NEUROENGINEERING

PART B

Author:
Gioele Migno - migno.gioele@gmail.com

01 - THE ELECTROENCEPHALOGRAM - PART 1

TOPICS

- Overview
- EEG instrumentation

Reference: Hari and Puce. MEG-EEG Primer. Oxford University Press. 2017

OVERVIEW

Reference: Hari, Puce, MEG-EEG primer, Chapter 4

EARLY BRAIN ACTIVITY RECORDINGS

Richard Caton was the first to record electrical activity from the brain. He made invasive recordings directly from the cortex of rabbits, cats and monkeys (Caton, 1875). Noninvasive human EEG began to progress significantly in the 1920s and 1930s following the reports published by Hans Berger, a psychiatrist in Jena, Germany.

We can distinguish two different kinds of brain activity recordings:

- **EEG SPONTANEOUS ACTIVITY:** Study of spontaneous brain activity
- **EVENT RELATED POTENTIALS (ERP) ACTIVITY:** Study of how the brain responses to certain stimulus, using also in this case EEG

Early EEG recordings (spontaneous activity)

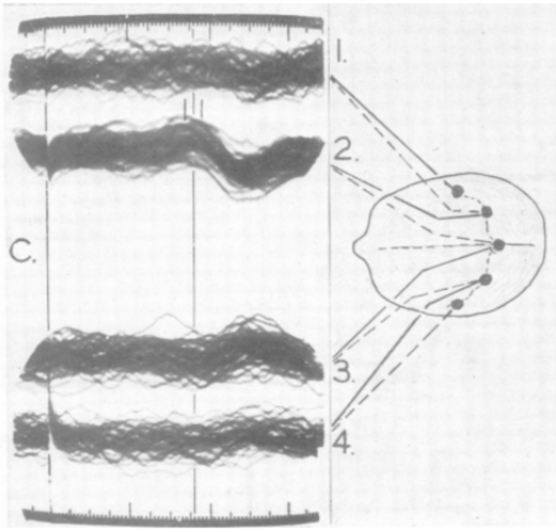
Hans Berger is considered the father of EEG, he was able to demonstrate a large and persistent 10-Hz rhythm that he
named “alpha.” Lord Adrian, a British neuroscientist became interested in EEG activity in the early 1930s, he observed that alpha oscillations appeared when the cortex had little to do, and he hence proposed, for the first time, an idling hypothesis for brain rhythms, namely the brain, when it is not working, it "kepts warm" himself, in this way it can quickly attain full capacity whenever needed.

The following image represents an example of a trace of 1 channel EEG recordings the alpha rhythm, its amplitude is about $20 - 50\mu V$:

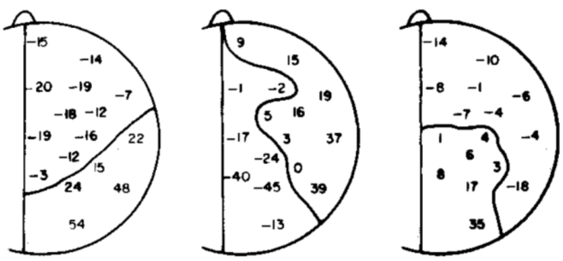
As we can see, alpha rhythm changes its amplitude when you open and close your eyes. As soon as a person open his eyes the rhythm decreases (disappears). This phenomena is called **alpha blocking**, so the alpha rhythm is blocked when open your eyes.

Early ERP recordings

A real leap in EEG applications occurred in the 1960s and 1970s when the first computers became routinely available and made it possible to average EEG signals time-locked to various sensory stimuli, thereby significantly increasing the signal-to-noise ratio. This leads to the study of event related potentials (ERP), since using a computer is possible to analyze the EEG removing all the signals not related to a specific stimulus.



Another important technical advance was the attempt to plot the EEG topographies as a function of time:



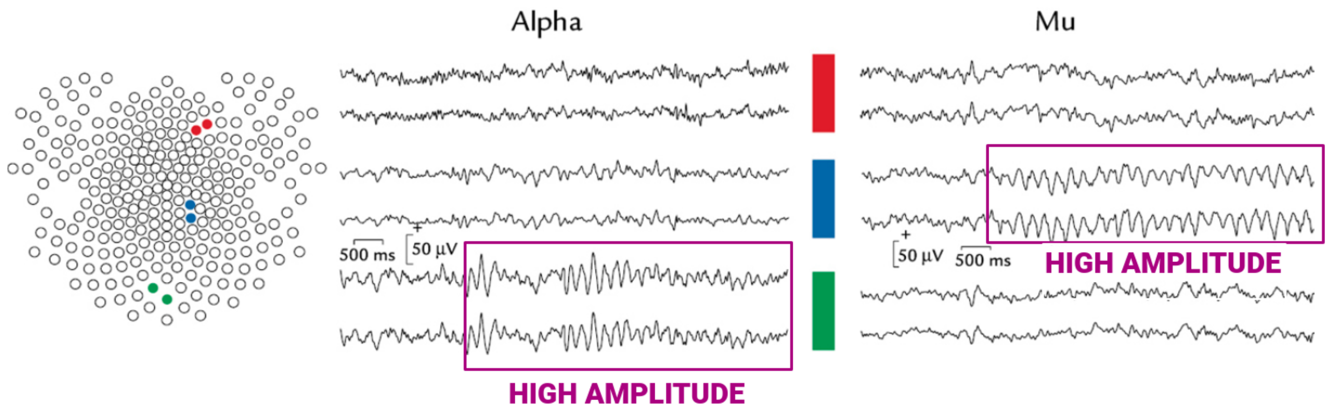
BRAIN RHYTMS

Lets go back to spontaneous activities, so we put electrodes on the head of a patient and record the EEG.

The two most known **brain rhythms** are:

- **Alpha rhythm:** It is strongest when we keep out eyes closed,. More visible at the back of the head
- **Mu rhythm:** It is strongest when we are totally relaxed and not making any movements. More visible in the Rolandic areas

In the following figure an high resolution EEG is shown. The EEG has been acquired using six electrodes placed in three different parts of the patient's head: front (red), central (blue) and back (green)



As we can see from the EEG, the alpha rhythmic is more visible in the back (green), the mu rhythmic instead, is more visible in the central area of the brain (blue). The variation of rhythmic amplitude is called *waxing and waning*.

Alpha rhythmic has a frequency of oscillation around $10Hz$ with a sinusoidal shape. Mu instead, has two frequency components, one around $10Hz$ and another around $20Hz$, this two components nature is shown by its arched shape.

Frequency bands

The alpha and mu rhythms just described, are the two first brain signals that have been recorded. Their names are related to a specific brain functions:

- **alpha rhythmic:** It generated by the visual cortical systems
- **mu rhythmic:** It generated by the motor cortical systems.

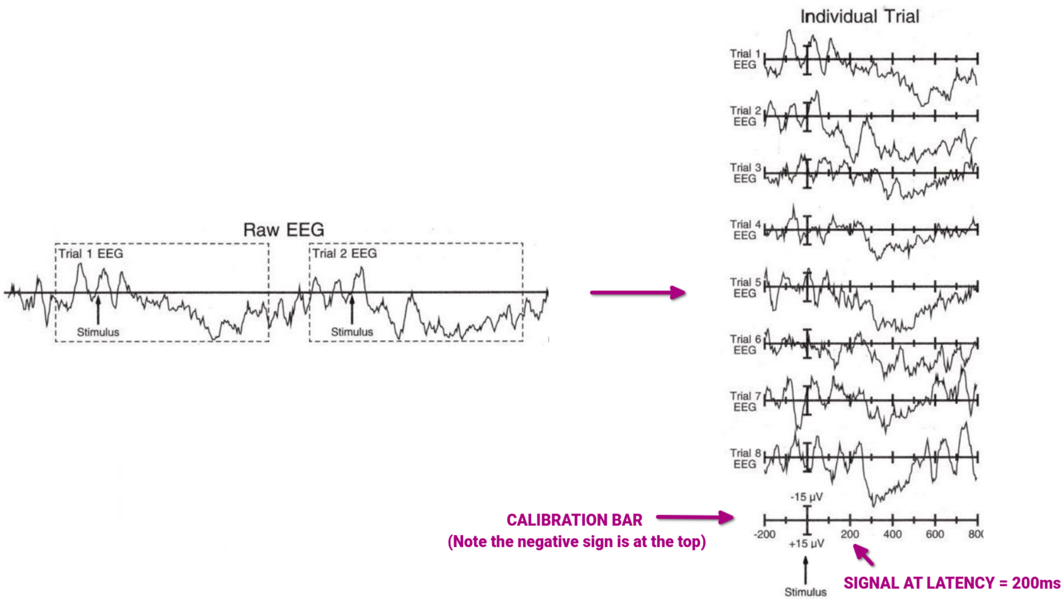
When EEG rhythms were increasingly used in clinical studies in the 1930s and 1940s, they were typically characterized in terms of their **dominant frequency** instead of their functions. The same frequency ranges are still in use today. These **frequency bands** are the following:

Name	Frequency Band
Delta δ	$< 3.5Hz$
Theta θ	$4 - 7.5Hz$
Alpha α	$8 - 13Hz$
Beta β	$14 - 30Hz$
Gamma γ	$> 30Hz$

> **Note** *(Function VS frequency band):*
> These names in the table refer only to a specific frequency band and NOT to the functionality of the rhythmic, indeed oscillations in the *alpha band* (i.e. between 8 and 13 Hz) will be called alpha, mu, or tau rhythms depending on whether they are generated by the visual, motor or auditory cortical systems.

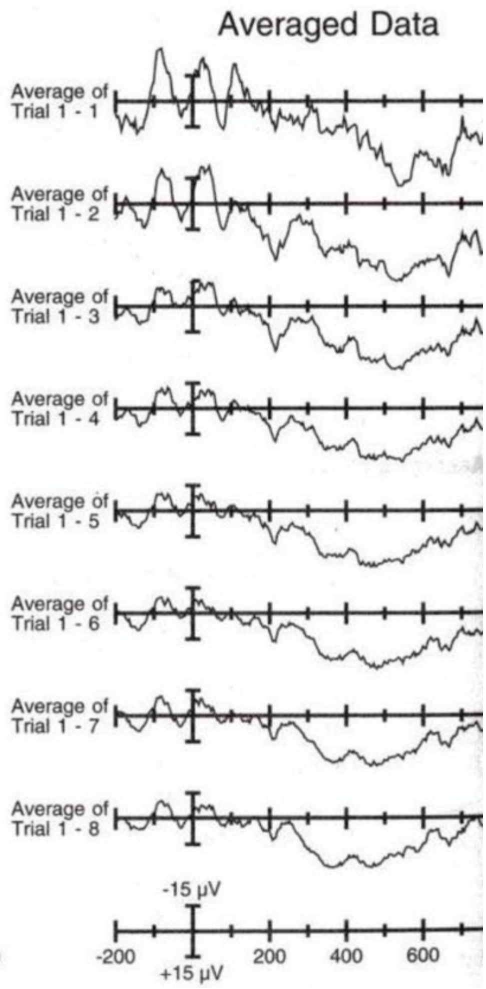
EVOKED AND EVENT-RELATED RESPONSES

Besides spontaneous brain activity, EEG can show time-locked (**evoked**) responses to various sensory stimuli. Spontaneous activities have a signal amplitude of about $10 - 100\mu V$, **event-related responses** instead, are in the order of $1\mu V$ or less. For that reason a single trial is usually not enough to record the response, therefore from a raw EEG all the trials are extracted and aligned respect to the stimulus:



The distance of a value from the stimulus is called **latency**, the latency after the stimulus can range from a few milliseconds to several hundreds of milliseconds.

Once aligned all the trails, for each latency we can compute the average value by considering an arbitrary number of trials:



Averaged Data
Average of Trial 1 - 1
Average of Trial 1 - 2
Average of Trial 1 - 3
Average of Trial 1 - 4
Average of Trial 1 - 5
Average of Trial 1 - 6
Average of Trial 1 - 7
Average of Trial 1 - 8
-15 μV
-200
200
400
600
+15 μV

TRUE STATEMENTS

- The alpha rhythm is an oscillatory component of the spontaneous EEG
- The frequency of oscillation of the alpha rhythm is around 10 Hz
- The amplitude of the EEG in the alpha band decreases when the generating region of the cerebral cortex becomes engaged in a functional task (visual, motor)
- The proper (visual) alpha rhythm is generated in the occipital lobe of the cerebral cortex.
- The mu rhythm is an oscillatory component of the spontaneous EEG, whose frequency is in the alpha band
- The mu rhythm is generated in the central regions of the cerebral cortex
- The oscillations of mu rhythm are more “arc-shaped”, rather than resembling a regular sinewave
- The amplitude of the alpha rhythm is not constant, but rather “waxes and wanes” with irregular periods, with changes occurring often after an interval in the order of 1 second.
- The alpha band of the EEG is conventionally limited between 8 and 13 Hz
- The delta and theta frequency bands identify frequencies lower than those in the alpha band
- The beta and gamma frequency bands identify frequencies higher than those in the alpha band
- The alpha rhythm can be observed by filtering the spontaneous EEG signal using a narrowband filter, with cutoff frequencies at 8 and 13 Hz (approximately)
- Evoked Potentials are deflection of the EEG signal following the presentation of a sensory input.
- Event related potentials include evoke potentials, as well as EEG responses to motor or cognitive events.

EEG INSTRUMENTATION

> **Reference**: Hari, Puce, MEG-EEG primer, Chapter 5

Let's see the main components in a **EEG instrumentation**

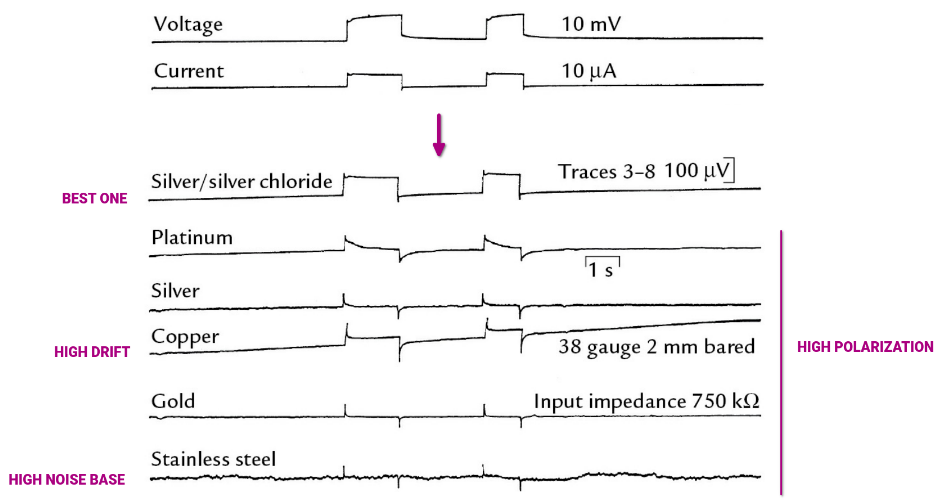
ELECTRODES

The first step during the acquisition of an EEG signal, is the conversion from an electrical signal used in a biological system and based on ions current, to a standard electrical signal that we use in electronics based on the flow of electrons. This conversion is performed by the **electrodes**.

The most commonly used EEG electrodes are made of silver and silver chloride (Ag/AgCl), the most important properties of AgCl are:

- **Non reactivity** with biological tissue
- **Accurate reproduction** of extremely slowly changing potentials
- **Low polarization potentials**
- **Low drift**
- **Low noise**

To understand what these properties mean, let's look at the following figure that compares several electrodes, based on different metals, respect to their capabilities to reproduce a signal:

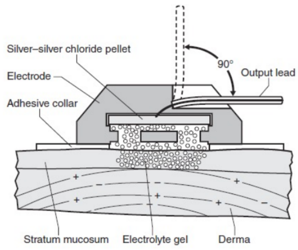


As we can see, silver is the best metal one, the other metals are not able to perform an accurate reproduction and they acts as capacitor (high-pass filter), especially gold (high polarization). Copper suffers also of a high drift effect, namely the mean value of the signal increases over time. The worst metal regarding to the noise, is the stainless steel. However, this metal is usually used to made intramuscular electrodes since it guarantees an high level of hygiene due to its resistance in an autoclave (sanitizer).

> **Note:** *Different electrodes*
> During an EEG it is important that all the electrodes used are made of the **same material**

Skin interface

The human skin is very dry, so it does not allow a good conduction of electricity (high impedance). To reduce its impedance, a gel based on silver is used, it avoids a direct contact between the electrodes and the skin and most important, is able to penetrate in the skin allowing a better conduction.



Impedance is a measure of the quality of the contact between the scalp (head), electrode, and conducting medium, this quantity must be measured with alternating current ($10 - 20Hz$), otherwise high polarizable metals like gold would have an infinity impedance ($R = \infty$).

EEG AMPLIFIERS

Once acquired the electrical signal from the brain through the electrodes, it is necessary to amplify the signal reducing as much as possible any kinds of noise. Moreover, since we want to amplify a voltage signal, the amplifier must have an high input impedance (ideally infinite).

So, the requirements for a good EEG amplifier are:

- **Low noise**
- **High gain**
- **High input impedance** (hundreds of $M\Omega$)
- **High CMRR** (~100 dB)

These properties can be guaranteed by using a **differential amplifier** that amplifies the difference between two signals: one of **interest** and one other used as **reference**.

Let's delve now into these characteristics.

Common-Mode Rejection Ratio (CMRR)

A measure of interest is the amplifier's **common-mode rejection ratio** (CMRR), which is the ratio between the differential gain (how much the difference between the inputs is amplified) and the common-mode gain (how much the common signal in both inputs is amplified).

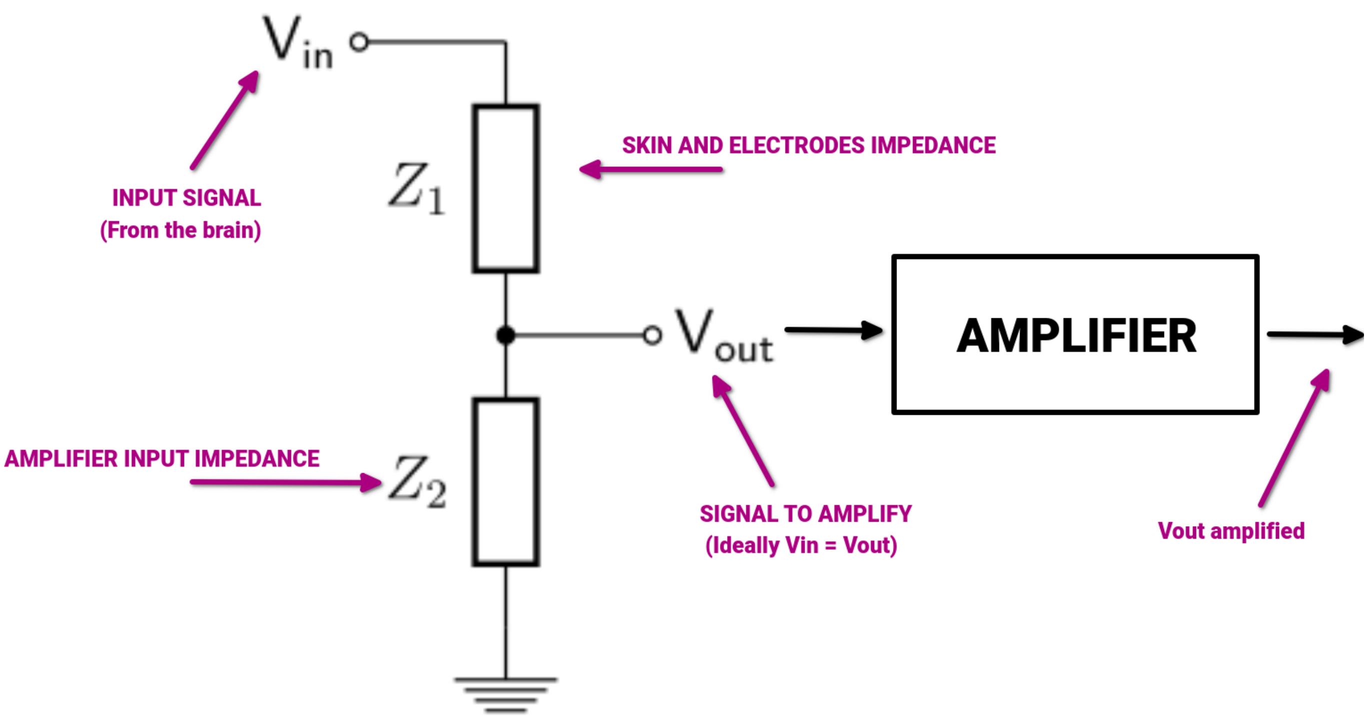
CMRR is expressed in decibels as follows:

$$CMRR = 20\log_{10}\left(\frac{\text{gain for difference signal}}{\text{gain for common-mode signal}}\right)$$

Examples of **unwanted common-mode signals** could be power-line noise or noise arising from a cardiac or nerve stimulator in another part of the body. A good amplifier should reject, or strongly suppress, all common-mode signals and amplify only the signals of interest.

Input Impedance

An important feature of the amplifier is its **input impedance**, which should be many orders of magnitude higher than the electrode impedances so that the amplifier remains as insensitive as possible to small changes in electrode impedances. To understand why, let's look at the following figure:



The first impedance Z_1 represents the summed impedance of the skin and the electrodes, as said before, this impedance must be very small. Z_2 instead, represents the input impedance of the amplifier. Our goal is to have $V_{in} = V_{out}$, in this way the biological potential is not dampened (voltage divider) and we can amplify the "real" signal from the brain V_{in}

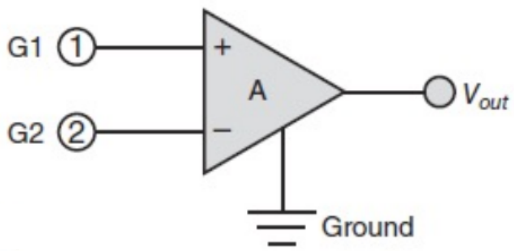
EEG guidelines suggest to keep Z_1 impedance below $5k\Omega$, however modern amplifiers with input impedances Z_2//(hundreds of $M\Omega$) allow to relax this requirement.

> **Note:** *(Impedance symmetry)*
> Impedances should be comparable across electrodes, because CMRR is reduced by any asymmetry in the measuring circuit

Although modern amplifiers have an high input impedance, it is important to use electrodes with a low impedance, in this way the possible difference of impedance between electrodes is kept small.

Differential Amplifiers

All the previous characteristics are satisfied by a **differential amplifier**. This device has two input pins: a **non-inverting** $(+)$ and an **inverting** one $(-)$. In output we have the difference of the input signals amplified.

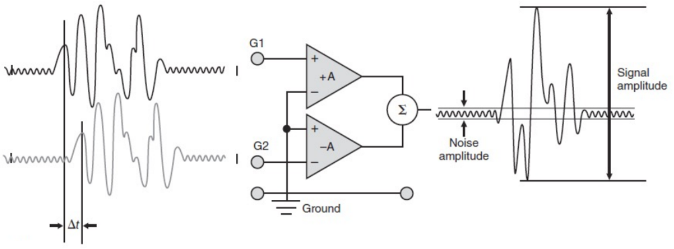


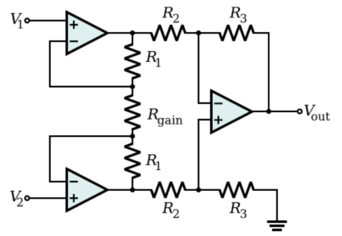
Let's try to clarify where the input pins must be placed:

- **Non-inverting pin** $(+)$ (G1): It used as **active** pin, namely is connected to the area of the skull to measure
- **Inverting pin** $(-)$ (G2): It used as **reference** pin, it is usually placed on the skull or nearby
- **Ground pin** (GND): Overall ground of the circuit

> **NOTE:** *(Ground electrode)*
> The ground electrode, also fixed to the subject, is required to feed the two halves of the differential amplifier. The ground electrode location can be freely chosen, as far as it does not introduce a high common mode signal.

The differential amplifier is usually used in a configuration called **instrumental amplifier**. This configuration makes the amplifier particularly suitable for use in measurement and test equipment. The most important characteristics include very low DC offset, low drift, low noise, very high open-loop gain, very high common-mode rejection ratio, and very high input impedances. Instrumentation amplifiers are used where great accuracy and stability of the circuit both short- and long-term are required.





(Standard instrumental amplifier)

STANDARD ELECTRODE POSITIONS

When we perform an EEG is important to use a **standard electrode positions**. This positions can be used to place both the **active** electrodes G1s and the **reference** ones G2s.

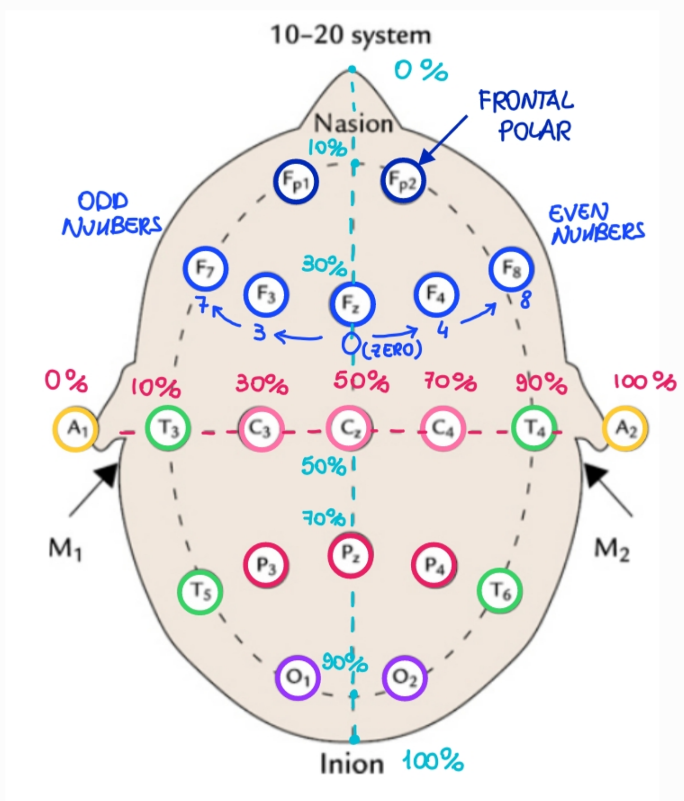
There exist two different international electrode placement and labeling: 10-20 system and 10-10 system. Let's see both of them.

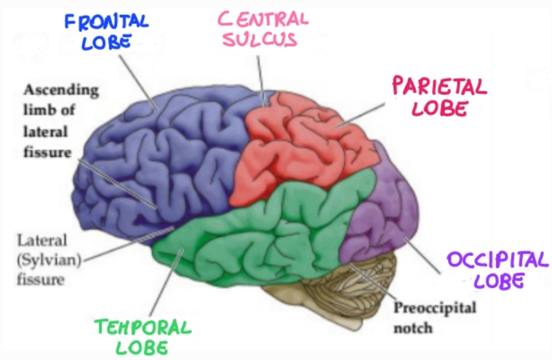
10-20 system

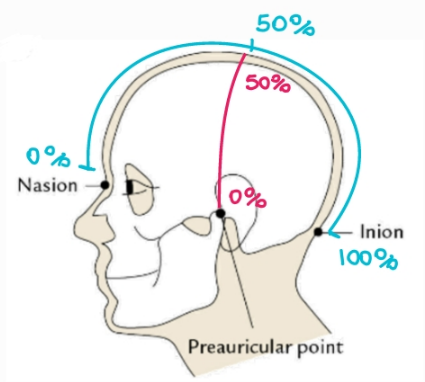
First at all, the head of the subject must be measured. We take two different measurements:

- **Nasion-Inion:** "Vertical line" from the nose to the back of the head
- **Preauricular:** "Horizontal line" between the ears

Then each line is divided in percentage, at the begin and the end we increase of 10%, in the middle points of 20%, namely: 0%, 10%, 30%, 50%, 70%, 90%, 100%. This is the reason behind the name **10-20 system**.





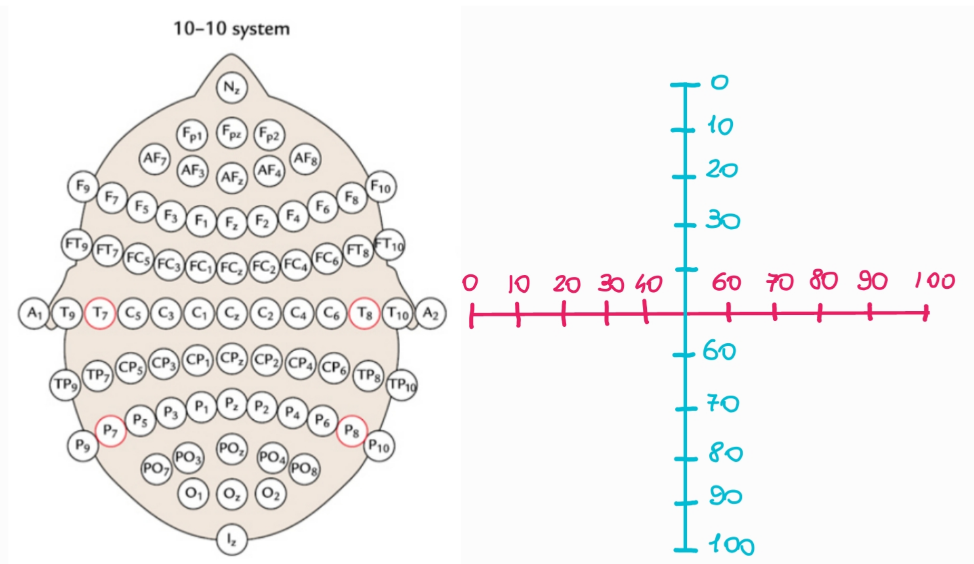


> **OBS:** (*index of labels*)
> The name of each position depend by its relative brain area (lobe) and by an index. We can note that at the left part we have only odd indexes and in the right part instead, only the even indexes. Moreover, at the middle we use *Z* as index to indicate the zero.

> **OBS:** (*number of electrodes needed*)
> We have already said that for a single measurement, three signal are necessary: $active$, $reference$ and $ground$ signal. So to acquire a **1-channel EEG** are necessary at least three electrodes.
> A **2-channel EEG** instead required at least four electrodes: $active_1$, $active_2$, $reference$, $ground$.
> We said "at least" since we can use different reference signals as we will later on.

10-10 system

The second standard is called **10-10 system** is similar to the previous but increases the total number of electrodes.



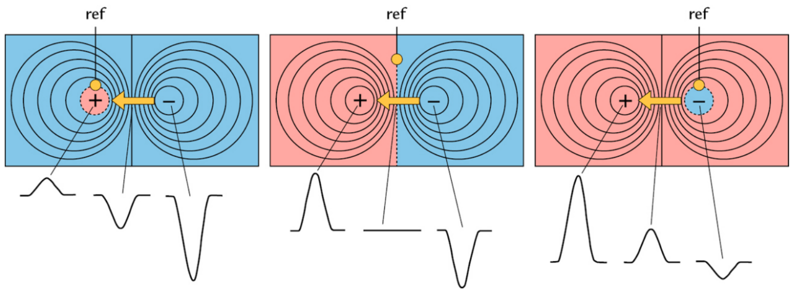
EFFECT REFERENCE ELECTRODE ON POTENTIAL DISTRIBUTION

In this part we are interesting into understand how the choose of the $reference$ signal influences an EEG.

Ideally, the reference electrode should be totally neutral and not contribute neural activity to the measurement. Unfortunately, there is not point in the human body with a zero electrically activity, therefore it is not possible to have an inactive reference ($REF \neq 0$). In practice this is not a huge issue, the important thing is to known the potential REF used at reference. Indeed we recall that this value must be taken into account since an EEG is acquired by measuring the voltage differences (potential) between two electrodes G1, G2 (in addition to the ground).

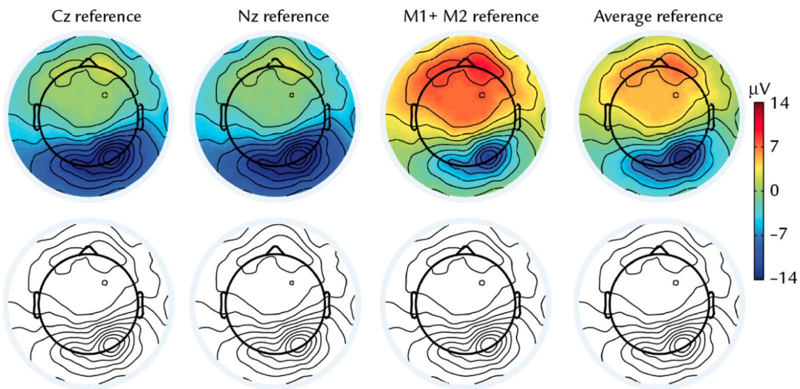
Reference electrodes are typically sited in places that are assumed to be far from the putative activity of interest, and traditionally popular places for reference electrodes have been earlobes, mastoids, and the nose. Note, these electrodes are not inactive.

To understand how the choose of $reference$ signal affect the measurement, consider a water basin where we put the two terminals $(+)$ $(-)$ of a $12v$ battery. Now using a voltmeter we measure the potential at each point inside the water using as reference a point ref inside the basin. Ideally, there is a point between the two terminals in which the potential is exactly zero, this point is called **point at infinity**. Using it as reference ref, we obtain the measurement in the middle of the figure in which we measure a positive potential near the positive terminal and a negative potential near the negative terminal of the battery. In practice scenarios we are not able to place ref at the point at infinity, therefore our actual measurement will be similar to the one on the right or the other on the right.



> **NOTE:** *(Fixed potential destruction)*
> It is important to note when we change the reference, we change only an offset therefore the potential distribution still the same but each value is shifted.

The following figure shows how an EEG changes with respect to the reference used. The data were collected with a 256-channel geodesic net with respect to the vertex, Cz (top left), and then digitally re-referenced with respect to nasion, Nz (top right), linked mastoids, M1 + M2 (bottom left), and an average (bottom right) reference.



The best approximation of the *reference at infinity* is obtained in the last case when the average between all electrodes is used as digital reference. This technique is called **Common Average Reference (CAR)**.

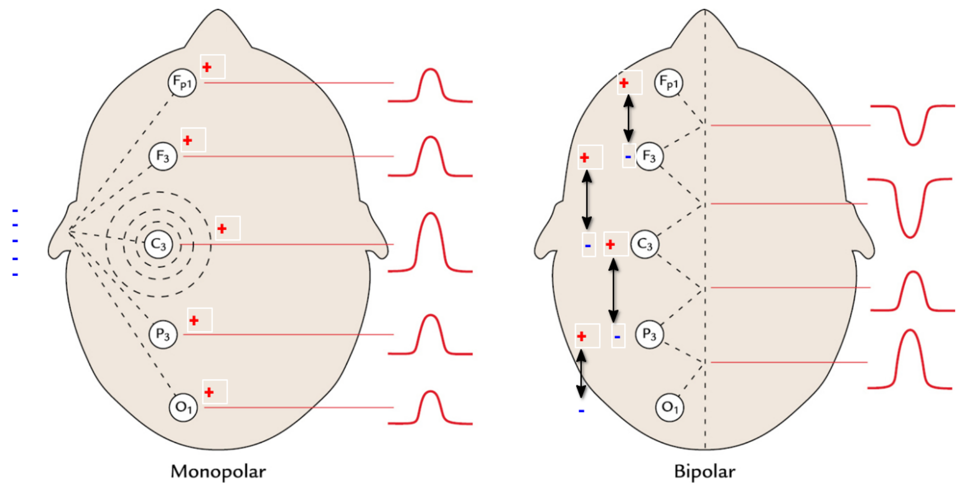
> **OBS:** *(Same potential distribution)*
> The figures at the bottom are always the same and are used to show that the potential distribution remains constant when we change the reference.

> **NOTE:** *(mastoids as reference)*
> In general cases, when M1, M2 pins are used as reference, each of them must be acquired singularly and NOT short circuited.

MONOPOLAR AND BIPOLAR RECORDING

So far we have simplicity referred to **monopolar recording** namely measure the potential at each electrode with respect to a single electrode used as reference. In a clinical setting instead, it is usually used another type of measurement called **bipolar recordings** in which potentials are measured between neighboring electrodes in different standardized pairs of configurations.

The following figure shows the difference between the two types of recording during the acquisition of the same potential maps. In the left it is used a monopolar derivation with left ear as reference, in the right instead, a *longitudinal* bipolar derivation is used. In the figure is also shown where non-inverting $(+)$ and inverting $(-)$ pins of the amplifiers are connected.



The most interest difference is that in the monopolar derivation we have an amplitude variation. In the bipolar one instead, we can see clearly the change of polarity between the front and back part of the scalp.

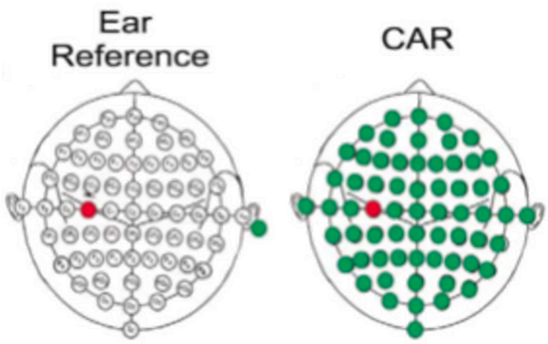
Re-Referencing Relative to an Average Reference

// to do

• The integral of the potential over the surface of a sphere that contains only concentric inhomogeneities is zero,

• The summed potentials of evenly spaced electrodes across the entire surface would be null as well

• In practice these conditions are never exactly met



TRUE STATEMENTS

- EEG electrodes made of silver with a layer of silver chloride (Ag/AgCl) are non-polarizable, thus allow recording of extremely slow-changing potentials
- In the EEG terminology, impedance is a measure of the quality of the contact between electrode and scalp, through the conductive gel
- Contact impedance of the electrodes is measured in kiloOhm ($k\Omega$) and must be measured using an alternating current.
- Gold electrodes are polarizable, thus they should not be used to measure slowly changing potentials.
- The Common-Mode Rejection Ratio (CMRR) of a bipolar amplifier characterizes the ratio between the gain of the difference of potential between input electrodes and the gain of the common potential with respect to ground.
- The CMRR is usually expressed in decibel (dB) and high values characterizes better amplifiers.
- The advantage of a high CMRR amplifier is that it suppresses common-mode disturbances such as powerline (50 Hz) noise.
- The measurement of a single EEG signal requires three electrodes – two as input to the differential amplifier and one to provide the ground potential.
- The input impedance of a biosignal amplifier must be many orders of magnitude higher than the contact impedance of the electrodes. It is usual to have input impedances in the order of 108Ω.
- EEG guidelines suggest keeping the electrodes' contact impedance below 5 $k\Omega$. The use of modern amplifiers with high input impedance allow to relax this requirement.
- One of the reasons why one should keep the electrodes' contact impedance much lower than the amplifier's input impedance is that the resulting a voltage divider would otherwise reduce the measured potential.
- The difference of contact impedances of electrodes should be small compared to the input difference of the differential amplifier, otherwise the resulting unbalance compromises its common-mode rejection capability.
- The electrode labels of the International 10-20 System uses the first letter of the four lobes of the cerebral cortex (Frontal, Parietal, Occipital, Temporal), plus "C" (central) to designate electrodes over the central sulcus.
- In the electrode labels of the International 10-20 System, odd/even number designate electrodes on the left/right side, respectively
- The International 10-20 System for EEG electrodes placement takes its name because the distance between adjacent electrodes is 10% or 20% of the distance between pairs of reference points on the skull (Nasion and Inion, or preauricular points)
- In monopolar EEG recordings, a single reference potential is used for all recorded channels. The reference electrode is placed on scalp position that are assumed to be far from the electrical sources of interest, such as the earlobes.
- The position of the reference electrode can strongly influence the shape and amplitude of EEG potentials. The profile (i.e. disregarding the actual potential value) of scalp topographies are not influenced.

- In bipolar EEG recordings, each channel is the difference of potential between two adjacent electrodes. It is mostly used in clinical settings to highlight features of interest at visual inspection of the waveforms.
- EEG signals recorded in monopolar configuration can be re-referenced to the Common Average Reference (CAR), by subtracting from each channel the instantaneous average of all channels. In ideal conditions, this would approximate taking the reference potential at infinity.

02 - THE ELECTROENCEPHALOGRAM - PART 2

> **TOPICS**
>
> > - Practicalities of data collection
> > - Artifacts
> > - EEG analysis

> **Reference**: Hari and Puce. MEG-EEG Primer. Oxford University Press. 2017

DATA COLLECTION

In this section we will see some good practices to follow during an EEG recording session.

GENERAL PRINCIPLES OF GOOD EXPERIMENTATION

During the acquisition of an EEG the signal measured are very small, therefore it is important to record EEG data from alert and cooperative subjects in as artifact-free and noise-free conditions as possible. Artifacts are signals no related to the EEG that we want to reduce as much as possible, we will see later that to reduce them, the collaboration of the subject is crucial.
A good practice to follow in order to maintain active the subject and avoid to throw away all the data due to an error, it is to divide the recording session to smaller *runs*, or blocks (e.g. 5/10 minutes). With the term *run* we refer to a continuous flow of uninterrupted signals.
During the session is important to monitor the brain signals continually online for any artifacts or technical issues and to remind subjects about relaxing their muscles, minimizing blinking, and avoiding head and body movements.

ELECTRODES PREPARATION

The first thing to do is to ask subjects to wash their hair (and skin) prior to arriving for an EEG study and to refrain from applying any cosmetic products on the skin or hair/scalp. This reduces skin impedance since dirty and pieces of dead skin are removed during the washing process, however head must be well dry.

Skin preparation and electrode application

Most electrodes have a hole at the top where conductive gel can be inserted with a blunted needle/syringe, the blunted needle also allows light abrasion of the skin of the scalp to be performed, if needed. Following skin preparation, impedances of less than 5-10 $k\Omega$ can be easily obtained.

Electrode-impedance measurement

Another important step is to measure impedance of electrodes, they must be low as possible and more important very similar between them (symmetry). As said previously, electrodes impedance must be measured using alternating current, to do that professional EEG equipments contains a specific circuit and provides an easy to use GUI.

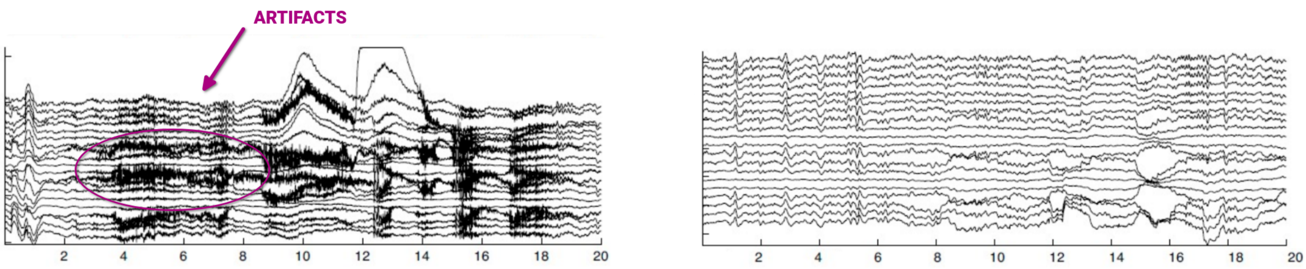
PLANNING AND LOGGING

Scientific experiments have strict protocols, even if conducting an exploratory session, you still need to:

- Plan your recording
- Accurately log the conditions occurring in each run (e.g. issues with some electrodes)
- Possibly save explanatory metadata (e.g. time markers) with the data

ARTIFACTS

With the term **artifacts** we refer to external signals that produces electrical activities not related to the brain. This
activities are recorded together the data and makes it difficult to read EEG data. For that reason in the first place we have to avoid the generation of artifacts and then remove them from our EEG recording.



EEG \W ARTIFACTS
(Real recording)

EEG W\O ARTIFACTS
(Desired recording)

More formally we can define **artifact** in EEG as any potential difference due to an extracerebral source.

INTRODUCTION

We can distinguish the artifacts of technical and biological origin:

- **Biological:** (Physiological) Eye blinks, eye movements (EOG), muscle activity including swallowing and teeth clenching (EMG), electrocardiogram (ECG).
- **Technical:** (Non-Physiological) Power supply, spurious electrical, noise from electrical engines, bad electrode contact, or its detachment, saturation of the ADC (analog to digital converter)

Body and head movements may induce not only muscle electrical activity, but also slow potential shifts due to the physical displacement of the ions double layer on the electrode surface.

> **NOTE:** *(Artifacts prevention*
> Prevention of artifacts is always preferable to removing or compensating for them post hoc during data analysis.

Let's some examples of both types

PHYSIOLOGICAL ARTIFACTS

There are several types of **physiological artifacts:**

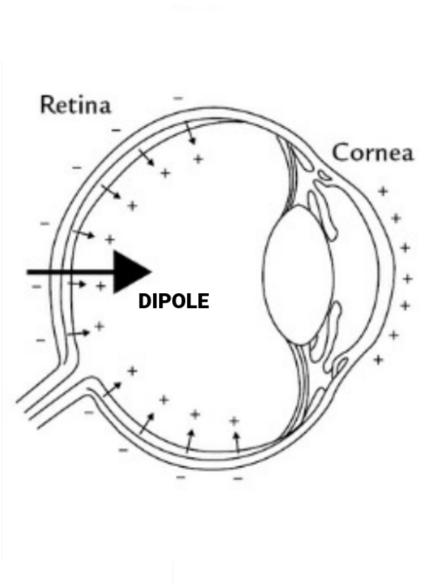
- Eye-related artifacts
- Muscle artifacts
- Cardiac artifacts
- Sweating artifacts

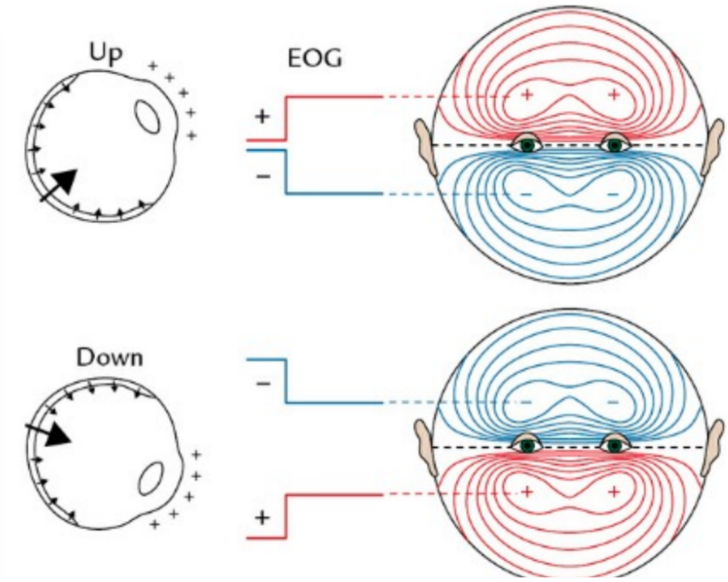
Eye-related artifacts

Ocular artifacts arise because the eye is an electrical dipole, with the cornea positively charged with respect to retina at the back of the eyeball.

Eye movement

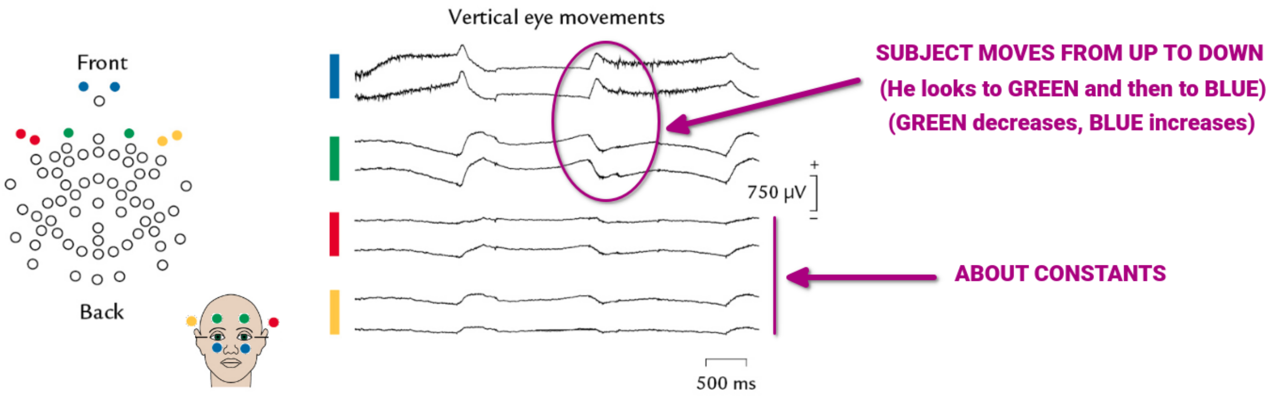
During **eye movements**, the eyeball moves within the volume conductor and changes potential measured on the scalp. Using reference at infinity, we have a positive potential in the direction pointed by the **cornea** as shown in the following figure:

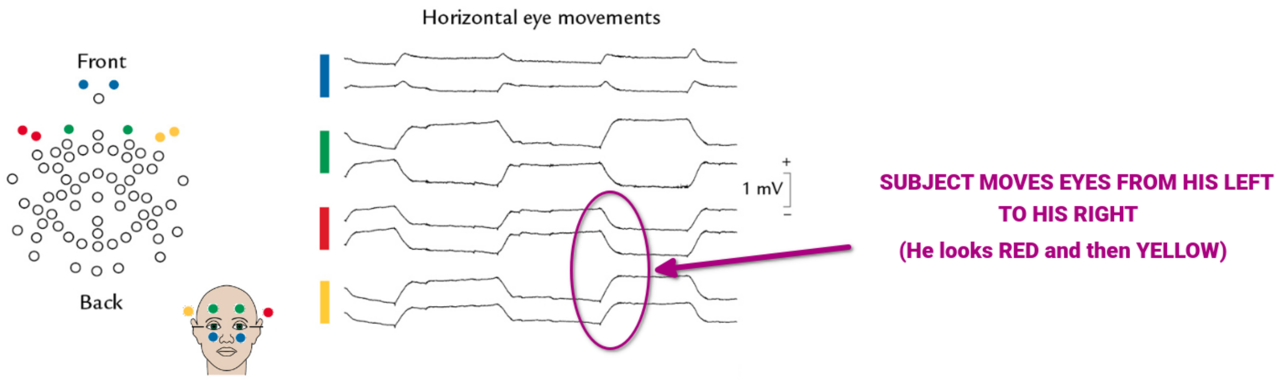




VERTICAL EYE MOVEMENTS
(Reference at infinity)

Let's see some example of eye movements artifacts:



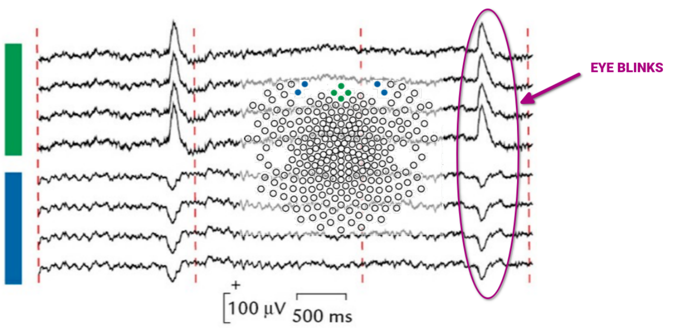


These ocular artifacts are of the order of $0.5mV$.

> **NOTE:** *(Relative eye movements)*
> In order to avoid saturation of ADC, most of EEG equipments apply an high pass filter to inputs, this implies that every constant signal goes slowly to zero. Therefore we are not able to know absolute eyes positions of the subject but only relative movements.

Eye blinks

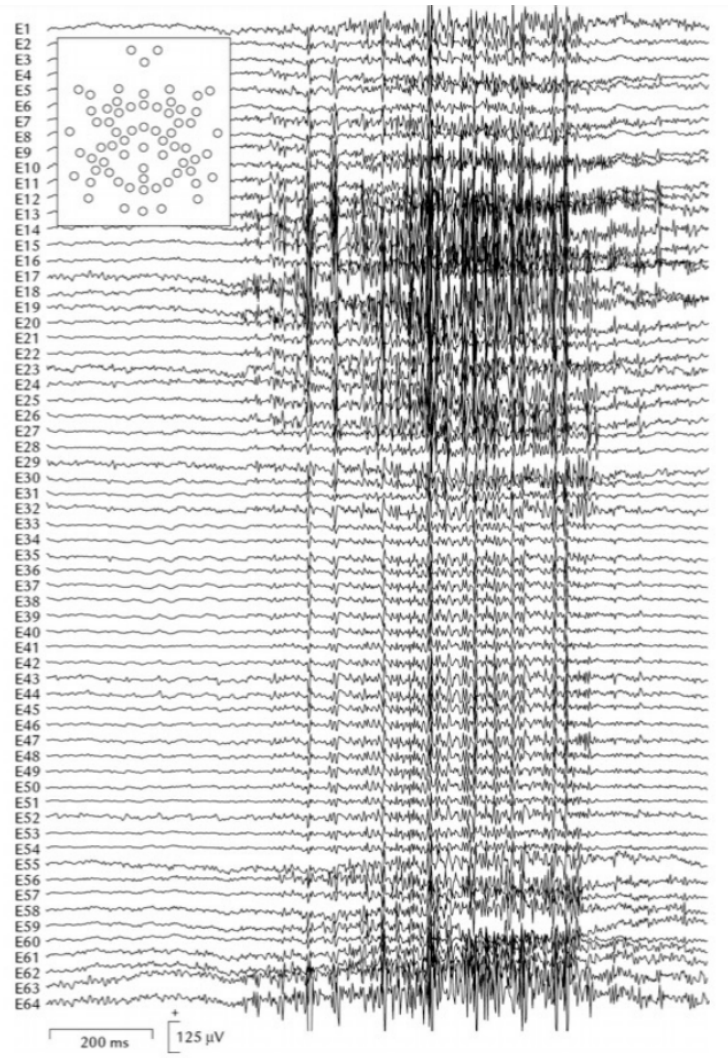
During eye blinks, the volume conductor changes because of movements of the eyelids that due to their moist inner surface provide a well-conducting pathway for current flow.



These ocular artifacts are of the order of $0.5mV$, monophasic, $200-400ms$

Muscle artifacts

Muscle contractions are seen as artifacts in the $100 - \mu V$ or $1 - mV$ range, with a wide frequency spectrum from tens of Hz to a few kHz, thereby in the same frequency range as beta- and gamma-band signals.
The experimenter should always ask the subject to relax the muscles and carefully inspect the EEG traces to check for contamination. Despite the instructions, subjects may be unable to relax and release muscle tension.

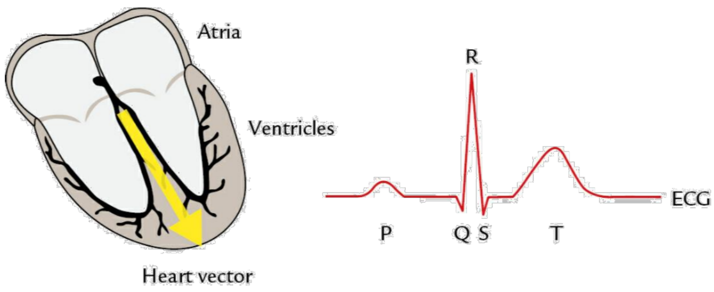


Cardiac artifacts

We distinguish two different types of cardiac artifacts, the first one is not a real issue, let's see both of them.

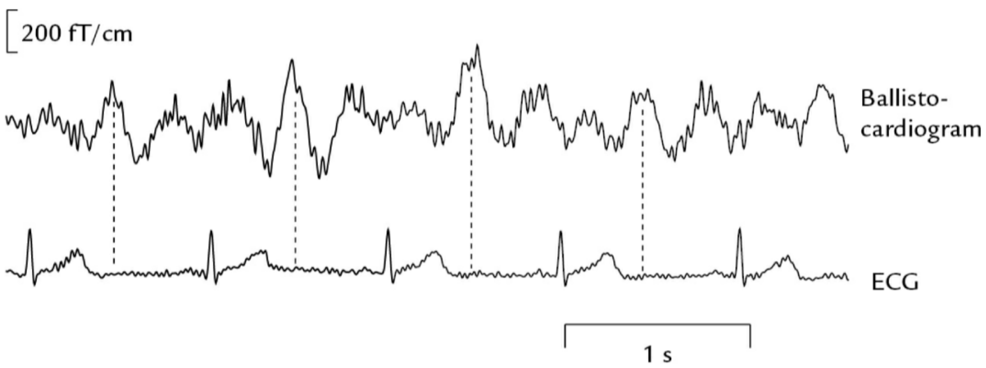
Heart electrical potential

Electric potentials generated in the heart can be easily recorded, and are known as the electrocardiogram (ECG). In the literature they can also be referred to by the German-derived acronym EKG.
In EEG recordings using noncephalic reference electrodes prominent ECG artifacts occur, but less so when the EEG reference is placed on the subject’s head since heart electrical potential is the same over the head. Therefore, these artifacts usually do not occur during a standard EEG recording.



Ballistocardiogram

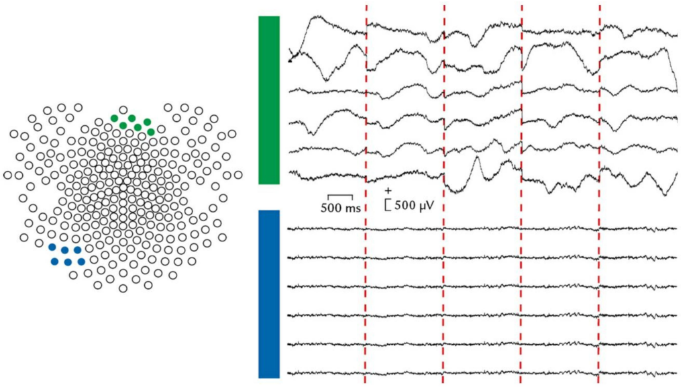
Another important cardiac-cycle-related artifact in EEG data results from the ballistocardiogram that reflects small electrode movements due to a nearby blood vessel in the scalp. This artifact is usually clearly visible during online monitoring, and it can be eliminated by moving the EEG electrode a small distance.



EMG (magnetoencephalography)

Sweating artifacts

Sweating can be a problem in EEG recordings as it is accompanied by electrodermal responses, high-amplitude slow (< 0.5 Hz) potentials.



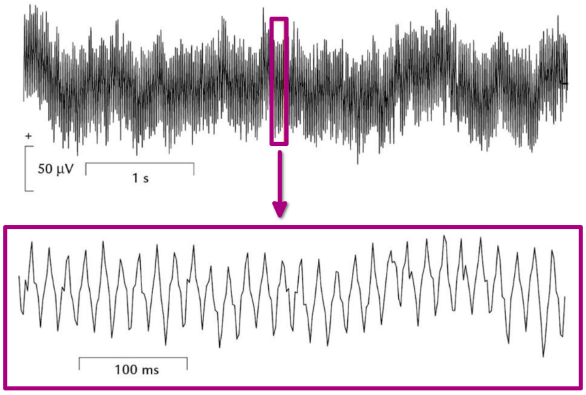
NONPHYSIOLOGICAL ARTIFACTS

There are several types of **non-physiological artifacts:**

- Powerline artifacts
- Electrode-related artifacts
- Motion artifacts

Powerline artifacts

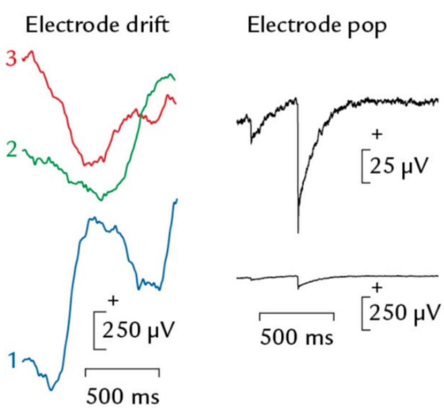
Power-line noise arises from the capacitive coupling of alternating current supplied to electrical wall outlets, thIs line noise occurs either at 50 Hz (or 60 Hz) and at their harmonics, thus, in the time domain, the signal is not necessarily a sinewave. Powerline artifacts are accentuated in EEG electrode pairs that have an impedance mismatch (reduced CMRR). Unless there are no other alternatives, a digital notch filter set to the appropriate line frequency (and its harmonics, if necessary) can remove this noise.



Electrode-related artifacts

A number of different artifacts can arise from improperly applied EEG electrodes:

- **Bad electrode:** artifacts as very slow baseline drifts
- **Electrode pop:** is an intermittent phenomenon that appears as a sharp, sudden rise followed by a gradual fall in the EEG signal of a single electrode. It can happen due to an accumulation of capacity in an electrode

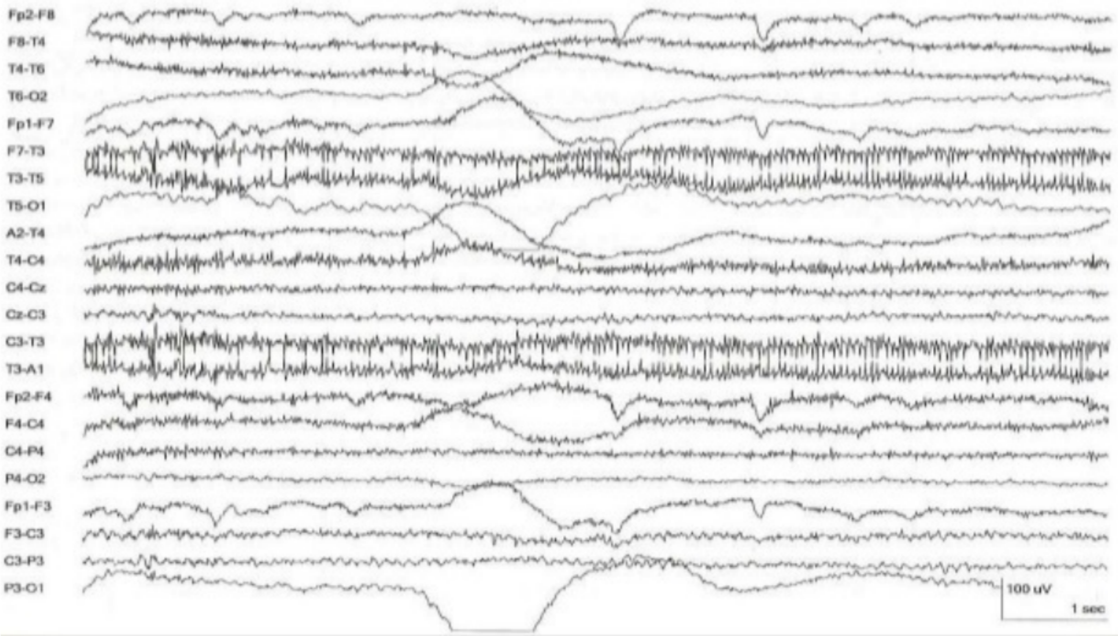


Regarding to electrode drift, sometimes it can happen that all your channels are good but there are 1 or 2 channel that are behaving differently. It means that you see something like sweating, and you see that one electrode is going up and down like 1, or 2 and 3.

In EEG recordings, artifacts can be created by the experimenter’s movements around the subject, and this problem can be exaggerated in dry air environment creating a problem with static electricity.

Motion artifacts

At the interface between metal (electrode) and ionic conductor (gel/skin) there is a displacement of ions that resembles a capacitor. Movement of the electrode (e.g. produced by movements of the subject’s head) disturbs the geometry of the "capacitor", producing a relatively strong potential.



TRUE STATEMENTS

- An EEG artifact is a potential difference due to sources outside the brain.
- Artifacts can have biological origin (such as eyes, muscles heart), can be due to external electromagnetic generators (power supply, engines, etc) or to events affecting the recording setup (electrode movements or loss of contact, saturation of the ADC).
- Artifacts can be partially attenuated or removed through signal processing during data analysis, but it is always preferable to make all efforts to prevent them when the signal is being acquired
- The eye is more positive in its frontal part (cornea) than its posterior part (retina), and thus its movements can generate large artifacts (EOG) on the EEG (of the order of 500 μV)
- During an eyeblink, the eyelid shortcuts the positive potential of the external surface of the eye, causing positive deflections of the EEG with amplitude of several hundreds of microvolt and duration of a few hundreds of milliseconds.
- A sudden upwards/downwards movement of the eyes generates a positive/negative deflection of EEG potentials (EOG) on frontal EEG channels, respectively. A sudden movement of the eyes to the right/left generates a positive/negative deflection of EEG potentials (EOG) on the EEG channel F8, respectively. The electrical activity of the muscles (EMG) has a spectral content starting at frequencies of 20 Hz and up, thus affecting the beta and gamma bands of the EEG signal
- The amplitude of the electromyogram (EMG) originated from muscles in the head can have amplitude ten times higher than the EEG signal (thus in the order of 1 mV).
- EMG artifact can easily appear on the EEG recording unless the subjects are specifically instructed by the experimenter on how to relax their face and tongue/throat muscles.
- The heart activity can contaminate an EEG recording because an electrocardiographic (ECG/EKG) artifact can directly affect the potentials, especially if the reference electrode is not placed on the head.
- The heart activity can contaminate an EEG recording because a ballistocardiographic artifact is indirectly generated by the pulse of a blood vessel causing movements of a nearby electrode.
- Sweating can affect the EEG, causing a slow changing and high amplitude artifact (below 0.5 Hz, up to a few mV)
- Powerline noise is an artifact caused by the capacitive coupling between (i) the conductors carrying the alternating (typically at 50Hz) current power supply and (ii) the recording setup including the subject.
- The powerline noise affects a very narrow frequency band of the recorded signal around 50 Hz (or 60 Hz, depending on the powerline frequency). Other frequency bands can be affected at multiple frequencies that at multiple (typically odd multiples) of 50/60Hz.
- Powerline noise is accentuated by asymmetries in the recording electrode pairs, such as impedances and cable path, because asymmetries prevent the noise to be rejected by the amplifier’s common-mode rejection capabilities.

- Notch filters effectively remove powerline noise because they selectively reject the narrow band affected by the artifact, preserving almost entirely the useful signal.
- Movement of the subject’s head may produce slow artifacts on the EEG recording, whose waveform is closely related to the timecourse of the movement. Since the potentials originate from the mechanical displacement of the charged double layer at the electrodes interface, these artifacts are less pronounced when nonpolarizable electrodes

EEG ANALYSIS

In this section we will see some very basic EEG data analysis techniques.

INTRODUCTION

Off-line data inspection

The first step before any kind of analysis is to perform a **(off-line) data inspection**, namely review raw data before processing and analysis. Bad EEG segments or channels must be removed from the data, in this step is important to understand which artifacts mostly affect the results of the analysis, so that we do not waste data by rejecting too much of it. From a practical point of view, dealing with bad channels or epochs will depend on the analysis software we are using.

Evoked potentials and Event-related potentials

EEG data are typically collected in experiments using multiple trials per condition, as the signals of interest may be quite small, but one of the basic assumptions of analysis is that the signal remains the same during the whole experiment although it is masked by noise.
With the term **evoked potential** (EP) we indicate the stimulus-driven activity. In contrast, the terms **event-related potential** (ERP) is used more generally to describe changes in EEG signals triggered by either *external* stimuli or related to internal mental or task-related events.

> **OBS:** *(Stimulus Types)*
> A stimulus can be also produced by the subject, for example when he moves a finger. So in general we can distinguish two different classes of stimulus:
>
> - **External Stimulus:** External stimulus independent by the subject
> - **Internal Stimulus:** Interval stimulus caused by the subject
>   --> **Internal mental events**
>   --> **Task-related events**

Types of EEG analysis

In this section we will see two different kind of EEG analysis:

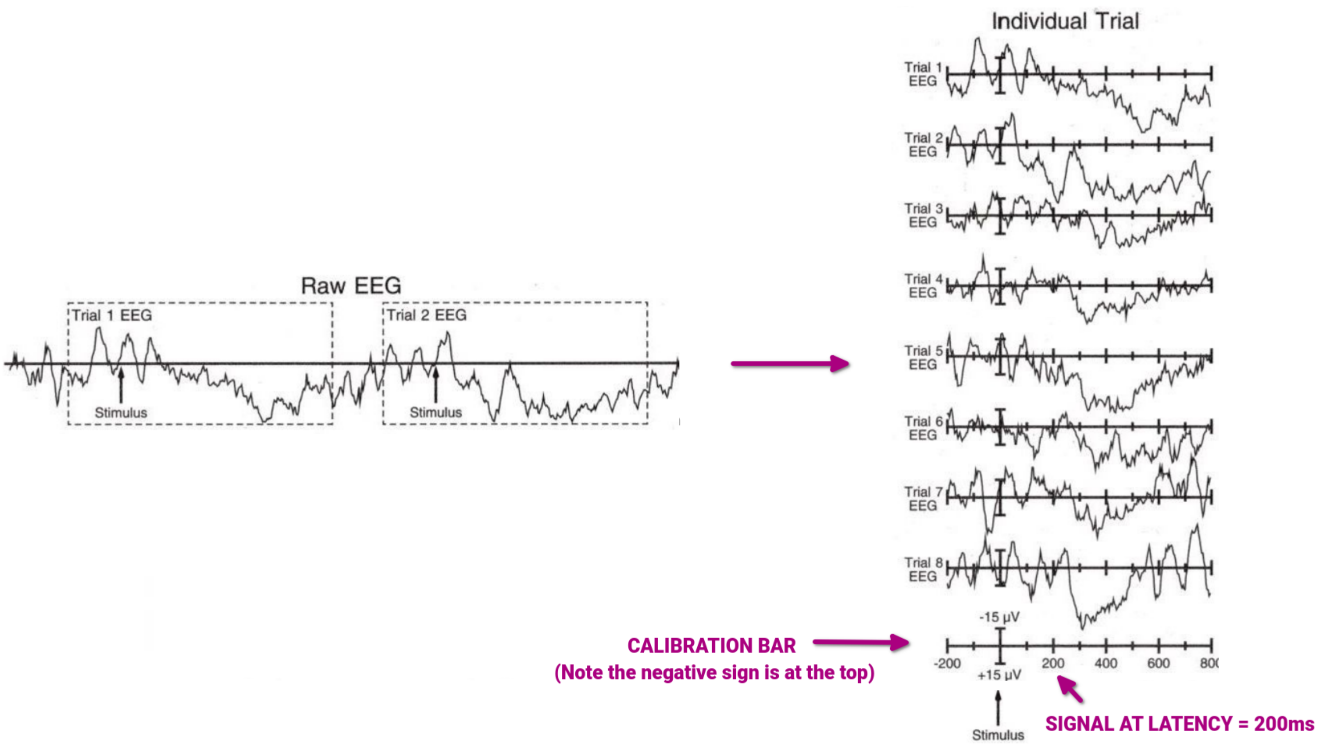
- **Analysis EP EEG:** (EP = Evoked Potential) Analysis of an EEG recording **during** *external* stimulation
- **Analysis Spontaneous EEG:** Analysis of an EEG recording **without** *external* stimulation

ANALYSIS EP EEG: SYNCHRONIZED AVERAGING

About the first type of EEG analysis, we see only **(synchronized) averaging** that we have already briefly introduced in previous sections.

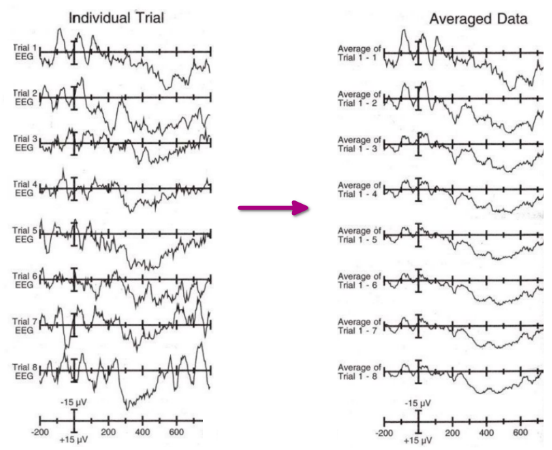
Segmentation

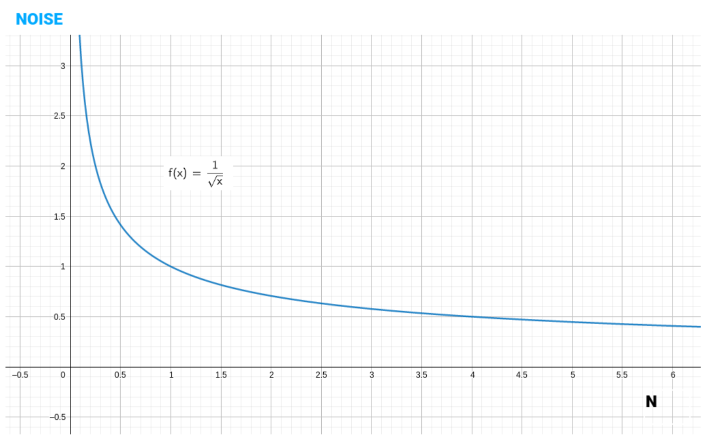
The continuous EEG record must first be “cut up” into epochs (or segments)—consisting of a pre-stimulus or a pre-event period, followed by a period of time suitably long enough to be able to highlight the particular activity of interest.



Averaging

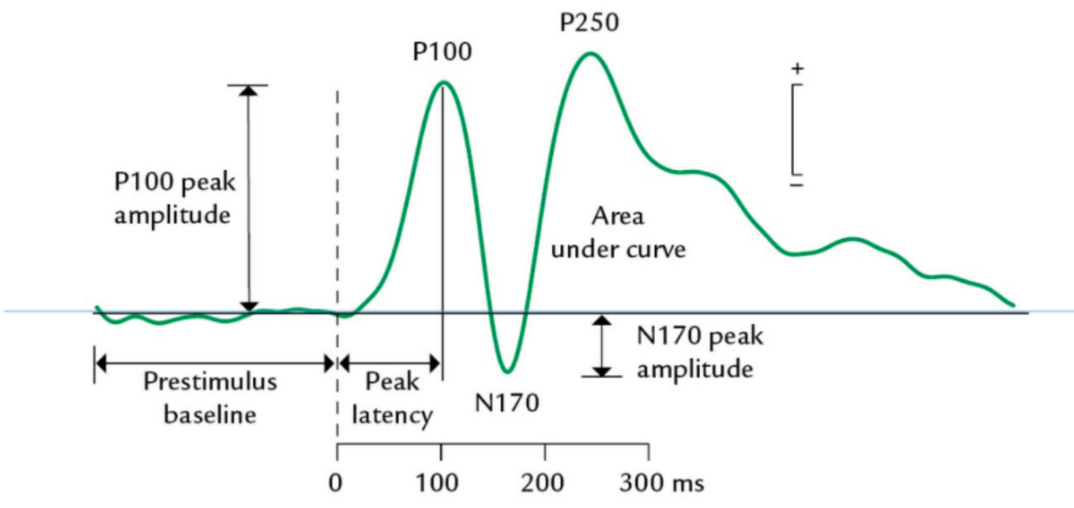
Signal averaging is based on the assumption that the recorded time-varying signal $x(t)$ comprises a distinct signal embedded in noise. Averaging N responses would improve the signal-to-noise ratio by $\sqrt{N}$, namely noise amplitude decreases as $\frac{1}{\sqrt{N}}$ where N is the number of trials.





Nomenclature of ERP components

The peak amplitudes (with respect to pre-stimulus baseline) and latency (with respect to stimulus onset, or relative to a motor event) are commonly measured.



In the vast EEG and MEG literature, the nomenclature of evoked responses is diverse and confusing. An old naming convention numbered successive deflections separately for scalp-positive (P) and scalp-negative (N) EPs—where the polarities refer to the "active electrode"—resulting in notations such as P1, N1, P2, N2, P3. To make matters worse, letters are sometimes added to these labels, such as P3a and P3b.
A less ambiguous way is to combine the polarity (N or P) of the response peak or trough with the nominal peak latency in milliseconds, for example, P60, N100, and P200

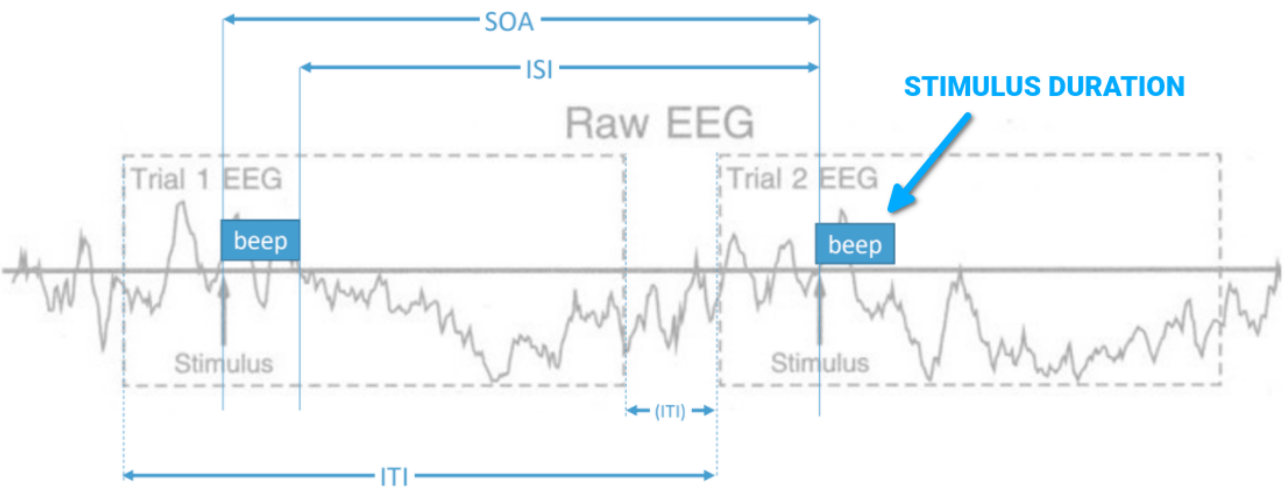
> **NOTE:** *(Nominal latency)*
> Note that in most cases it is clear enough to use in the response name the approximate (or nominal) latency (and not the measured individual latency). For example, the mean peak latency of the N100 deflection may vary between, say, 90 or 110 ms without causing any confusion in the nomenclature

Standard times

Let's see some standard time quantities:

- **Stimulus Onset Asynchrony (SOA):** Refers to the time between two successive stimulus onsets
- **Interstimulus interval (ISI):** Refers to the time between the offset of one stimulus and the onset of another
- **Intertrial interval (ITI):** Refers to the time between the beginning of subsequent trials (during which multiple stimuli may be presented).

NOTE: *(ITI definition)*
Beware, some authors define the ITI as the pause period between the end of a trial and the beginning of the next one!

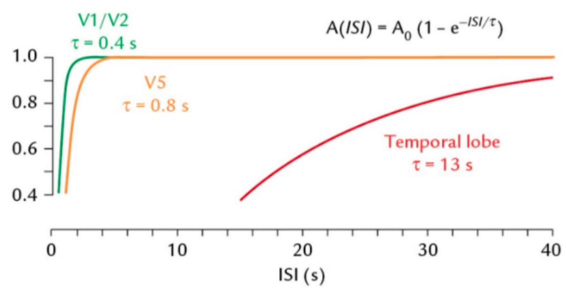


OBS: *(Long stimulus duration)*
SOA and ISI/ITI can be quite different from each other if the stimulus duration is long, indeed we have:

$$\text{ISI} = (\text{SOA} - \text{stimulus duration})$$

Effect of stimulus timing

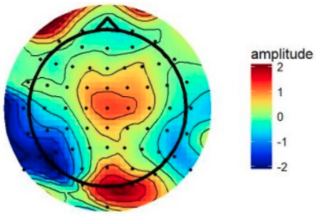
The shape and amplitude of the ERP may strongly depend on the timing of stimuli. In general, the longer the latency of the response, the more sensitive the response is to stimulus repetition rate, namely brain tends to respond less stronger to the same stimulus repeated several times in a short interval. Therefore, since we want to reduce the overall time of the EEG recording we have a tradeoff between number of stimuli (noise decreases in N) and SOA ($\text{total time} = N \cdot SOA$)



See also exam July 2020 part 2, question B.1

Mapping

In a 2D display of EEG data, topographic scalp voltage maps can depict interpolated voltages between the electrode locations at any time point. Most typically, such maps are displayed at times of response peaks and troughs.



ANALYSIS SPONTANEOUS EEG

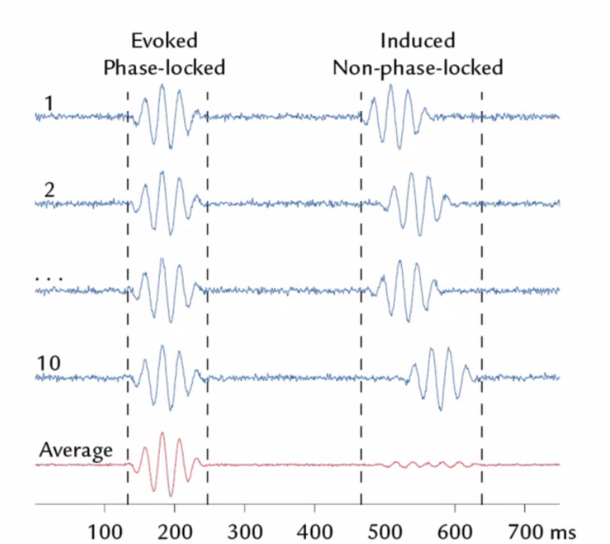
The second type of EEG analysis that we see is **analysis of spontaneous EEG** where we are interested to study EEG response after a **internal stimulus** namely a stimulus produced by the subject (e.g. finger movement or mental task)

Non-Phase locked response

The first huge different between external and internal stimulation is that in the last one, response time of subject can vary, namely time between stimulus delivery and EEG signal response is not always the same (**non-phase response**). This time quantity is called **jitter**.
Spontaneous or ongoing brain activity is always present irrespective of whether or not a subject performs a task, and it can influence how evoked and induced responses evolve and how they are related to perception and cognition

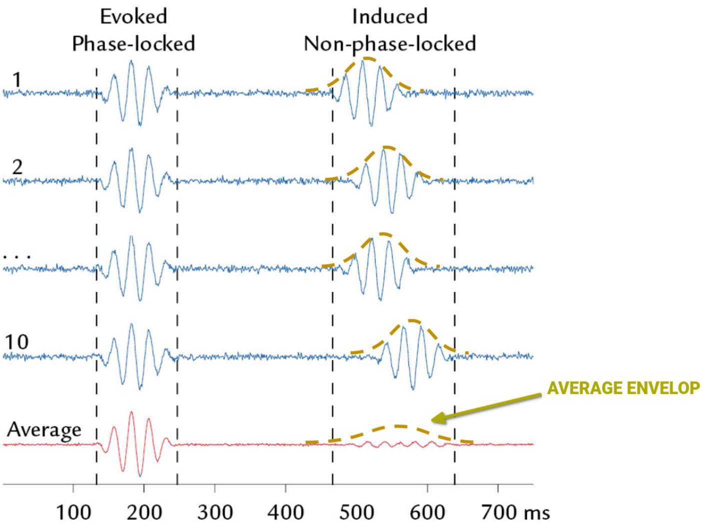
We have seen in previous EEG analysis, where the evoked activity is time-locked (*phase-locked*) to the stimulus, that EEG stimulus response can be uncovered by stimulus-locked averaging (synchronized averaging). However, this technique cannot be applied easily in this case due to jitter.

Let's compare phase-locked response and non-phase locked EEG responses assuming their shapes are the same at each trial.



As we can see from the previous figure (EEG non realistic), averaging non-phase-locked responses provided a resulting zero response that will be lost for sure behind ongoing EEG activity. To perform averaging properly we should align the response of each trial, this method is not so easy in practice and can be considered as an advanced technique.

A more simple technique to averaging non-phase-locked responses is to use compute the *average envelop*, doing that the final average will not be zero since each envelop is always positive:



Let's now see how this technique is applied in real scenarios where we call a negative envelop **event related desynchronization (ERD)** and a positive envelop **event related synchronization (ERS)**.

Event related desynchronization/synchronization (ERD/ERS)

Let's define more clarity these two quantities:

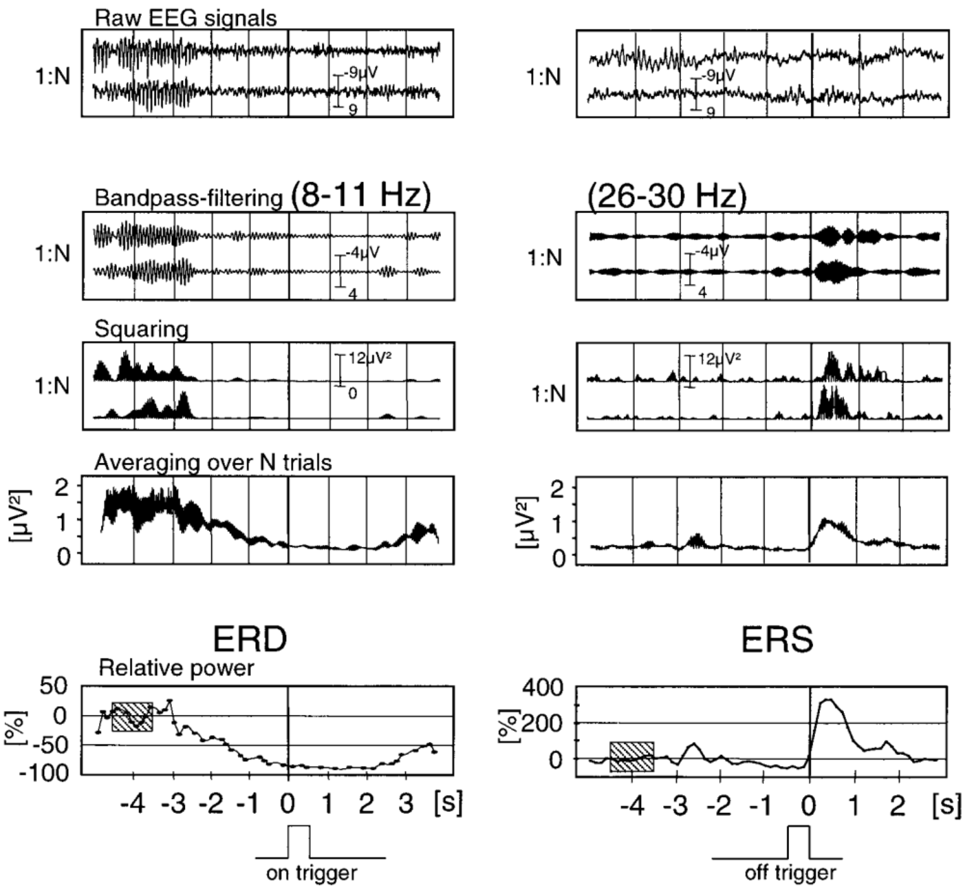
- **Event Related Desynchronization (ERD)**: *Negative* variation of the EEG power rhythms at one time relative to a *baseline* period at another time
- **Event Related Synchronization (ERS)**: *Positive* variation of the EEG power rhythms at one time relative to a *baseline* period at another time.

The steps to follow to obtain these quantities from **N** raw EEG trials are the following:

1. **Bandpass filtering** of all event-related trials
2. **Squaring** of the amplitude samples to obtain power samples
3. **Averaging of power samples** across all trials
4. **Averaging over time samples** to smooth the data and reduce the variability

At the end the result is expressed as percentage variation of the signal respect to the baseline.

In the following figure one example for each type of envelop is shown: one example of dominant ERD in the alpha band on the left side and one example with dominant ERS in the beta band on the right side.



In the case of ERD we are interested into mu rhythmic, indeed the stimulus correspond to a finger movement by the subject.

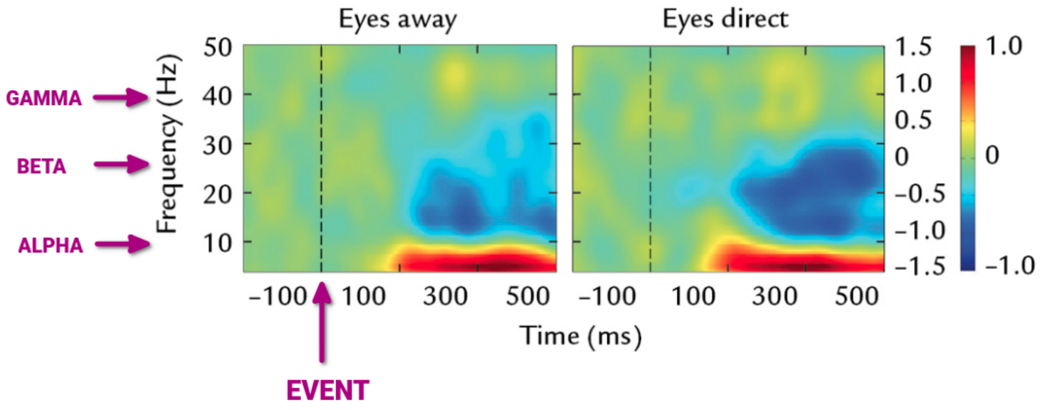
> **NOTE:** *(Large time window)*
> As we can see from the previous figure, it is necessary to start EEG signal recording several seconds before the stimulus, this is necessary to detect the baseline signal. Baseline is "faraway" because the subject's brain prepares in advance to move the finger, indeed this is a voluntary movement.

> **REFERENCE:** *Event-related EEG/MEG synchronization and desynchronization: basic principles* - G. Pfurtscheller, F.H. Lopes da Silva (1999)

TIME-FREQUENCY EEG ANALYSIS

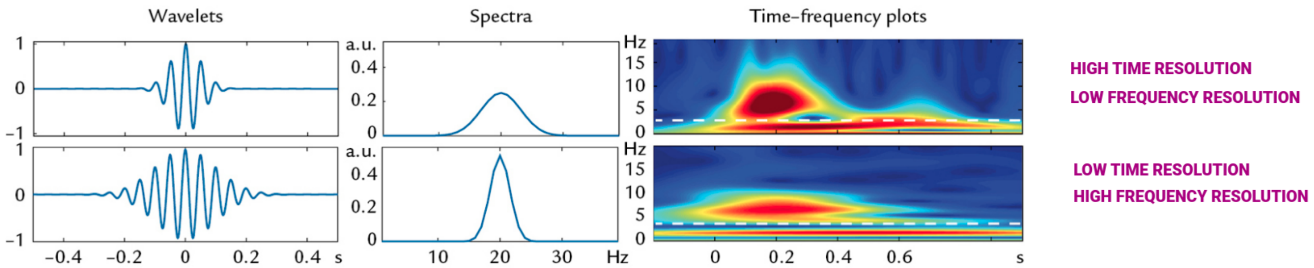
Time–frequency analyses go one step further by computing and visualizing the spectral or amplitude content of the signal as a function of time simultaneously for all frequencies of interest. Fourier transforms, Hilbert transforms, and wavelet-based approaches can be used to calculate MEG/EEG signal power (amplitude). Using this procedure, features in MEG/EEG data can be visualized in both time and frequency.

The following figures show examples of time-frequency plots, in particular they show the frequency of EEG signals during a visual task with two conditions: a face gazing away (Eyes away) and at the subject (Eyes direct) as a function of time. Vertical color scale indicates power in dB. The two conditions show very similar profiles of activity: a prolonged increase in theta and alpha activity and a prolonged *decrease in beta activity after the stimulus (vertical broken line). Passband 0.1-200 Hz. Data are from a nine-electrode cluster centered on the left sensorimotor scalp.*



Resolution tradeoff

Improving time resolution worsens spectral resolution (and vice versa)



FURTHER READING:

- Ch 12. auditory responses, including steady-state
- Ch 13. visual responses, including steady-state
- Ch 14. somatosensory responses
- Ch 16. motor function
- Ch 17. change detection (CNV, MMN, P300, ErrN, ...)

03 - BASICS OF SIGNAL PROCESSING - PART 1

TOPICS

- to do

Reference: Steven W. Smith The Scientist and Engineer's Guide to Digital Signal Processing CHAPTER 3 https://www.dspguide.com/pdfbook.htm

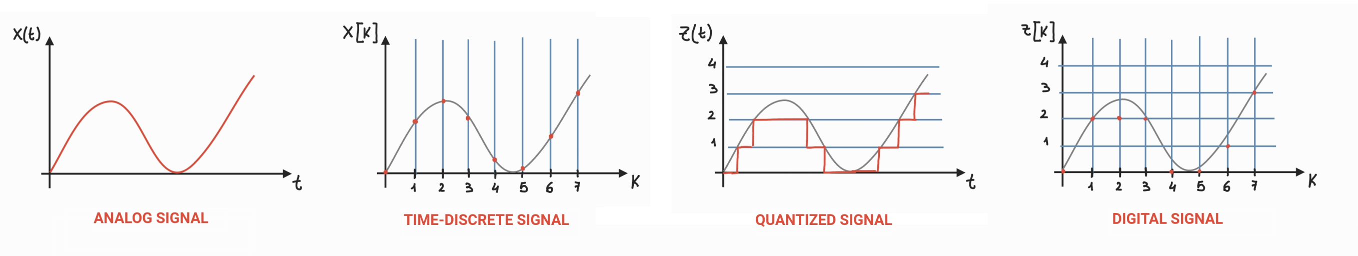
ANALOG TO DIGITAL CONVERSION (ADC)

During an **analog to digital conversion** we perform two different kinds of discretization steps:

- **Sampling:** Discretization of the signal with the **respect to time**
- **Quantization:** Discretization of the signal with the **respect to its value** at each instant

$$\text{ADC} = \text{SAMPLING} + \text{QUANTIZATION}$$

$$\text{DIGITAL SIGNAL} = \text{QUANTIZED SIGNAL} + \text{TIME-DISCRETE SIGNAL}$$



Let's see each of them separately starting from **quantization**.

QUANTIZATION

Quantization step consists in the discretization of signal amplitude.
Three components must be taken into account during quantization procedure design:

- **Noise:** Introduced during analog to digital values conversion
- **Precision (LSB):** of conversion
- **Full scale (Range):** indicates the highest analog input amplitude manageable

Let's see briefly all of them.

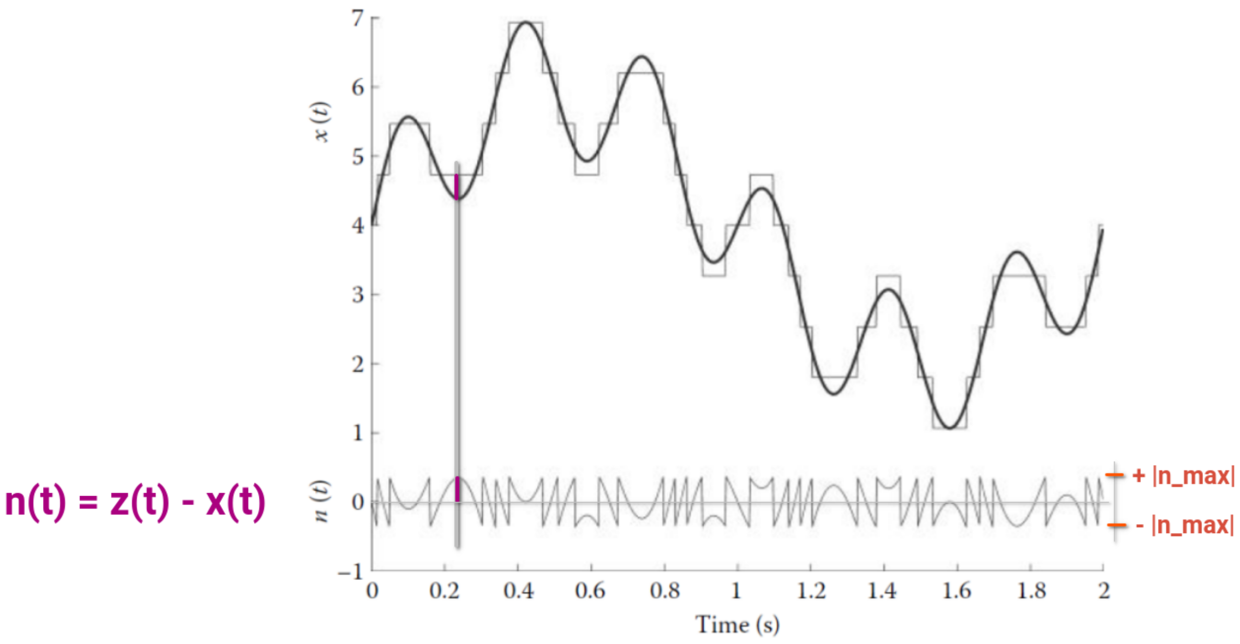
Noise

During quantization, at each instant t we have to approximate analog input signal amplitude to a specific digital value. This approximation from analog amplitude to digital amplitude introduces an error in acquisition pipeline that we can model as a random **noise** $n(t)$.
We define this noise as the difference between quantized signal $z(t)$ and analog input signal $x(t)$:

$$n(t) = z(t) - x(t)$$

At each instant, the noise amplitude is limited and depends on instrumentation specifications, we can assume it is **uniformly distributed** in a range of value $[\ +|n_{max}|, -|n_{max}|\ ]$.

The following figure shows an example:



> So we model error introduced by quantization step as a limited **additive** random noise uniformly distributed:
>
> $$n(t) = z(t) - x(t) \Leftrightarrow z(t) = x(t) + n(t)$$

> **Examples**
>
> - 8 bit ADC -> RMS noise = 0.29/256 ~= 1/900 Range.
> - 12 bit ADC -> RMS noise = 0.29/4096 ~= 1/14,000 Range
> - 16 bit ADC -> RMS noise = 0.29/ 65536 ~= 1 /227,000 Range

Precision (LSB)

An important technical specification of the instrumentation is **precision** that indicates the distance between adjacent quantization levels. In jargon, precision is called **LSB** (Least Significant Bit) and it is usually measured in $Volts$.

Quantization noise depends on precision, indeed maximum absolute noise amplitude is equal to half LSB, namely half of the distance between two digital levels. Moreover, also its **standard deviation** σ depend on LSB.

$$n(t) \in \left[+\frac{1}{2}LSB, -\frac{1}{2}LSB\right]; \quad mean(n(t)) = 0; \quad \sigma(n(t)) = \frac{1}{\sqrt{12}}LSB$$

In addition to quantization noise, we have also an independent **intrinsic noise** caused by measurement procedure, therefore the LSB must be small enough so that quantization noise is smaller than the input signal's intrinsic noise, in this way the quantization noise is negligible. We can understand why this is required by looking at the standard deviation of the overall noise:

$$\sigma_{int+quant} = \sqrt{\sigma_{int}^2 + \sigma_{quant}^2}$$

> **Probability recall**
> Variance $Var[]$ of a random variable X is the square of its standard deviation σ:
>
> $$Var[X] = \sigma_X^2$$
>
> The variance of the sum of two **uncorrelated random variables** X, Y is the sum of their variances:
>
> $$Var[X+Y] = Var[X] + Var[Y]$$

Full scale (range)

Another important technical specification of instrumentation is range of value (**full scale**) that can accept in input, values outside are saturated.
The **range** depends on precision (LSB) and the number of bits used to store the amplitude signal digitally:

$$Range = LSB * 2^{bits}$$

SAMPLING

Sampling step consists in the discretization of signal over time. (no over signal amplitude)
The most important components that must be taken into account during sampling procedure design is:

- **Sampling frequency:** Number of sample acquired in a second

Sampling frequency (sampling rate)

The proper **sampling frequency** to use is given by **Nyquist-Shannon theorem**.

Sampling frequency: Nyquist-Shannon theorem

> A continuous signal can be "properly" sampled, **only if** it does not contain frequency components above **Nyquist frequency**.

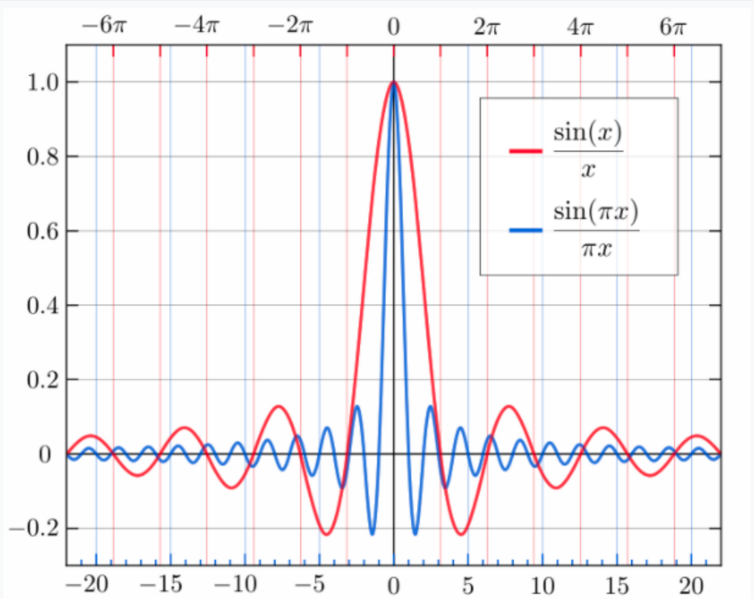
> **Nyquist frequency** $f_{Nyquist}$ is defined as **one-half** the sampling rate $f_{sampling}$
>
> $$f_{Nyquist} = \frac{f_{sampling}}{2}$$

With the term **properly** we mean that we are able to reconstruct the analog signal from the sampled signal with no error.

> **NOTE:** *(Exact reconstruction, no quantization)*
> Nyquist-Shannon theorem is a very powerful mathematical result, it guarantees no loss of information only if we sample the signal **without quantize** it. If we quantize the signal, then theorem does not hold anymore.

The signal can be reconstructed by placing a shifted and scaled $sinc()$ function for each sample and then "summing" all functions (convolution).

Sinc



Part of the normalized sinc (blue) and unnormalized sinc function (red) shown on the same scale

> **Proper sampling rate**
> From Nyquist theorem, given a signal with the highest frequency component f_{max}, the

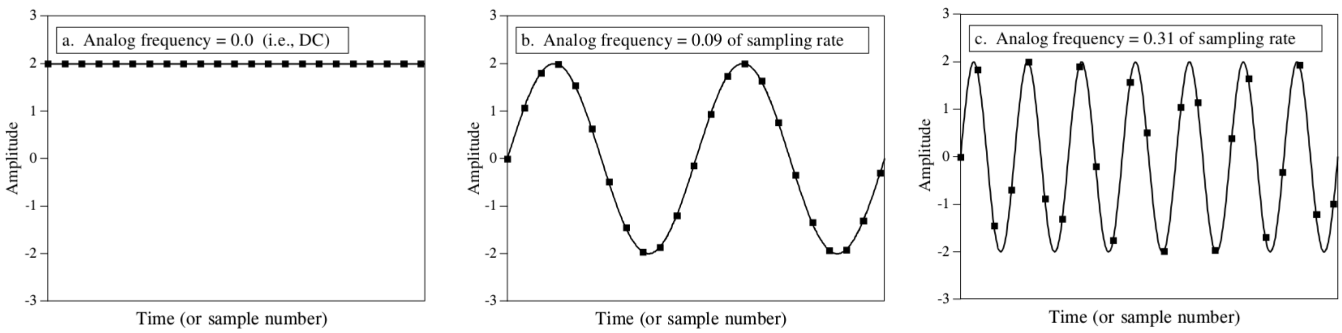
> sampling rate $f_{sampling}$ must satisfies:
>
> $$f_{sampling} \geq 2f_{max}$$

Sampling frequency: Aliasing

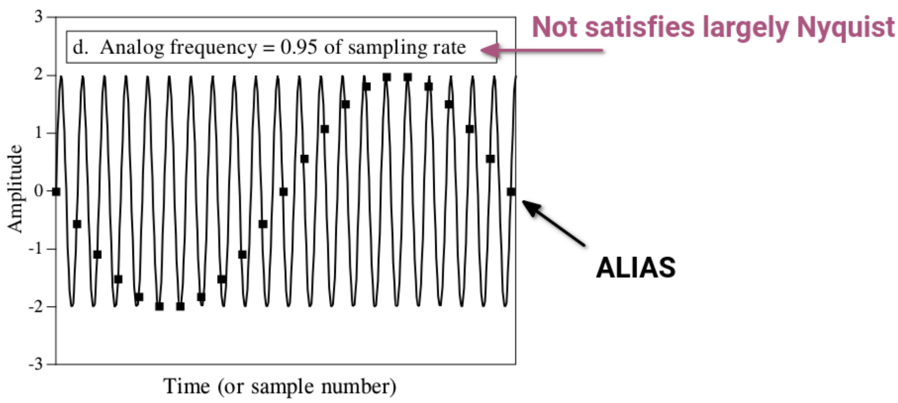
When we do not respect Nyquist theorem, we are not able to reconstruct the sampled signal properly, indeed during digital to analog (re)conversion we do not reconstruct the original analog signal but an other signal that acts like as *alias* of original one. This phenomena is called **aliasing**.

Let's see an example of aliasing, we start with a proper pair of signal frequency and sampling rate. Then we increase the frequency of the analog input until we broke Nyquist theorem.

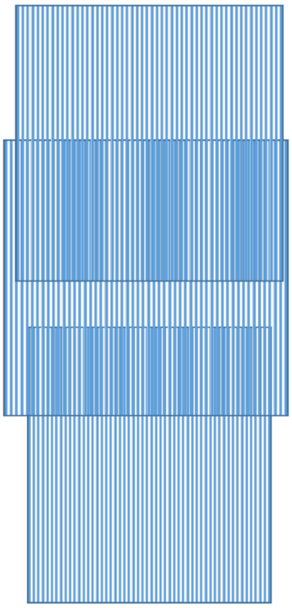
In the following, we do not have aliasing:



In this last case instead, input signal frequency is too high respect to sampling rate, so we will be no able to reconstruct original signal but an its alias.
Since the input signal frequency is 0.95 of sampling rate, it means that we are able to take only 5% of input signal samples, therefore the reconstructed signal (alias) has a frequency equals to 0.05 of original signal.



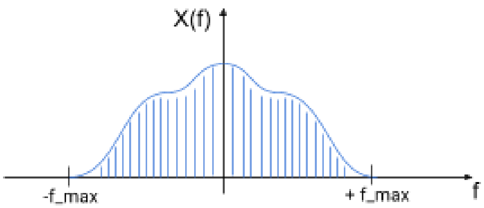
Visual aliasing
We are used to have experience with **visual artifacts**, here an example:



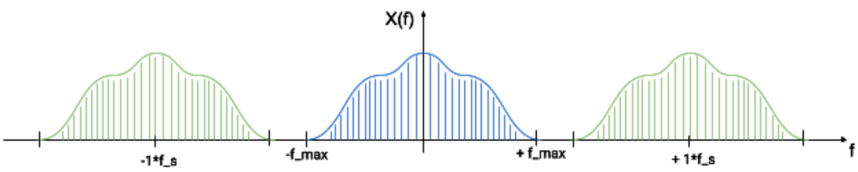
Aliasing: Frequency spectrum

Let's try to understand the reason why behind Nyquist theorem.

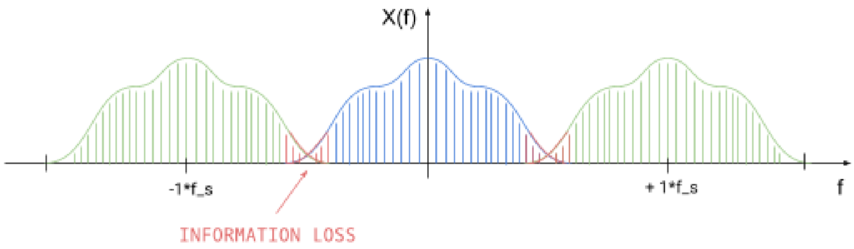
A signal can be represented with each spectrum of frequency, let's suppose the following one is the spectrum of input analog signal to be sampled:



When we perform sampling, we replicate signal spectrum along frequency axis by shifting original spectrum to multiples of sampling rate (f_s), if we respect Nyquist theorem we have:



Instead, if we do not satisfy Nyquist theorem we have overlapping between spectrums that cause loss of information, namely we have aliasing:



> **OBS:** *(Aliasing filter)*
> Since real analog signals have a huge spectrum frequency (also infinity), in practice an analog filtering is applied before ADC circuit.

ANALOG SIGNAL PROCESSING AND FILTERS

Real ADC are limited by (among other factors):

- Input range
- Maximum sampling frequency

The analog signal must be conditioned so that range and bandwidth do not exceed what is strictly needed for a faithful representation of the signal of interest. So we have to remove non-interesting part of the signal (a.k.a. noise) by performing **analog** filtering.

TRUE STATEMENTS

- The Shannon's theorem (sampling theorem) states that a continuous signal can be properly sampled only if it does not contain frequency components above one-half of the sampling rate.
- In Analog to Digital Conversion (ADC), the Nyquist frequency equals half of the sampling frequency.
- The reconstruction of an analog signal from its sampled version is equivalent to the sum a set of a set of $sinc()$ functions, one for each sample, each centered on the time of the respective sample, whose amplitude equals the sample value.
- Aliasing occurs when an analog signal is sampled outside the conditions set by the Shannon's theorem.
- When aliasing occurs in ADC, a sinusoidal component with frequency $f_0 \in (f_{Nyquist}, f_{sampling})$ is reconstructed as a sinusoidal component at

$$f_{aliasing} = f_{sampling} - f_0 \in (0, f_{Nyquist})$$

- Aliasing can be prevented by applying an analog low-pass filter with cutoff frequency lower than $f_{Nyquist}$ to the analog signal (i.e. before it is converted).
- Quantization (i.e. approximation of the analog value of a sample to the nearest among the allowed quantization levels) introduces a noise whose amplitude is proportional to the width of the quantization interval:

$$\sigma_{quant} = \frac{1}{\sqrt{12}} LSB$$

- Quantization divides the input range of the ADC into (approximately) 2^{N_BITS} intervals, where N_BITS is the number of bits of the ADC.
- Given a fixed number of bits N_BITS of the ADC, choosing a large input range increases the quantization error, while choosing a small input range increases the chance that the signal is clipped (i.e. the input range is saturated).
- Appropriate application of an analog filter (i.e. before the analog signal is converted) may prevent saturation by removing high amplitude artifacts in specific frequency bands.

04 - BASICS OF SIGNAL PROCESSING - PART 2

TOPICS

- Basic measures for signal characterization
  -- Mean, variance, RMS, ...
  -- Examples of deterministic and stochastic signals
  -- Amplitude distributions, including normal (gaussian) distribution
- Central limit theorem (CLT)
  -- Why we care so much about normal distributions
  -- Why noise amplitude is reduced by a factor 1/ N when we take an average of N values

Reference: Steven W. Smith The Scientist and Engineer's Guide to Digital Signal Processing - CHAPTER 2 - https://www.dspguide.com/pdfbook.htm

STATISTICS, PROBABILITY AND NOISE

INTRODUCTION

Signal can be entirely deterministic, or they can be known only by means of their statistical properties. We will introduce a number measurements to characterize both types of signals.

Probability theory deals with the mathematical modelling of random variables (numbers) and processes (signals). Statistics deals with the description of empirical observations, and with the estimation of the (usually unknown) parameters of the mathematical models of the variables and processes (stochastic signals).

Fundamental to several analysis algorithms, the Central Limit Theorem states the relevance of Normal (Gaussian) distributions in empirical sciences.

BASIC MEASURES FOR SIGNAL CHARACTERIZATION

Let's see some basic measures.

Mean

$$\overline{X} = \frac{1}{N}\sum_{i=0}^{N-1} x_i$$

When the number of samples N tends to infinity, the *empirical* mean $\overline{X}$ tends to the *mathematics* mean of the stochastic variable X:

$$\lim_{N\to+\infty} \overline{X} = \mu_X$$

So $\overline{X}$ is essentially an *estimator* of μ_X

Amplitude

There are two quantities that tell us about signal power, they are the following:

> **Average Rectified Value (ARV)**
>
> $$ARV_X = \frac{1}{N}\sum_{i=0}^{N-1}|x_i|$$

> **Root Mean Square (RMS)**
>
> $$RMS_X = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}x_i^2}$$

Deviation

Also in this case we have two different quantities used to measure **deviation** of the signal with the respect to signal mean. The first one has been defined by engineers, the second one instead is the mathematical definition.

> **Average Deviation**
>
> $$AD_X = \frac{1}{N}\sum_{i=0} N-1|x_i - \overline{X}|$$

> **Variance and Standard Deviation**
> The estimator of the mathematical **variance** is the following:
>
> $$s_X^2 = \frac{1}{N-1}\sum_{i=0}^{N-1}(x_i - \overline{X})^2$$
>
> Similarly to the mean:
>
> $$\lim_{N\to+\infty} s_X^2 = \sigma_X^2$$

NOTE: *($N-1$ division)*
At denominator we have $N-1$ instead of N due to mathematical reasons. In a nutshell, since s_X^2 is computed though an estimator $\overline{X}$, in order to be *unbiased*, s_X^2 has $N-1$ at denominator. Anyway, if we know exactly μ_X then we can use it and divide by N instead $N-1$.

So actually, average deviation is not mathematical correct but still be used for historical reasons.

Then, we have estimator for **standard deviation**:

$$s_X = \sqrt{s_X^2} = \sqrt{\frac{1}{N-1}\sum_{i=0}^{N-1}(x_i - \overline{X})^2}$$

In this case, estimator of standard deviation does not converge exactly to the mathematical standard deviation σ_X, this is due to root square that is a non-linear operator that transforms an unbiased quantity to a biased one.

$$\lim_{N\to+\infty} s_X \cong \sigma_X$$

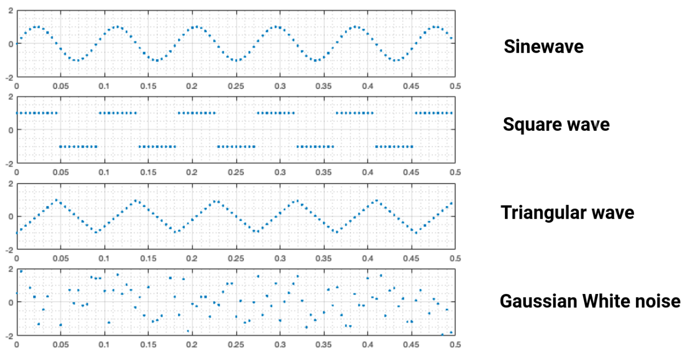
NOTE: *(Zero mean signal)*
Note that when a signal has zero mean, $RMS_X \cong \sigma_X$, and $ARV_X \cong AD_X$

EXAMPLE SIGNALS

We introduce here four signals that we will analyze in the following:

1. Sinewave (Deterministic)
2. Square wave (Deterministic)
3. Triangular wave (Deterministic)
4. Gaussian white noise (Stochastic)

The first three are deterministic waveforms, oscillating at fundamental frequency . The last one is a stochastic signal, characterized by having uncorrelated samples (whiteness, i.e. no statistical prediction can be made on the value of a specific sample by knowing the value of the others) and Gaussian distribution of the sample values (see below).



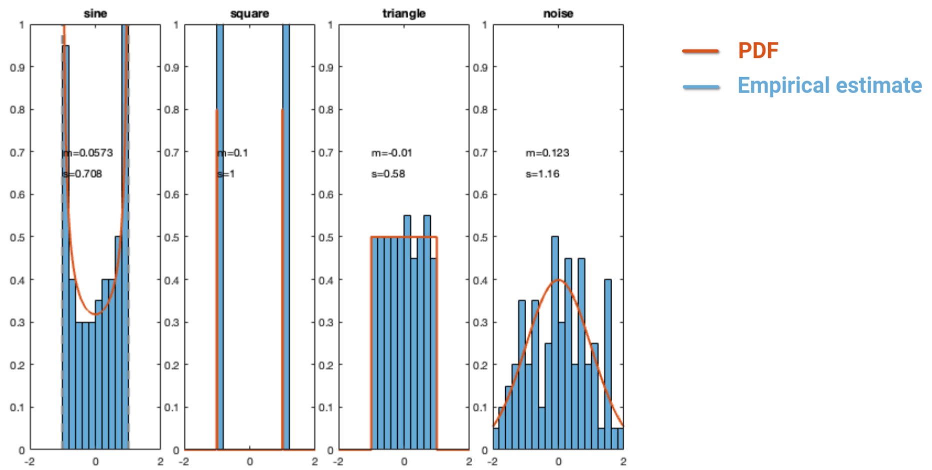
> **NOTE:** *(Uncorrelated samples)*
> The main difference between the first three signals and the last one, is that in the first ones each sample is correlated to their neighbors, namely given a sample, we can predict the next one. This is not possible in an uncorrelated signal like Gaussian white noise.

Amplitude distribution

We want to characterize here how the samples are distributed on the vertical axis: the central tendency (mean), the dispersion (standard deviation), the shape of the distribution (probability density function, PDF).

Note that we have perfect knowledge of the mathematical model we used to generate the deterministic and stochastic signals. Nevertheless, even for deterministic signals there are sources of non-deterministic outcome (e.g. finite number and uncontrolled position of time samples) which make the empiric and ideal results to overlap only in part.

As for the stochastic signal, each time we repeat this simulation, we obtain different values of mean, standard deviation, and histogram. We can only assess the compatibility of empirical observations with the mathematical model in a statistical sense.



> Values in for PDF of square signal are not correct due to an intermediate normalization step performed in MATLAB.

> **NOTE:** *(PDF Continuous variable)*
> Probability Distribute Function (PDF) in the case of continuous variables, is used to get the cumulative probability of a set of values that the variable can assume, namely the probability for which the variable assumes values on a specific range. Therefore, the PDF in a single point x is infinitesimal (about 0), however is defined in a small interval like $[x-\epsilon, x+\epsilon]$.

THE CENTRAL LIMIT THEOREM (CLT)

When N independent and identically distributed random variables X_i are averaged, the resulting variable Z tends toward a **normal distribution** even if the original variables themselves are not normally distributed. (Normal distributions are sometimes called Gaussian distributions.)

$$Z = \frac{1}{N} \sum_{i=1}^{N} X_i \to \mathcal{N}(\mu_Z, \sigma_Z^2)$$

The mean of Z equals the mean of X_i, the variance decreases by a factor N, the standard deviation by a factor $\sqrt{(}N)$:

$$\mu_Z = \mu_X \quad \sigma_Z^2 = \frac{1}{N}\sigma_X^2 \quad \sigma_Z = \frac{1}{\sqrt{N}}\sigma_X$$

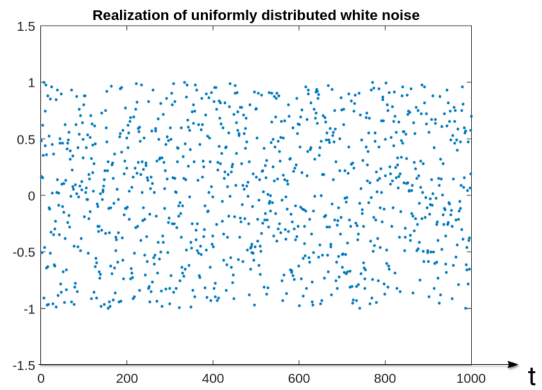
Example Noise

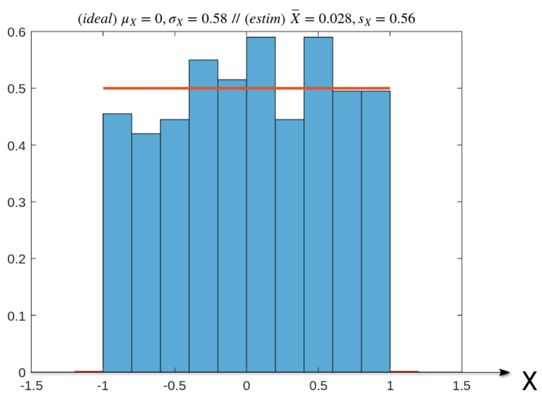
To visualize the consequences of the CLT, we first show the amplitude distribution of a (uniformly distributed) white noise, and its standard deviation. By design, the expected value of this noise's mean and standard deviation are respectively:

$$\mu_X = 0 \quad \sigma_X = \frac{2}{\sqrt{N}} \cong 0.57735$$

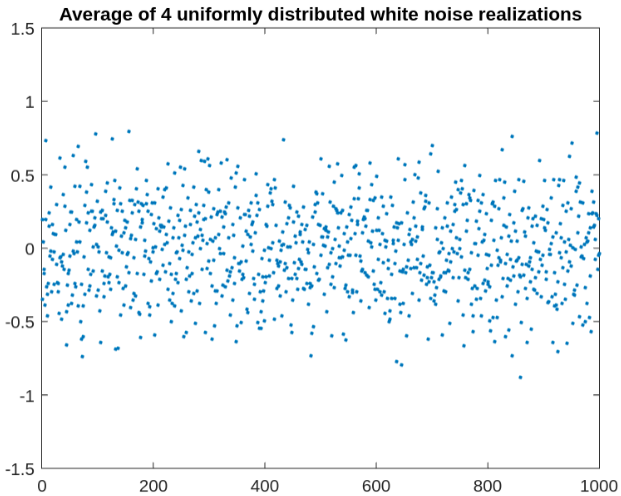
> N.B. when we estimate these parameters from empirical data we can only approach the modeled values.

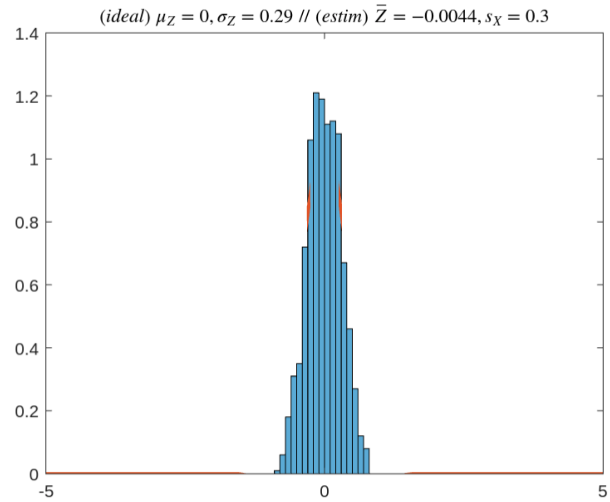
Let's consider a trial in which we collect only a white noise without any signal of interest in the recording:

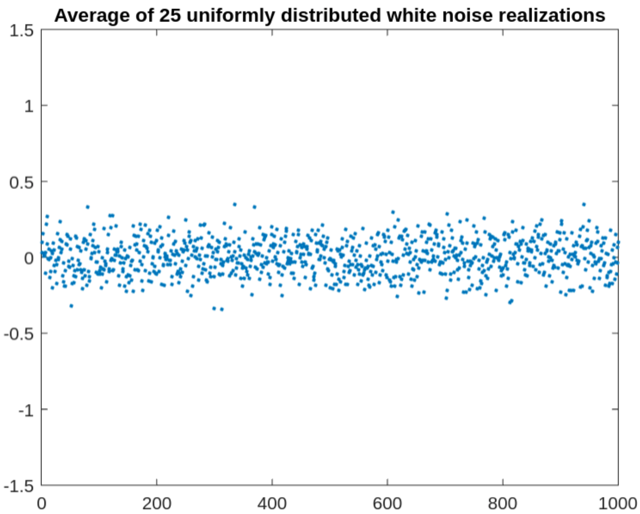


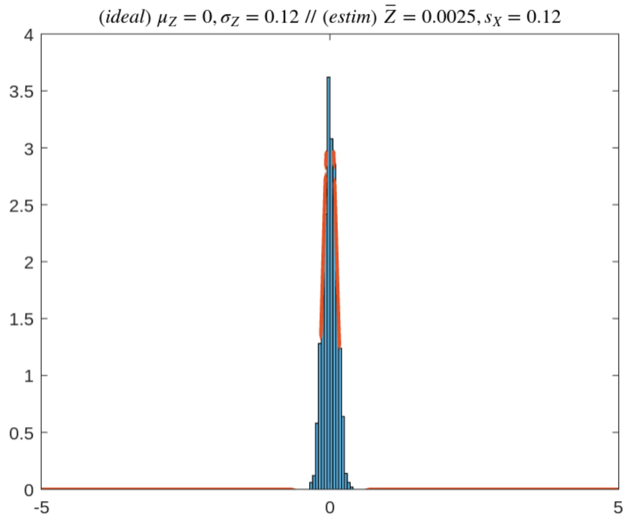


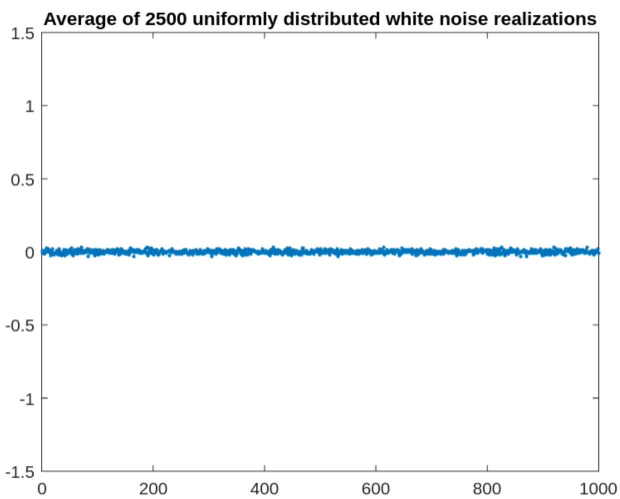
Then computing the average of N trials we can observe CLT effect, namely the random average signal is a normal distribution with the same mean of the white noise and with a lower standard deviation.

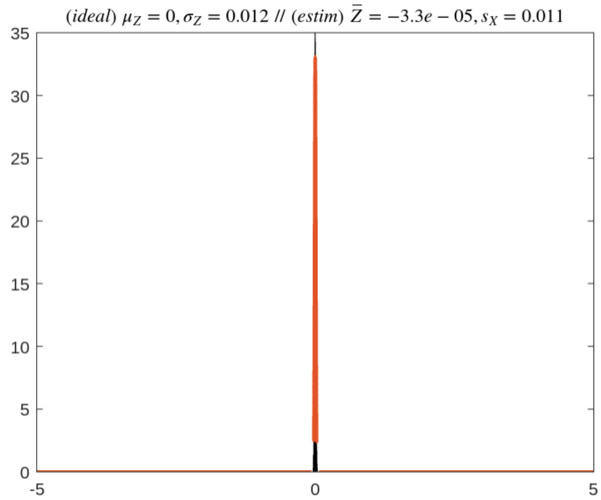

Average of 4 uniformly distributed white noise realizations


(ideal) $\mu_Z = 0, \sigma_Z = 0.29$ // (estim) $\bar{Z} = -0.0044, s_X = 0.3$


Average of 25 uniformly distributed white noise realizations

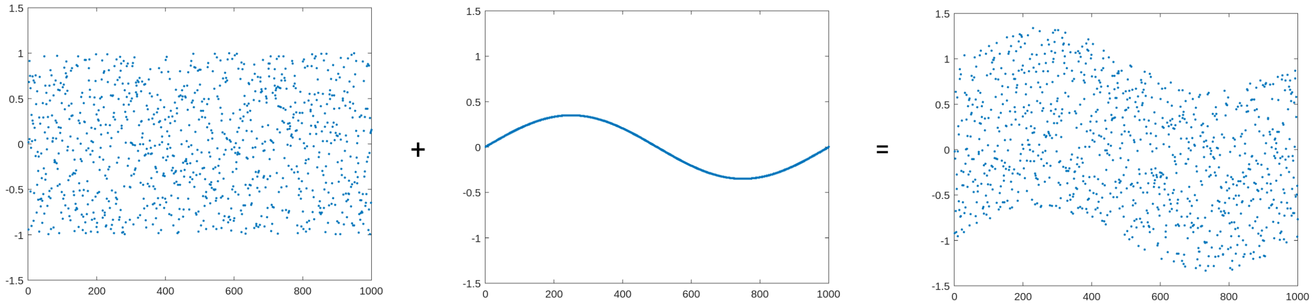

(ideal) $\mu_Z = 0, \sigma_Z = 0.12$ // (estim) $\bar{Z} = 0.0025, s_X = 0.12$


Average of 2500 uniformly distributed white noise realizations

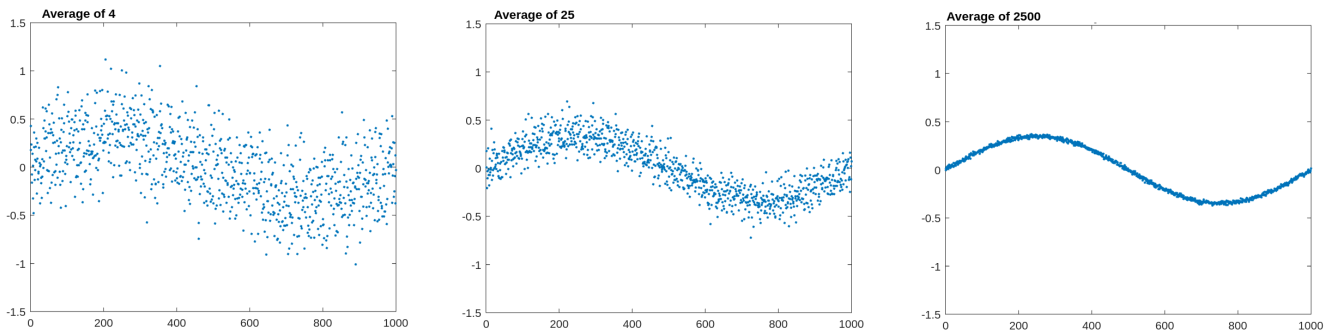

(ideal) $\mu_Z = 0, \sigma_Z = 0.012$ // (estim) $\bar{Z} = -3.3e-05, s_X = 0.011$

Example Noise + Signal

Let's consider now a trial in which we collect data affected by a white noise during the recording:



If we collect and average an high number of trials, then we are able to extract signal of interest from noisy recordings



TRUE STATEMENTS

- The Average Rectified Value (ARV) is a measure of the amplitude of a signal, and it is obtained by summing the absolute values of all samples and dividing the result by the number of samples.
- ARV_X is defined as:

$$ARV_X = \frac{1}{N}\sum_i |x_i|$$

  where the sum extends on the N samples of the signal X
- The Root Mean Square (RMS) is the square root of the average of the squared value of the samples of a signal
- The variance of a signal is estimated by summing the square of all deviations of the N sample values from the sample mean, and then dividing by (N – 1)
- s_X^2 is defined as:

$$s_X^2 = \frac{1}{N-1}\sum_i (x_i - \overline{X})^2$$

  where the sum extends over the N samples of the signal X.
- The variance σ^2 and the square of the RMS of a zero-mean signal have the same value. (Consider $N \to +\infty$)
- The standard deviation σ of a signal is the square root of its variance.
- In white noise, all samples are uncorrelated, i.e. when given the value of one sample we have no increased knowledge to predict the value of another sample.
- The frequency spectrum of white noise is flat, i.e. it has the same power at any frequency.
- In a Gaussian noise, the probability [density] that a sample has a given amplitude value follows the normal (Gaussian) distribution with zero mean.
- Given two ranges of equal width $A = [-0.5, +0.5]$ and $B = [0.5, 1.5]$, it is more likely that samples of a Gaussian noise will have amplitude in A rather than B [because the gaussian probability density function is highest around 0]
- Given two ranges of equal width $A = [-0.1, +0.1]$ and $B = [0.8, 1.0]$, it is less likely that samples of a sinewave $x = \sin(t)$ will have amplitude in A rather than B.
- The amplitude of the samples of a triangular waveform have uniform probability density function, i.e. samples have the same probability [density] of taking a value between the –A and +A (being A the peak value of the waveform) and zero probability of taking a value outside [–A, A].
- The Central Limit Theorem (CLT) states that the probability distribution of the average of N independent and identically distributed random variables tends to a normal distribution for N approaching infinity.
- The probability distribution of the average of N independent and identically distributed random variables is a normal distribution independently of the value of N.
- Given N independent and identically distributed random variables with variance (or standard deviation) equal to σ^2 (or σ), the variance (standard deviation) of their average is: σ^2/N (or $\sigma/\sqrt(N)$)

> - The synchronized average of N trials containing only spontaneous EEG whose $RMS_{trial} = \sigma^2$ is a signal $RMS_{avg} = \sigma^2/N$

05 - BASICS OF SIGNAL PROCESSING - PART 3

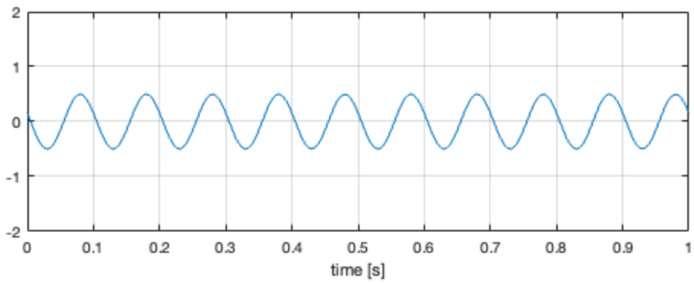
> **TOPICS**
>
> > - Fourier Transformation
> > - Use Discrete Fourier Transformation
> >   -- Zero padding and windowing
> >   -- Power Spectral Density

> **Reference**: Steven W. Smith The Scientist and Engineer's Guide to Digital Signal Processing - CHAPTER 2 - https://www.dspguide.com/pdfbook.htm

SPECTRAL ANALYSIS

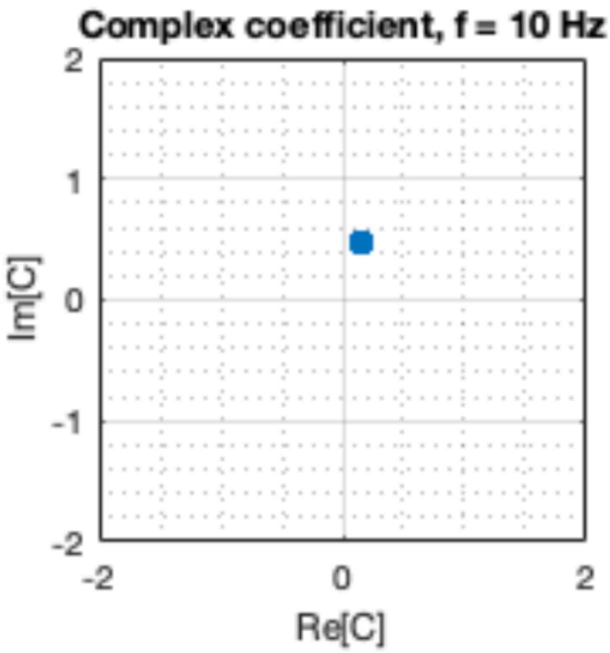
A sinewave is a function of time, with parameters: frequency f , amplitude A , and initial phase ϕ

```
sinewave = @(t, f,A,phi) A * cos(2*pi*f*t + phi);
t = (0:.005:1)'; % seconds
f = 10; % Hz
A = 0.5;
phi = 0.4 * pi; % radians
```



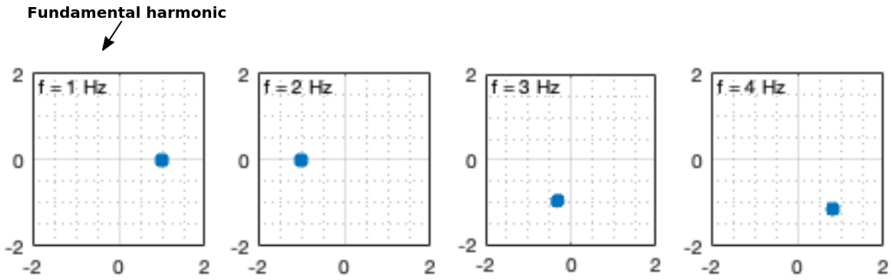
A sinewave can be represented as a complex number C:

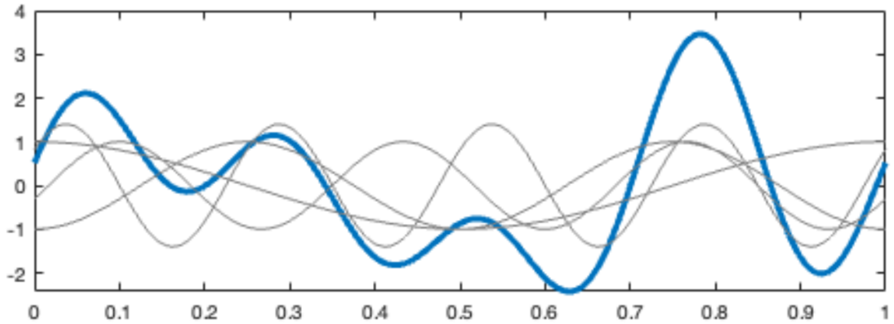
$$C = Ae^{j\phi} \Leftrightarrow A = |C|, \phi = \angle C$$



Sinewaves can be composed to create more complex waveforms:

```
f1 = 1;  % Hz
ff = f1 .* [1 2 3 4];
CC(1) = 1 * exp(1j * 0 * pi);
CC(2) = 1 * exp(1j * 1 * pi);
CC(3) = 1 * exp(1j * -0.6 * pi);
CC(4) = 1.4 * exp(1j * -0.3 * pi);

waveform = @(t) ...
    sinewave(t, ff(1), abs(CC(1)), angle(CC(1)) ) + ...
    sinewave(t, ff(2), abs(CC(2)), angle(CC(2)) ) + ...
    sinewave(t, ff(3), abs(CC(3)), angle(CC(3)) ) + ...
    sinewave(t, ff(4), abs(CC(4)), angle(CC(4)) );

figure(2)
clf
for k=1:4
    subplot(2,4,k)
    plot_complex_coefficient(CC(k), ff(k));
end

subplot(2,1,2)
plot(t, waveform(t), "Linewidth", 3)
hold on
for k=1:4
    plot(t, sinewave(t, ff(k), abs(CC(k)), angle(CC(k))), "Color", "#808080" );
end
hold off
```





.FOURIER TRANSFORMATION

Fourier Analysis, i.e. decomposition of signals into sum of sinewaves. Since each sinewave carries power at exactly one frequency, the decomposition can be used to analyze the signal in the frequency domain.

Specifically Discrete Fourier Transform can be used to transform the (time-limited and sampled) time- domain representation of a signal into its (bandwidth limited an sampled) representation in the frequency domain.

•

07 - BASICS OF SIGNAL PROCESSING - PART 5

TOPICS

-

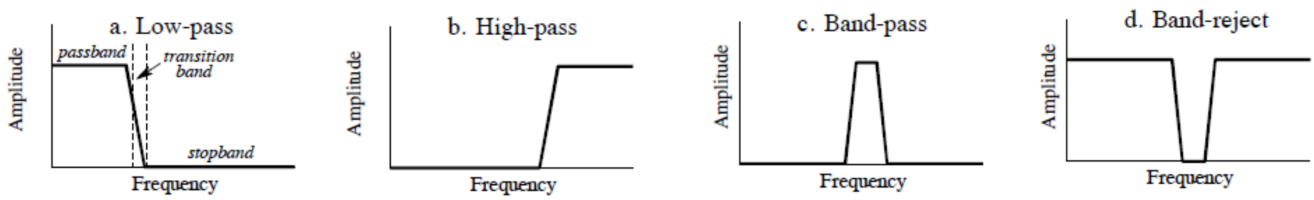
Reference: Semmlow, Biosignal and medical image processing, Chapter 4 (https://www.dspguide.com/pdfbook.htm, Chapter 14)

FILTERS

The purpose of a filter is to allow some spectral components of a signal to pass (almost) unaltered, while (almost) blocking other spectral components. For some purposes like anti-aliasing or anti-saturation filtering, analog filter are used, however in most of the cases also a digital (software) filter is used for final signal processing.

BASIC FERQUENCY RESPONSES

The figure below shows the four basic frequency responses:



We can distinguish three different regions in each frequency response:

- **Passband** refers to those frequencies that are passed (GAIN = 1)
- **Stopband** contains those frequencies that are blocked (GAIN = 0)
- **Transition band** is between passband and stopband (GAIN $\in (0, 1)$)

A **fast roll-off** means that the transition band is very narrow.

Cutoff frequency

The division between the passband and transition band is called the **cutoff frequency**.

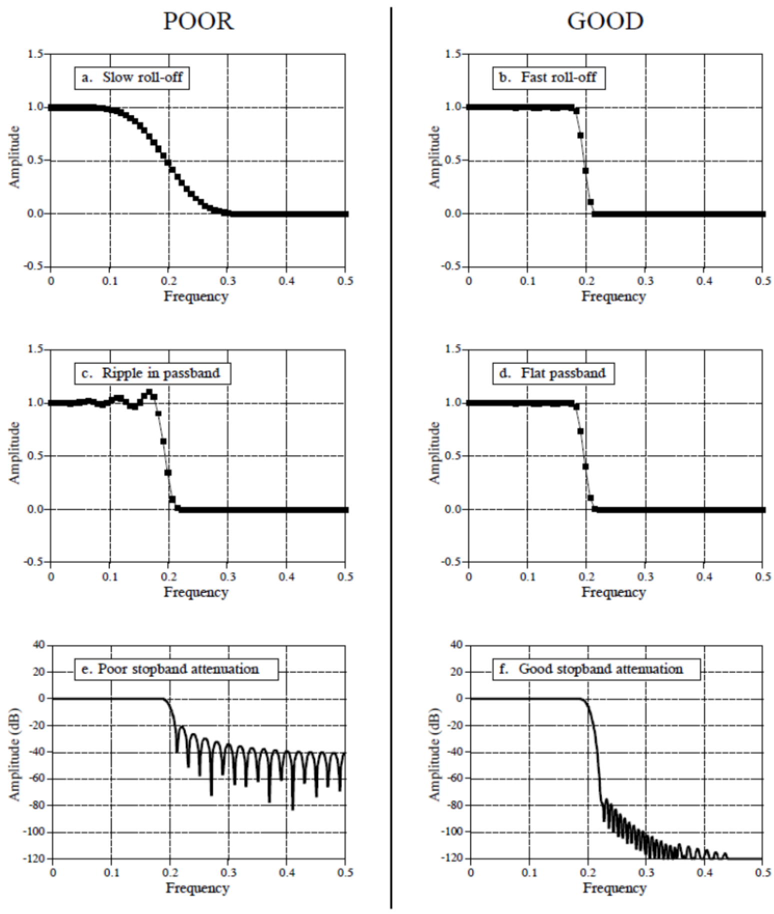
In **analog filter** design, the cutoff frequency is usually defined to be where the amplitude is reduced to 0.707 times the input (i.e. -3dB).

Digital filters often specify different attenuation at the cutoff frequency, depending on the synthesis process.

POOR-GOOD FILTERS

The figure below shows three parameters that measure how well a filter performs in the frequency domain:

- To separate closely spaced frequencies, the filter must have a **fast roll-off**, (top row).
- For the passband frequencies to move through the filter unaltered, there must be **no passband ripple**, (middle row).
- To adequately block the stopband frequencies, it is necessary to have **good stopband attenuation**, (bottom row).



> **OBS:** *(Logarithmic Y-scale)*
> Note that in the last figure (stopband attenuation) we use a logarithmic scale for amplitude axis in order to be able to appreciate small variationS (ripple)

> **NOTE:** *(Normalized X-Axis)*
> The X-axis of the previous figures use a normalized scale respect to the sampling frequency $f_{sampling}$, indeed they represent frequency response in the frequency interval $[0, f_{Nyquist}]$. (Recall $f_{Nyquist} = 0.5 f_{sampling}$)

In the real world is impossible to have a perfect filter (No free lunch), when we design a filter we have a tradeoff between three factors:

- **Ripple tolerance:** Maximum ripple amplitude admissible
- **Order:** (Complexity)

- **Roll-off slope**

In practice, only two of them can be chosen. For instance, if we want a filter with fast roll-off and low ripple tolerance, then our filter must have an high order (high design complexity).

MAIN FILTERS FAMILIES

Finite impulse response and Infinite impulse response are the two main families of filter types. The main difference between them is their response to a finite impulse.:

- Given in input a finite impulse to **FIR filter**, it responses with a **finite signal** that eventually reaches zero
- Given in input a finite impulse to **IIR filter**, it responses with a **infinite signal**

Finite Impulse Response (FIR)

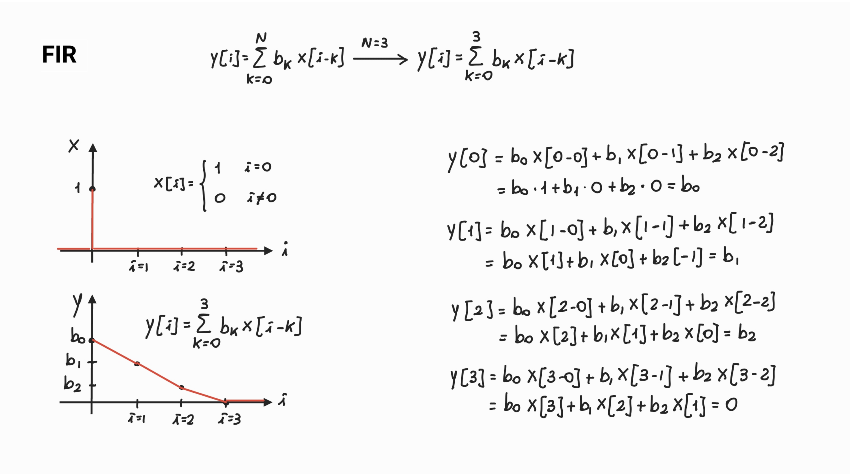
In a FIR filter, the output $y[i]$ at time discrete i is computed by combining samples of the input x. We essentially perform a convolution operation:

$$y[i] = \sum_{k=0}^{N} b_k x[i-k]$$

Where:

- b_k Convolution coefficients
- N Filter order (indicates how many recent samples are used to compute the output)

Example *(FIR)*
N = 3



We can note that y goes to zero after a while when we have an impulse x in input.

Common used FIR Filters
Examples of design methods for FIR filters:

- **Windowed sinc** (using the same window types used in spectral analysis)

- **Minimax** (or Parks-McClellan)

Infinite Impulse Response (IIR)

In a IIR filter, the output is computed by combining samples of the input x and previous samples of the output y. In this case we have a recursive formula:

$$y[i] = \sum_{k=0}^{N} b_k x[i-k] - \sum_{k=1}^{M} a_k y[i-k]$$

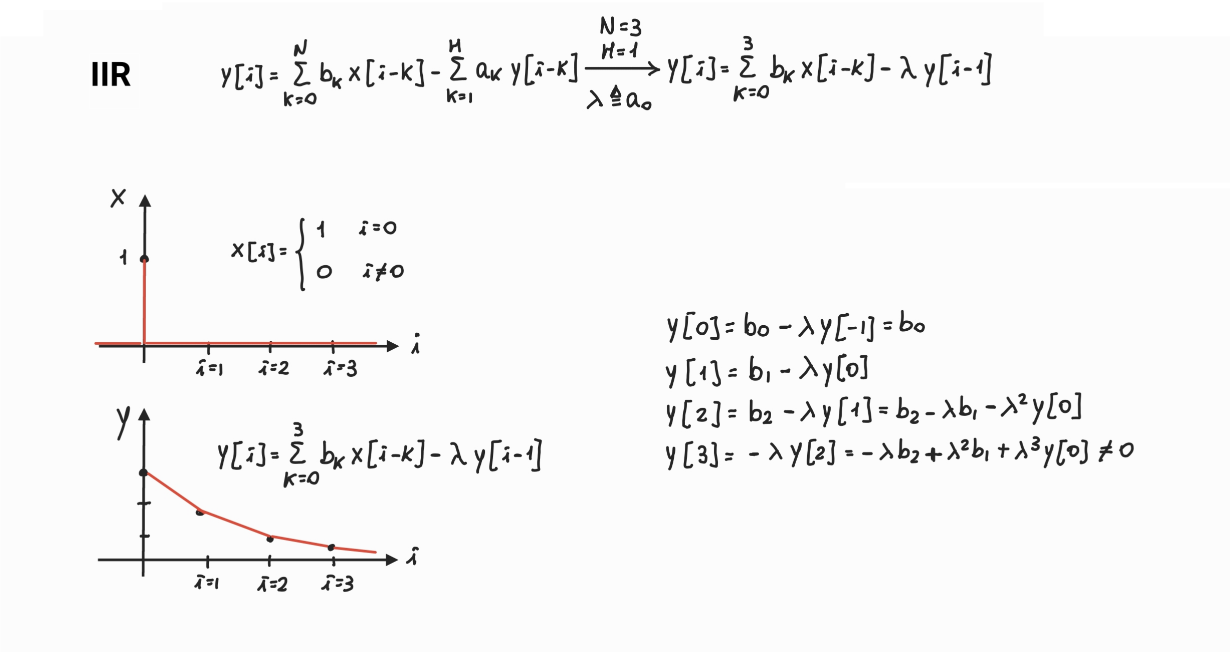
Where:

- N: Number of input samples to use
- M: Number of previous output samples to use
- b_k Convolution coefficients
- a_k Recursion coefficients

In this case the filter order is given by:

$$\text{FILTER ORDER} = \max(M, N)$$

Example *(IIR)*
N = 3, M=1



We can note that y does not reach zero when we have an impulse x in input.

Common used IIR Filters

Examples of design methods for IIR filters:

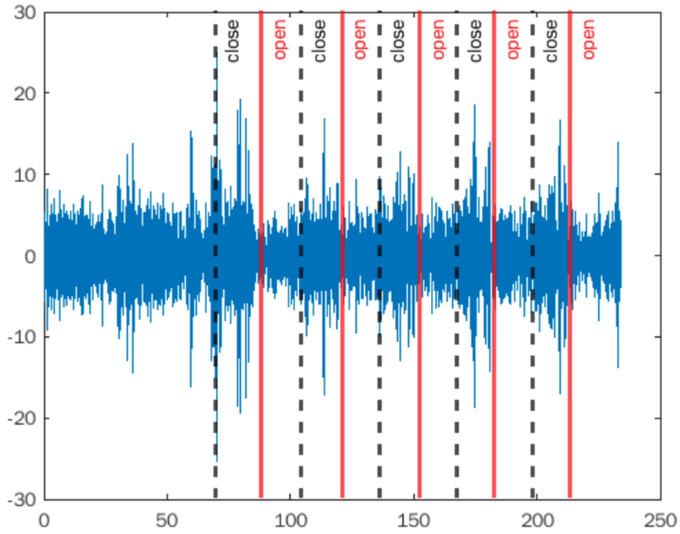
- **Butterworth**
- **Chebyshev**
  -- Type I (It has ripple in passband)
  -- Type II (It has ripple in stopband)
- **Elliptic**

FILTERS APPLIED TO EEG SIGNAL

// to do
// add some examples

EXTRACT ALPHA RHYTHM

```
load eeg.mat eeg ec_start_samp eo_start_samp
fs = 512;
num_samp_eeg = length(eeg);
time_eeg = (0:num_samp_eeg-1) / fs;
ec_trial = eeg - mean(eeg); % fix overall baseline

alpha_filt = designfilt('bandpassfir', 'FilterOrder', 256, 'CutoffFrequency1', 8,
'CutoffFrequency2', 13, 'SampleRate', fs);

alpha = filtfilt(alpha_filt, eeg); % filtfilt introduces no delay

figure(3)
clf
plot(time_eeg, alpha)

for smp = eo_start_samp(:)'
        xline(smp/fs, 'r-', "open", "LineWidth", 2);
end

for smp = ec_start_samp(:)'
        xline(smp/fs, 'k--', "close", "LineWidth", 2);
end
```



30
20
10
0
-10
-20
-30
0
50
100
150
200
250
close
open
close
open
close
open
close
open
close
open

TRUE STATEMENTS

- The purpose a filter is to allow some spectral component of a signal to pass (almost) unaltered, while (almost) blocking other spectral components
- Filters are categorized into four types depending on the basic shape of their frequency response: (i) low-pass; (ii) high pass; (iii) band-pass; (iv) band-reject (or band-stop)
- The passband is the interval of frequencies in which the gain of the filter is close to 1. In the stopband the gain is close to 0. In the transition band the gain has an intermediate value between 0 and 1.
- The roll-off of a filter is the slope of its frequency response in the transition band. It is high when the transition band is narrow.
- The cutoff frequency (or corner frequency) designates the limit of the passband. The gain of the filter at cutoff frequency is approximately 0.71 (i.e. $1/\sqrt{2}$, -3dB) [The gain value at the cutoff-frequency might be different for some filter designs, but this concept is beyond the scope of this course]
- The gain in the passband can monotonically decrease below 1 when the frequency approaches the cutoff frequency, or it might ripple above and below 1.
- The frequency response of a filter in the stopband should not be evaluated from a graph where the gain axis is in linear scale, because a gain of 0.001 can hardly be distinguished from a gain of 0.0001. Rather, a vertical axis in logarithmic scale (i.e. the gain is expressed in dB) should be used.
- Good features of a filter include: (i) fast roll-off; (ii) flat passband (i.e. no ripple); (iii) strong stopband attenuation (e.g. gain below −40dB, but the specific value may change depending on applications)
- Digital filters are categorized into two types depending on their implementation: (i) Finite Impulse Response (FIR); (ii) Infinite Inpulse Response (IIR)
- The output of FIR filters is the linear combination of samples of the input. The output of IIR filters combines both samples of the input and past samples of the output.
- The order of a filter measures the number of recent samples of the input (or the output) are combined to compute the output. [slightly incorrect, but will do for this exam. A more accurate definition that is beyond the scope of the course states that the order of a filter is the maximum delay applied to an input or output sample, whichever is largest]
- The Butterworth filter is a design method in the family of IIR
- The corner frequency of a high-pass filter is called low cutoff frequency. The corner frequency of a low-pass filter is called high cutoff frequency. Band-pass and band-stop filters have both a low cutoff frequency and a high cutoff frequency.
- A notch filter is a type of band-stop filter which removes only attenuates the input in a narrow band around the notch frequency. Its most common use is to remove the powerline artifact at 50 Hz and its harmonics (60 Hz in some other countries).
- An IIR filter is more efficient than a FIR filter, meaning that the latter needs to be of a higher order to achieve the same quality specifications.
- A FIR filter can be designed to have "linear phase", meaning that it will not introduce time-domain distortions in the waveform of the output signal. IIR filters cannot have linear phase.