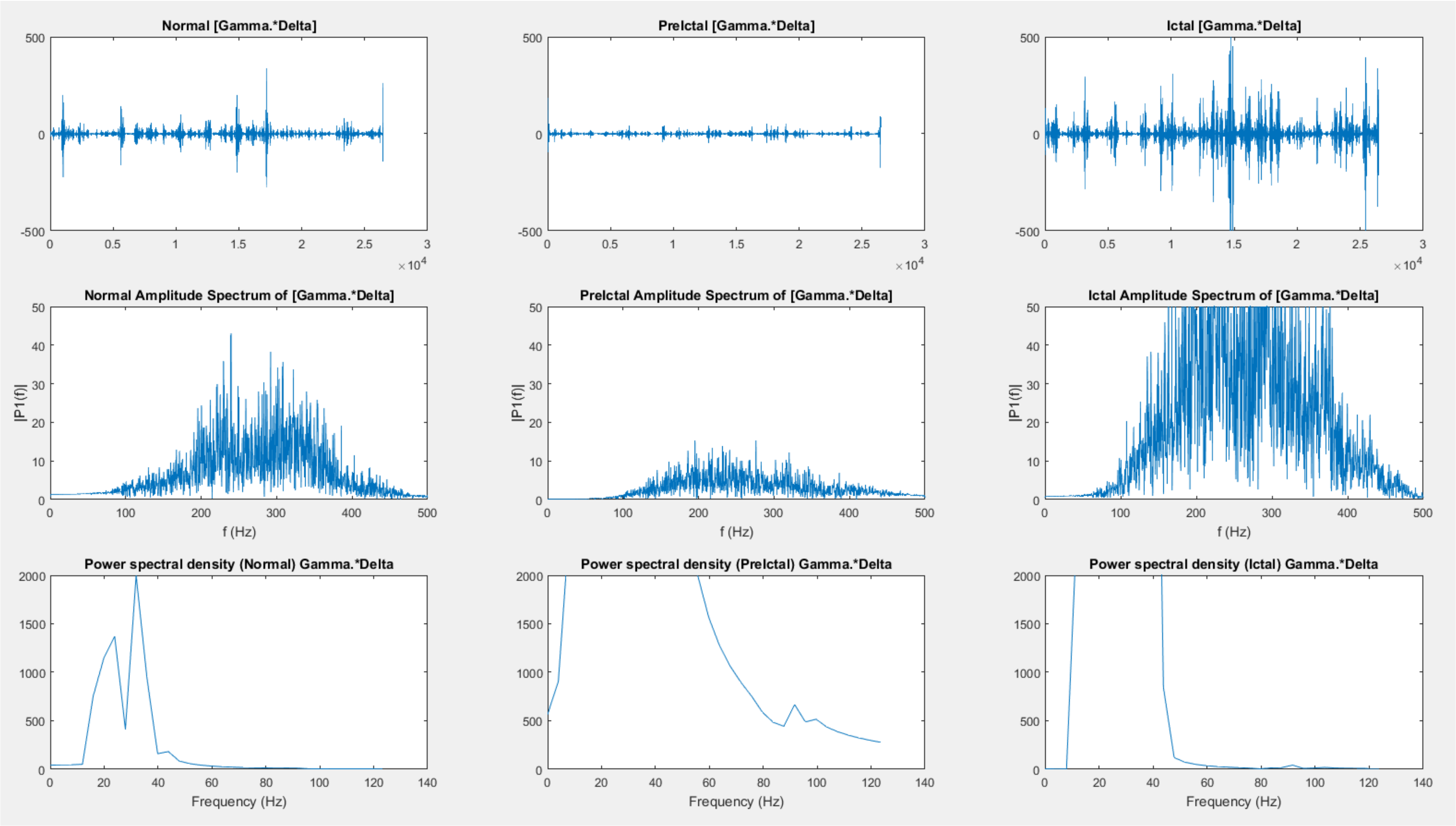
- Too Little: Anxiety, poor emotional awareness, stress
- Too Much: Depression, hyperactivity, impulsivity, inattentiveness

Brain Waves - Delta wave

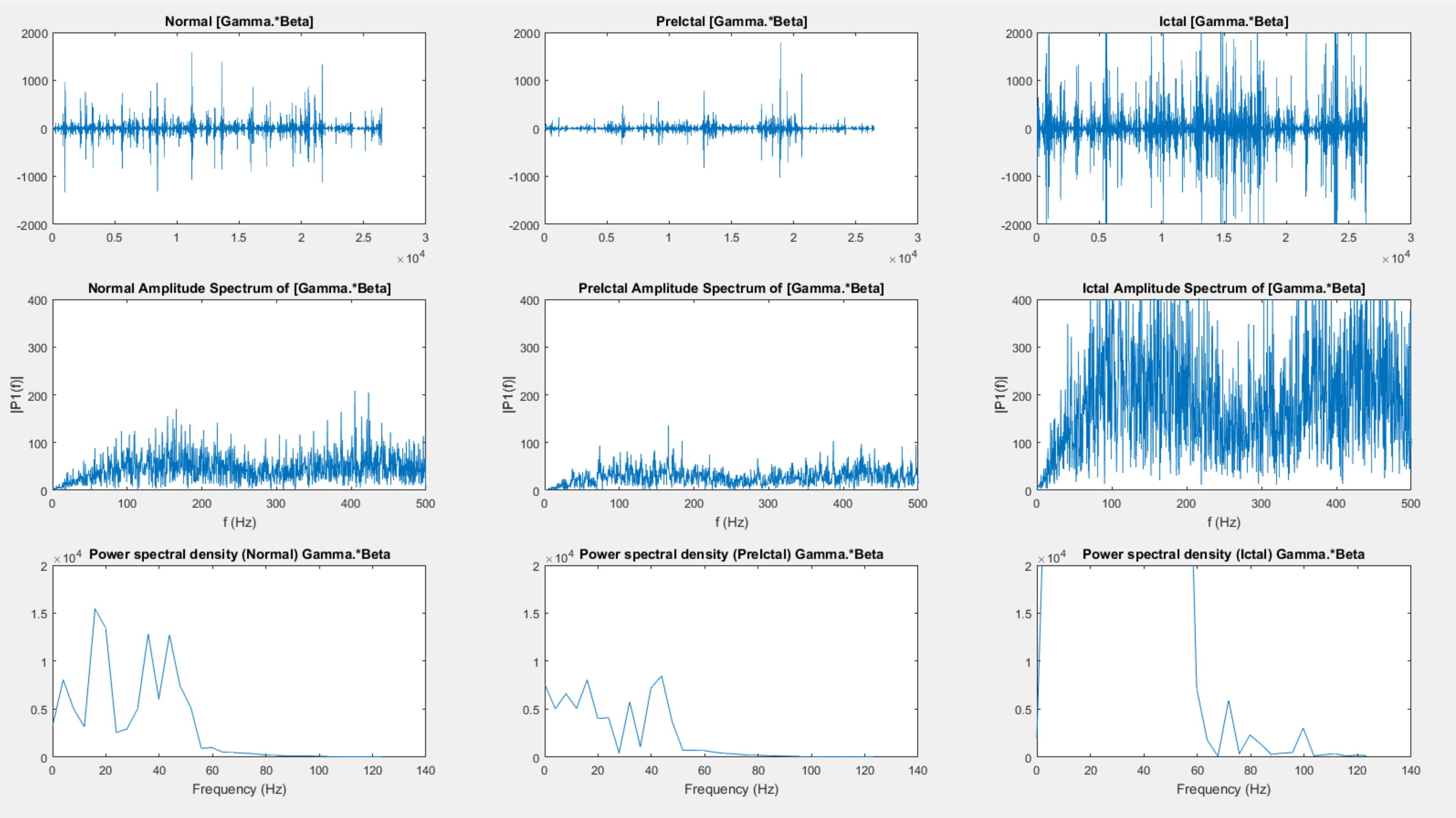


- Too Little: Inability to rejuvenate body, inability to revitalize the brain, poor sleep
- Too Much: Brain injuries, learning problems, Inability to think

Brain Waves – Gamma.*Delta



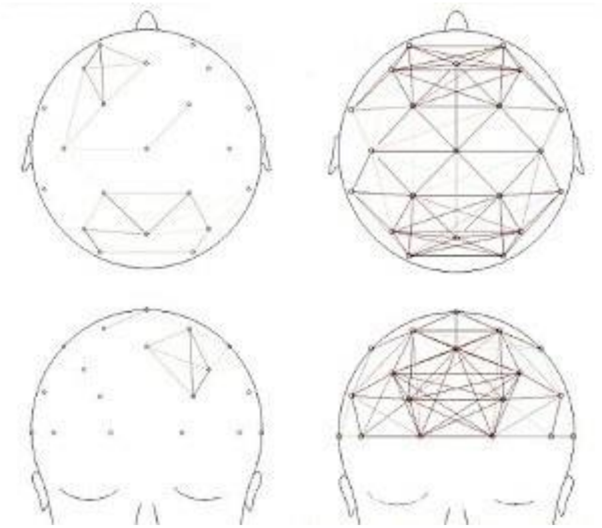
Brain Waves – Gamma.*Beta



Coherence

- Coherence based on spectrum analysis, which can describe the synchronization of electric brain activities of different frequencies between brain areas

- Extract transient characteristics of interactions among brain areas



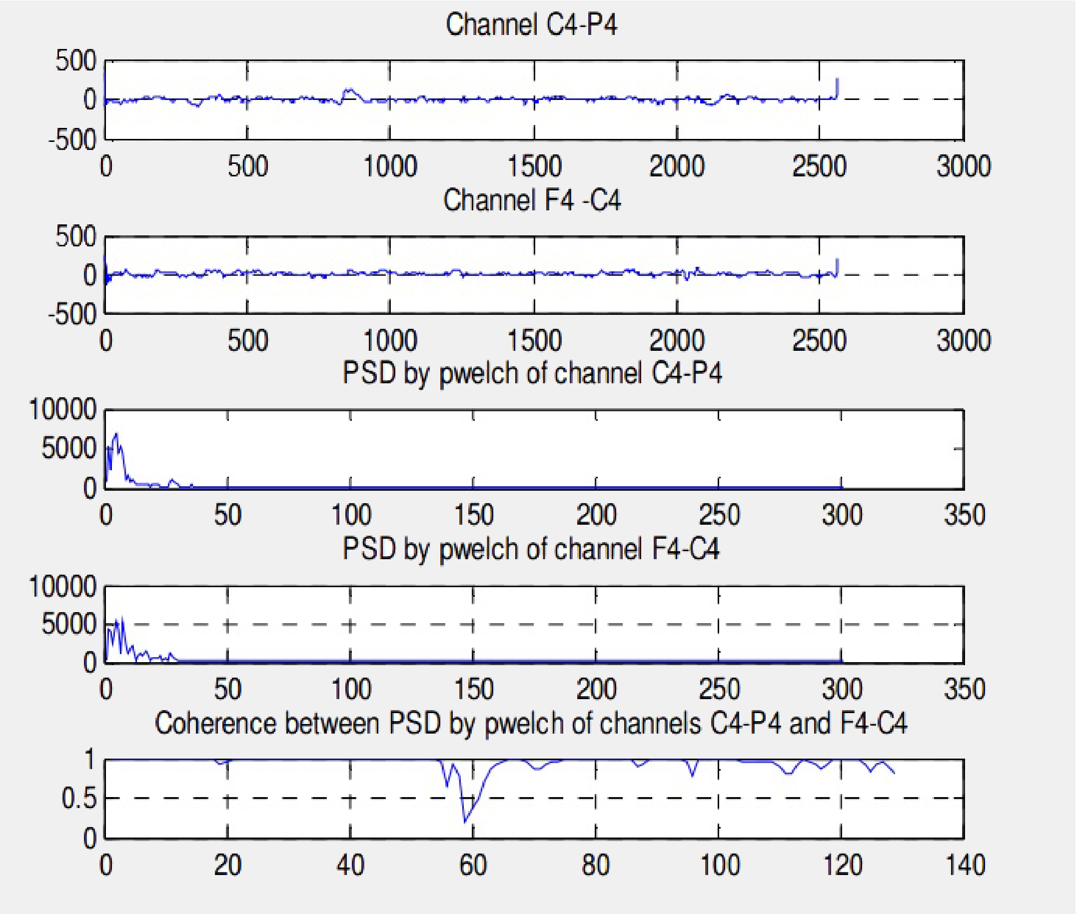
- Describes the temporal, spatial and frequency relationships of brain activities.

Coherence Estimation

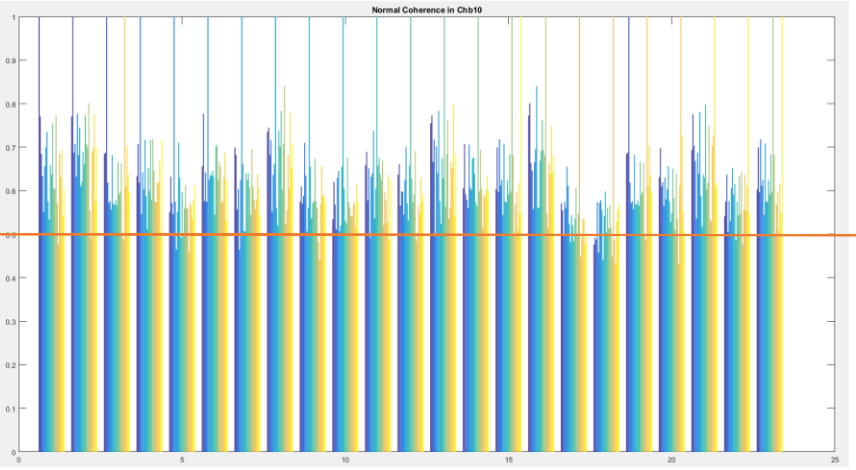
- In this analysis discuss the frequency-varying coherence of EEG to examine the coordination mechanism of the brain.
- The power spectral density (PSD) (Welch method) is the frequency-varying method to examine the coordination mechanism of brain areas.
- When two areas brain are more synchronous to each other the result is nearer to 1 that is, they are more coherent to each other, otherwise non-coherent means closer to 0.

$$\text{Coherence (f)} = \frac{\left| Cross - Spectrum(f)XY \right|^2}{(Autospectrum(f)(X))(Autospectrum(f)(Y))}$$

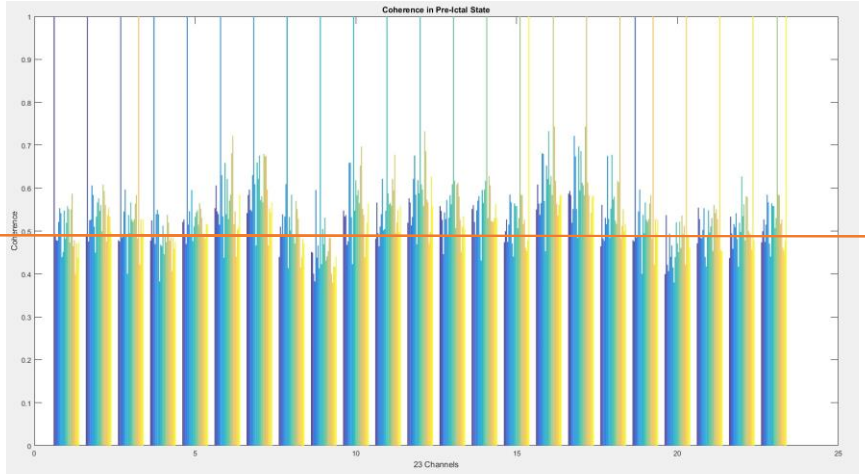
Coherence between C4-P4 and F4-C4



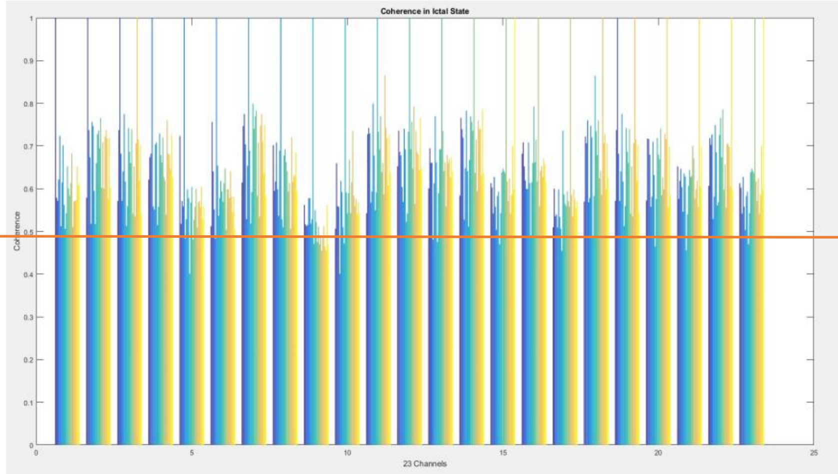
Coherence of all channels



Normal



Pre Ictal



Ictal

Hjorth Parameters

Hjorth parameters is a linear method which defines the temporal dynamics of a signal $X(t)$.

It indicates the statistical feature of a continuous EEG signal in time domain.

Hjorth Parameters

Activity: Estimates the variance of signals amplitude

$$Activity(y) = \sum_{i=1}^{N_s} \frac{(y(i) - a)^2}{N_s}$$

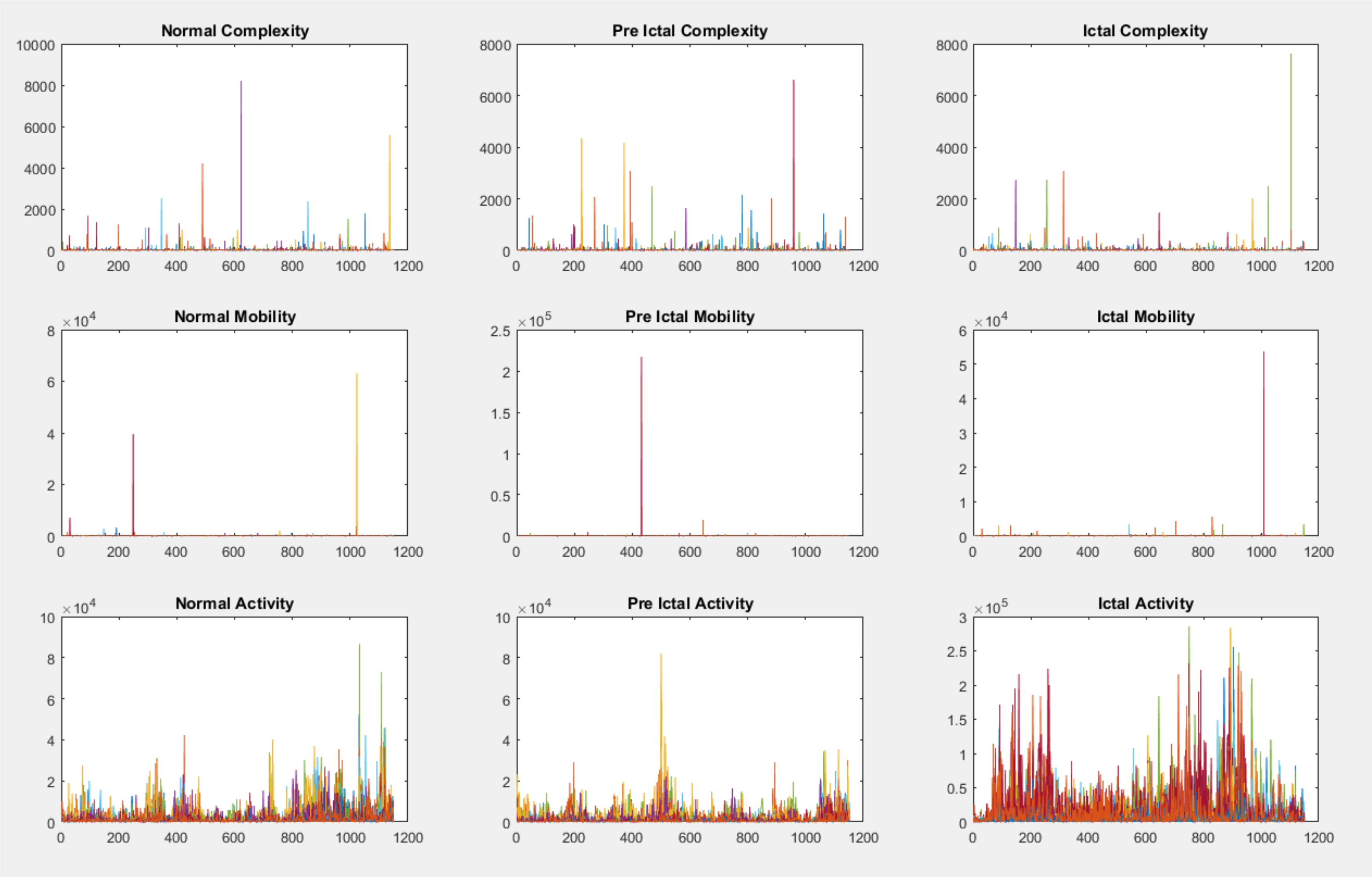
Mobility: Estimates the mean frequency

$$Mobility = \sqrt{\frac{var(y')}{var(y)}}$$

Complexity: Estimates of the bandwidth of the signal

$$Complexity = \frac{Mobility(y')}{Mobility(y)}$$

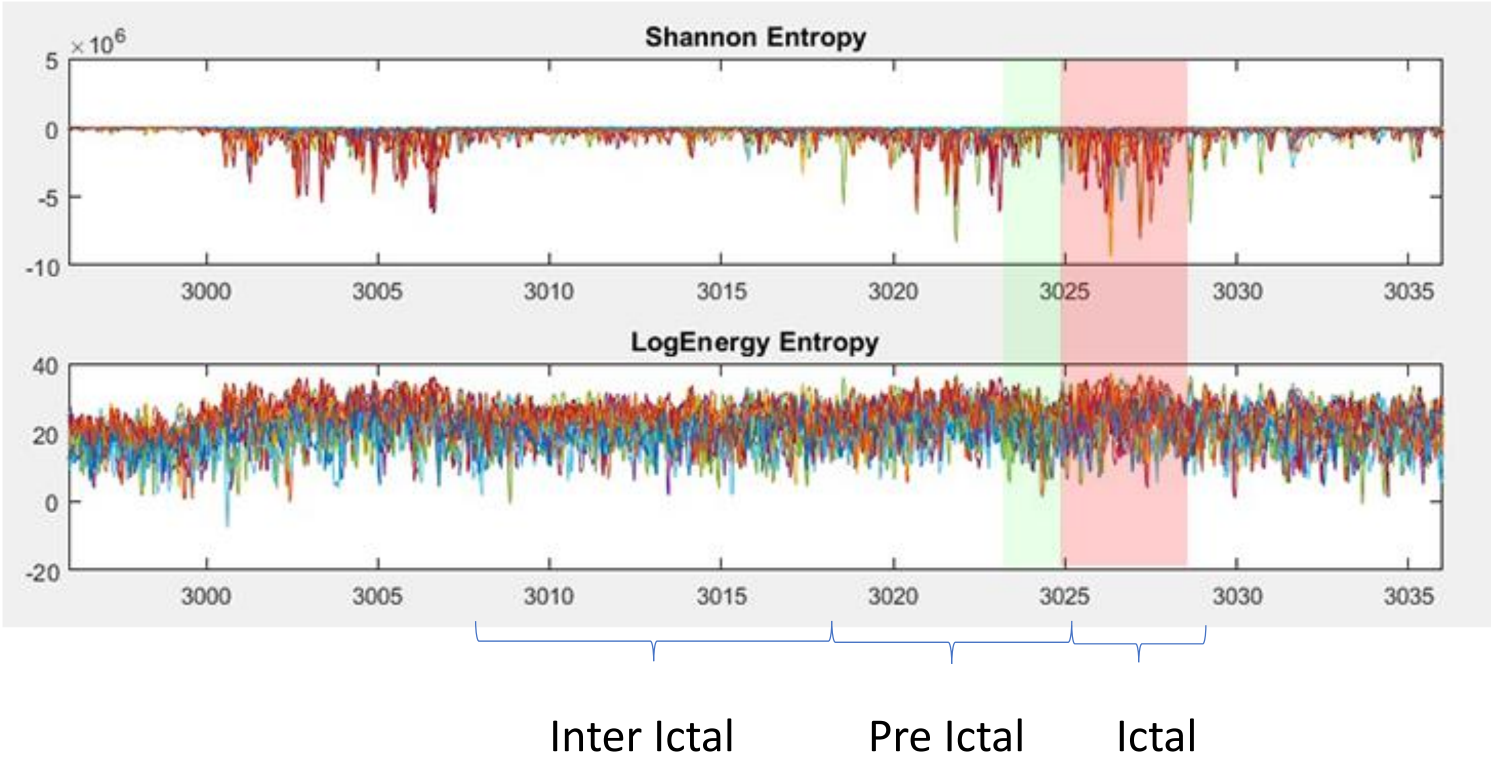
Hjorth Parameters



Entropies approach

A system exhibits high entropy values equal to high levels of disorder of a system, whereas low values describe a more ordered system.

- The information theory of time series was recommended as a metric of seizure detection and a recognize measure for stress or fatigue (prodromal symptoms).



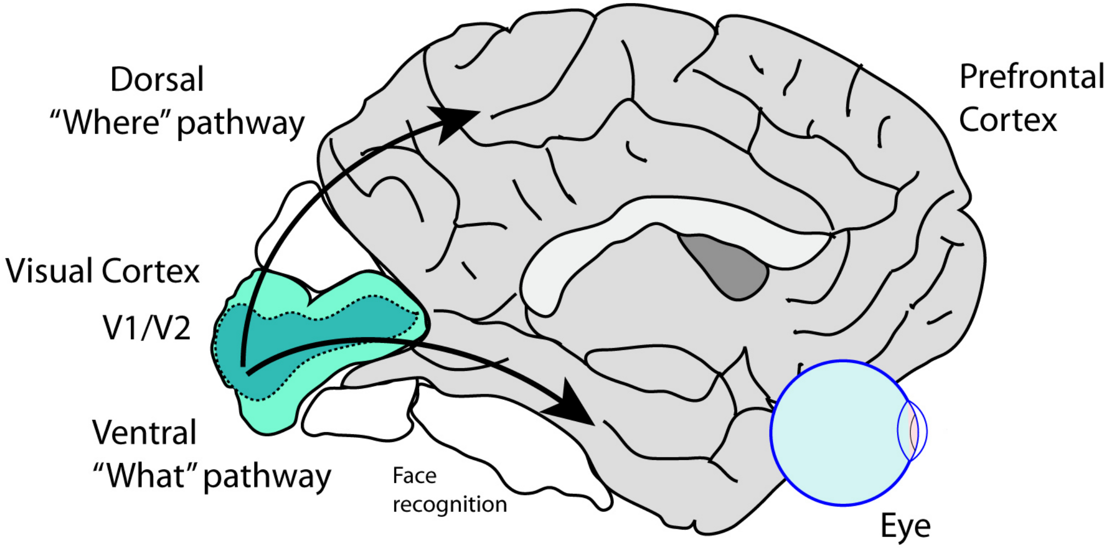
Limitations

- Different types of epilepsy disorders
- The age between youths and adults patients
- No ideal algorithm to work and adapt in more than one datasets
- Functional, biochemical and computational tools should be used to characterize therapeutic effects. With this knowledge, we may be able to develop closed-loop therapies that can restore normal function.

Extra Work

1. Definition of Sparsity

- Theoretical studies suggest that primary visual cortex (area V1) uses a sparse code to efficiently represent natural scenes.
- Sparse coding is a way of representing some phenomenon with as few variables as possible.
- Usage: *Compression, Analysis, Denoising*



2. Sparse Representation

$$x \approx D * a$$

$x \in \mathcal{R}^n$

is a signal

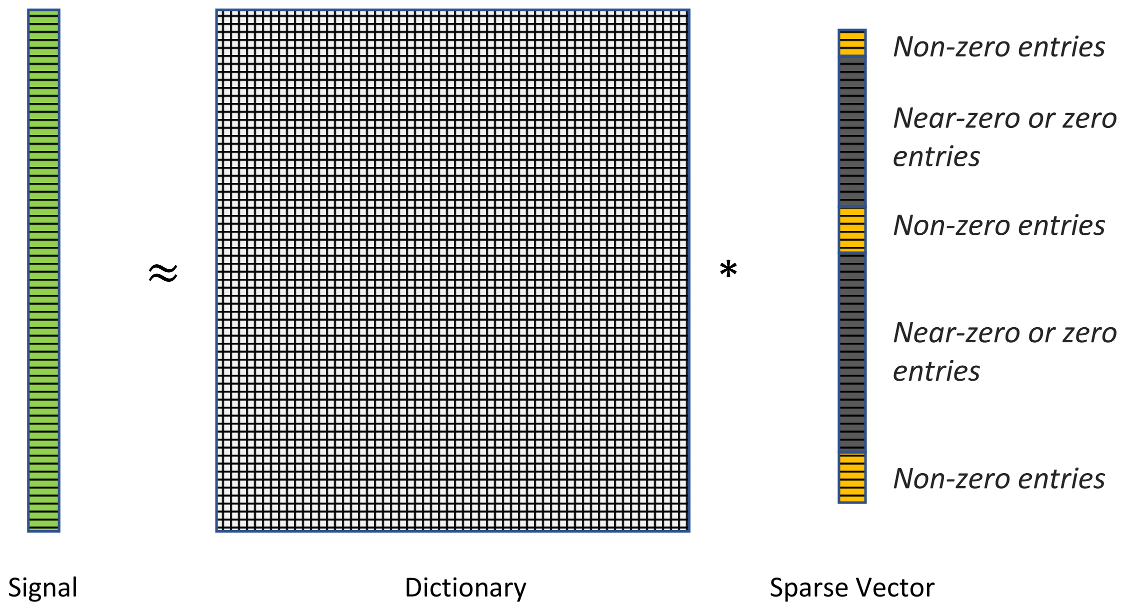
$D = [d1, d2..dn] \in \mathcal{R}^{nxm}$

$d^T d = 1$

is a set of normalized "basis vectors" or "atoms"

$a \in \mathcal{R}^m$

is a sparse vector that solves the equation

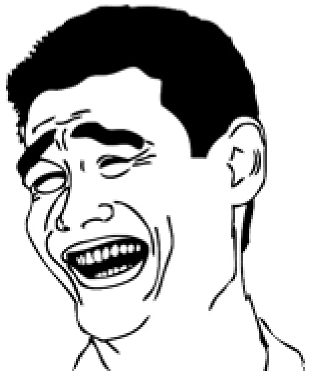


3. Why Sparsity?

- Provides the ability to represent large amount of data by keeping less information. (saves memory/storage)

- But the main property that sparsity has…

"If you can create a model such that the objects can be represented with much less information than originally, but still have high fidelity to the original, it means that you have created a good model, because you have replaced redundancies in the original representation."



4. Sparse Representation Overview

- When using sparse representation as a way of feature extraction, you may wonder:
  - Does there exists the sparsity property in the data?
  - Does sparse feature really come up with better results?
  - Does it contain any semantic meaning?

- In a short answer "sometimes, not always"

- **Below is a list of successful applications:**
  - Image Denoising
  - Deblurring
  - Inpainting
  - Super-Resolution
  - Restoration
  - Quality Assessment
  - Classification
  - Segmentation
  - Signal Processing
  - Object Tracking
  - Texture Classification
  - Image Retrieval
  - Bioinformatics
  - Biometrics
  - And Many Other Artificial Intelligence Sections

4. Sparse Representation Overview

- Assumes that there exists a sparse property in the data, otherwise sparse representation means nothing.
- **D** is used to be the feature of **x**
- **D** can be used to efficiently store and reconstruct **x**

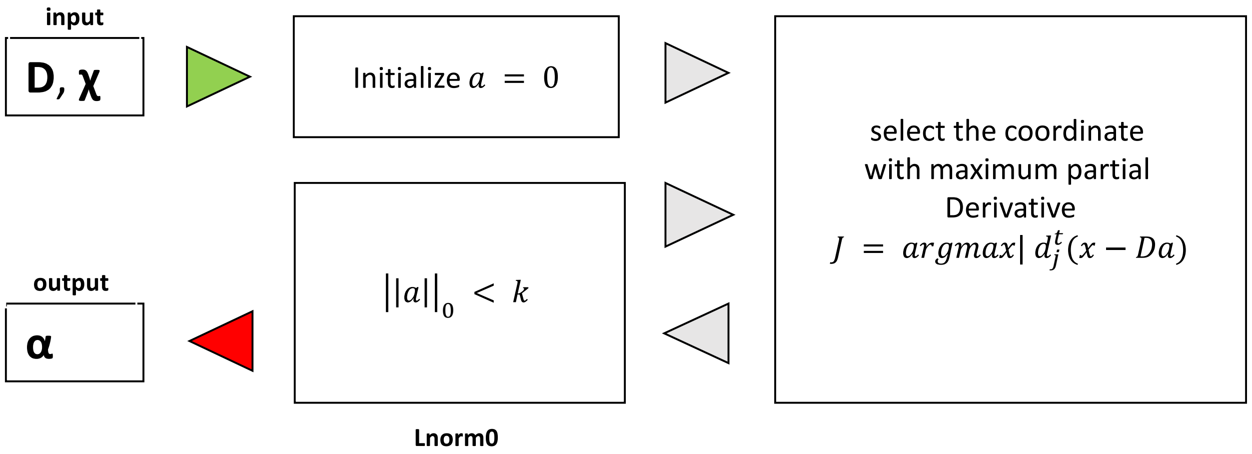
5. Methods of Solving Sparse Coding

The following algorithms are trying to find a sparse vector **α** of a given signal **x** and dictionary **D.**

- Greedy Optimization Strategy
  - **Matching Pursuit**
  - **Orthogonal Matching Pursuit**
  - **Orthogonal Matching Pursuit for large matrices**
- Constrained Optimization Strategy
  - **Gradient Projection Sparse Reconstruction**
  - **Interior-point method based sparse representation strategy**
  - **Alternating direction method (ADM) based sparse representation strategy**
- Proximity Algorithm Based Optimization Strategy
  - **Soft thresholding or shrinkage operator**
  - **Sparse reconstruction by separable approximation**
  - **Iterative shrinkage thresholding algorithm**
  - **l1/2-norm regularization based sparse representation**
  - **Fast Iterative shrinkage thresholding algorithm**
  - **Augmented Lagrange Multiplier based optimization strategy**
- Homotopy Algorithm Based Sparse Representation
  - **LASSO homotopy**
  - **BPDN homotopy**
  - **Iterative Reweighting l1-norm minimization via homotopy**

5.1. Matching Pursuit Algorithm

It is a coordinate descent algorithm that iteratively selects one entry of the current estimate α at a time and updates it. It is closely related to projection pursuit algorithms from the statistics literature.



6. Dictionary Learning

Now let’s look at the reverse problem: could we design dictionaries based on learning?

In the preceding slides, we generally assume that the (over-complete) bases **D** is existed and known. However in practice, we usually need to build it.

Our goal is to find the dictionary ***D*** that yields sparse representations for the training signals.

6. How can we get the dictionary D?

- Ideal case of Dictionary
  - Expect: $X - D * a = 0$
  - Until now **no algorithm is guaranteed** to find a global minimizer.

- Know categories of dictionaries:
  - Predefined dictionary based on various types of wavelets
  - Dictionary Learning

Future Work

- Try to learn a proper dictionary from the input signals and perform all the previous data preparation using the sparse representation of the signals and the dictionary.
- Why? This will increase the feature selection and minimize the computational requirements.

Questions