HEC MONTRÉAL

Reducing the magnitude of measurement error in EEG data by modelling the cyclic behavior of electrical brain signal
par
Marin Opari

Marc Fredette
HEC Montréal
Directeur de recherche

Sciences de la gestion
(Spécialisation Science des données et analytique d’affaires)

Mémoire présenté en vue de l’obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)

December, 2024

Résumé

Cette thèse analyse une méthode visant à réduire le nombre d'essais nécessaires pour l'analyse des Potentiels Évoqués (PE), tout en maintenant des niveaux comparables de rapport signal/bruit (SNR). L'objectif est d'améliorer la validité écologique des recherches sur les Potentiels Évoqués (PE). L'étude explore la suppression des composantes périodiques dominantes de chaque essai à l'aide de la Transformée de Fourier Rapide (TFR). Dans un premier temps, les données EEG de base ont été analysées à l'aide de modèles de régression sinusoïdale, ce qui a démontré que les composantes périodiques pouvaient être capturées dans des segments de 2 à 3 secondes. Par la suite, en nous basant sur cette première analyse, nous avons recentré notre attention sur des segments de données plus longs, directement influencés par les tâches expérimentales. En appliquant la Transformée de Fourier Rapide (TFR) pour supprimer les fréquences les plus puissantes, les résultats ont indiqué que cette approche préservait les caractéristiques essentielles des Potentiels Évoqués (PE) tout en atteignant des niveaux de SNR similaires avec moins d'essais. Après avoir comparé cette analyse à la méthode de Transformation Baseline (BT), les résultats suggèrent que l'approche TFR permettait une convergence plus rapide vers le PE cible. Cela suggère une amélioration de l'efficacité de l'analyse des PE, contribuant à une validité écologique accrue.

Mots clés : Données EEG, PE, Composants Périodiques, TFR, SNR, Validité écologique

Abstract

This thesis analyses a method for reducing the number of trials required in Event-Related Potential (ERP) analysis while maintaining comparable Signal-to-Noise Ratio (SNR) levels. The goal is to enhance the ecological validity of ERP research. The study explores the removal of dominant periodic components from each trial using the Fast Fourier Transform (FFT). Initially, the baseline EEG data was analysed, using sinusoidal regression models and this demonstrated that periodic components could be captured within 2-3 second segments. Subsequently, based on the initial study, we shifted the focus to longer data segments directly influenced by experimental tasks. By applying FFT to remove the most powerful frequencies, the results indicated that this approach kept the essential features of the ERP and at the same time achieved similar SNR levels with fewer trials. After comparing this analysis with the Baseline Transformed (BT) method, the results suggest that the FFT approach achieved faster convergence to the target ERP. This suggests that the efficiency in ERP analysis is improved, contributing to increased ecological validity.

Keywords: EEG data, ERP, Periodical component, FFT, SNR, Ecological Validity

Table of Contents

List of Tables and Figures

CHAPTER 1

Introduction

Since the beginning of the civilization, human beings have always been drawn to the idea of exploring their inner world, their mind and its functioning. Ancient philosophers from the early Greeks to the eastern sages, have all been inquiring the human nature, nature of the mind and consciousness. In the West they debated the origins of knowledge and the processes of reasoning. Meanwhile, philosophers in the East, have been discussing about the nature of consciousness and the reality of the self. Despite this philosophical progress, the understanding of human cognition continued to remain mostly subjective and mystical for several more centuries. Only with the developing of modern science, the study of human cognition began to develop an empirical form.

The development of psychology as a separate science in the late 19th century represented a shift from previous philosophical speculation to an entirely new approach of empirical approach to study the mind. The technological developments of the 20th century allowed the development of new tools. As a result, the first electroencephalogram (EEG) was invented. In 1924 the first EEG recording on humans was performed by Hans Berger (Haas LF, 2003). This was followed by other type of devices designed to measure different elements of human physiology, such as: functional Magnetic Resonance Imaging (fMRI) which allows detailed brain imaging, eye-tracking technologies providing insights on visual attention, Galvanic Skin Response (GSR) sensors, which measure changes in the skin conductance related to changes in emotional arousal etc. These and other innovations have greatly developed our understanding of brain activity, emotions and behavior.

In parallel, the rapid growth in computer science and the increased use of computers in everyday life, opened new frontiers in cognitive research with emerging fields such as Human Computer Interaction (HCI) and NeuroIS.

NeuroIS represents a new field within IS research as an intersection of computer science and neuroscience (Dimoka et al., 2007). Its goal is to understand how the brain processes information during human-computer interactions. The term "NeuroIS" was first introduced in 2007 at the International Conference on Information Systems (ICIS) (Riedl and Léger, 2017). Researchers in this field use tools such as EEG, fMRI etc. to analyze how humans interact with technology. They identify mental and emotional processes that occur during this interaction. This provides feedback to design and development technology that will help to create effective and user-friendly systems.

1.1 - Electroencephalography (EEG)

Electroencephalography (EEG) is a method used to record an electrogram of the spontaneous electrical activity of the brain (Moran, 2004). It measures the voltage fluctuations coming from groups of neurons presenting the cerebral cortex. These signals are captured through electrodes which are placed on the scalp and detect the tiny electrical changes that occur as neurons communicate with each other.

The data collected by EEG provides a continuous record of brain activity, which can be analyzed to gain insights into various cognitive and neural processes. It can provide precise timing information, but its spatial resolution is quite low. This means that is relatively difficult to pinpoint the exact brain location where the recorded signal is coming from. Techniques such as fMRI are more suitable in these scenarios.

The electrodes are labeled based on their location: "F" refers to the frontal lobe, "C" the central lobe, "T" to the temporal, "P" to the parietal, and "O" to the occipital. While the odd numbers refer to the left hemisphere and even numbers to the right one. These are electrodes used in our experiment: C3, C4, F3, F4, F7, F8, FCz, Fp1, Fp2, Fpz, Fz, O1, O2, Oz, P3, P4, P7, P8 and Pz.

1.2 - EEG in Neuroscience

In the research domain EEG finds applications in studies of cognitive processes such as attention, memory, language etc. It allows to investigate how the brain responds to specific stimuli, how these responses change over time and under different conditions. It also has a wide application in clinical studies such as: diagnosing and monitoring epilepsy, sleep studies, assessing patients in coma or with brain damages etc.

EEG is able to detect and measure electrical rhythms coming from the brain known as brain waves. These are result of a synchronous neurons' communication and are categorized in different frequency bands in the literature (Abhang et al., 2016). This activity is measured in hertz or number of oscillations per second. Each band has is characterized by specific state and cognitive functions as we will describe below:

- **Delta Waves (0.5 - 4 Hz):** state of deep sleep or unconsciousness.
- **Theta Waves (4-8 Hz):** state of relaxation, early stages of sleep and meditative states.
- **Alpha Waves (8-12 Hz):** relaxed states, periods of reflection and daydreaming.
- **Beta Waves (13-30 Hz)**: active thinking, anxiety, external attention, engaged in a mentally demanding task.
- **Gamma Waves (30-60Hz):** high-level cognitive functions such as: perception, concentration and learning.

1.3 - EEG in NeuroIS

NeuroIS integrates neuroscience and information systems while helping to understand how humans interact with technology. EEG plays an important role in this context and helps to explore our cognitive load, emotional responses etc. For example, increase in beta wave activity might indicate that a user is concentrating on a demanding task. These findings can be used to design systems that are more aligned with the participant responses, improving the user experiences.

1.4 - Event-Related Potential (ERP)

Event-Related Potential (ERP) is the measured brain response that is the direct result of a specific sensory, cognitive, or motor event (Luck SJ, 2005). ERPs are calculated by averaging the EEG signal coming from a repeated event. This allows researchers to isolate the brain activity specific to that event from the background noise or other signal.

1.5 - ERP in Neuroscience and NeuroIS

In cognitive neuroscience, ERPs have been used to identify different stages of cognitive processing. For example, the P300 component that happens around 300 milliseconds after the stimulus, is associated with cognitive information processing (e.g. memory, attention, executive function) (van Dinteren R et al., 2014), N400 component is linked to the integration of meaning including visual and auditory words etc.(Kutas M, Federmeier KD, 2000), (Kutas M, Federmeier KD, 2011). ERPs contribute also in understanding the neural background activity of disorders such as schizophrenia, ADHD etc.

ERPs can be very useful in NeuroIS to study human responses when facing a specific event. For example, how quickly and efficiently users process information presented by a pop-up notification or a security warning.

1.6 - Problems Associated with ERP

Although ERPs are very effective because of their temporal precision, they can be affected by several issues. One of the main problems that can be faced in ERP research is the large number of trials needed to obtain an optimal “signal-to-noise ratio” (SNR), a measure reflecting the ability to distinguish signals from noise (Gratton 2007). Research designs based on the aggregation of a large number of trials are often inappropriate for NeuroIS research because this can compromise the ecological validity of the experiment (Fredette et al., 2015).

The need for many trials can be problematic for several reasons:

- Participant attention: Presenting the same stimulus repeatedly can lead to fatigue or loss of interest in the task. This can affect the cognitive response and the data reliability.
- Time constraints: This can happen with children or elders which are prone to lose attention during long tasks.

These or other reasons may end up compromising the authenticity of the task. For instance, in a NeuroIS study where the user reaction to a popup is studied, their reaction to increased number of tasks might not reflect a real-world scenario. This can affect the ecological validity of the experiment which we will discuss in the following section.

1.7 - Ecological Validity

The concept of ecological validity was first introduced by psychologist Egon Brunswik (Kieffer, 2017). He argued that to understand human behavior, researchers must study it in context, in environments that closely resemble the settings in which the behavior naturally occurs.

It evaluates if the research's conditions resemble a real-world scenario where the event is usually occurring. This can determine if the findings of the research can be generalized outside of the laboratory. In case ecological validity is infringed, the findings may be a limited reflection of the real world.

In neuroscience or NeuroIS environment, the artificial manner in which trials are repeated can make the participants lose their concentration, thing that could affect the quality of the data and ERP's components. In this context researchers should aim to find balance between achieving successful results while making sure that these results are true not only to the specific conditions where the experiment takes place, but also be able to generalize with other subjects and environments.

1.8 - Aim of the study

Building on Fredette et al. (2015), this study aims to provide guidelines and develop new statistical methodologies that will be constructed explicitly for the analysis of EEG data in NeuroIS research and that will attempt to improve the measurement of ERPs.

A single EEG trial segment of a particular subject on a single electrode can be expressed as:

$$\text{EEG Signal} = \text{Event-Related Activity} + \text{Periodic Activity} + \text{Noise}$$

It is a composition of the non-periodic activity (event-related activity) which we want to study, periodic activity and general noise. By modeling and quantifying the periodic component of the signal, we aim to reduce the number of trials needed for suitable signal-to-noise ratio (SNR) while enhancing the ecological validity of the experiment. By reducing the participants' cognitive load that comes from large number of task repetition, we can observe more natural user reactions, leading to insights that can inform the development of more effective information systems.

1.9 - Research Questions and Methodology

This study introduces a new approach which aims to improving the analysis of Event-Related Potentials (ERPs) by modeling and removing periodic components from EEG signals. Its primary goal is to enhance the signal-to-noise ratio (SNR) while reducing the number of trials needed. This will help improving the ecological validity of ERP studies. Below, we describe the foundational assumptions, the proposed methodology, and the key research questions that this study tries to answer.

1.9.1 - Assumptions and Signal Decomposition

For simplicity, and without loss of generality, we assume a scenario involving a single subject and a single electrode. The observed EEG signal at any given time t is decomposed as follows:

$$S(t) = \begin{cases} P(t) + U(t - t_j) + \epsilon(t) & \text{for } t \in [t_j, t_j + 1] \\ P(t) + \epsilon(t) & \text{for } t \in [t_j - 2, t_j] \end{cases}$$

Where:

- **S(t):** observed raw EEG signal at time *t*.
- **t_j:** time of the j^{th} marker.
- **P(t):** the periodic component of the signal.
- **U(t):** the signal induced at the marker.
- **ϵ(t):** random noise.

EEG segments of 3 second around the marker will be taken into consideration. Specifically, from 2 seconds before to 1 second after the event.

Objective: Estimate the event-related signal *U(a)*, where a ∈ [0;1].

1.9.2 - Baseline Approach

The current approach, the baseline approach for estimating $\hat{U}_k(a)$ involves averaging the signal segments over the first k markers, where $k \leq k_{max}$. The mean value of the timepoints from segment [-0.2; 0] is subtracted from every timepoint in [0; 1] segment. In this case the ERP estimation is given by:

$$\hat{U}_k(a) = \frac{1}{k} \sum_{j=1}^{k} \left[S(t_j + a) - \bar{S}(t_j) \right]$$

Where:

- $\hat{U}_k(a)$: the ERP estimate from the baseline approach.
- $\bar{S}(t_j)$: the average of *S(t)* between $[t_j - 0.2, t_j]$.

1.9.3 - New Approach Using Periodicity

This study proposes a new approach that explicitly models and removes the periodic component from the signal. It is expressed as:

$$\tilde{U}_k^P(a) = \frac{1}{k} \sum_{j=1}^{k} \left[S(t_j, t_a) - \hat{P}(t_j, t_a) \right]$$

Where:

- $\tilde{U}_k^P(a)$ **:** ERP estimate from the new method.
- $\hat{P}(t_j, t_a)$: estimated periodic function from the EEG signal within the interval $[t_{j-1}, t_j]$.

1.9.4 - Research Questions

This study tries to answer these two questions:

1. Convergence of Estimates with Increasing Trials:

Question: As the number of trials k increases (towards k_{max} = 80), do the estimates $\hat{U}_k(a)$ and $\tilde{U}_k^P(a)$ converge towards the same function?

Methodology: The convergence will be evaluated by calculating the Kullback-Leibler (KL) divergence between both functions as k increases.

2. Rate of Convergence:

Question: Which estimate, $\hat{U}_k(a)$ or $\tilde{U}_k^P(a)$, converges more rapidly as the number of trials increases?

Methodology: This involves studying the decrease in the KL divergence between $\hat{U}_k(a)$ and $\hat{U}_{k_{\max}}(a)$ and comparing it to the decrease in divergence between $\tilde{U}_k^P(a)$ vs $\tilde{U}_{k_{\max}}^P(a)$ as k increases.

Throughout this study, we will introduce the dataset used and attempt to answer these two research questions. It is important to note that a single dataset will be used. This suggests that the conclusions do not represent definite answers to these questions. Instead they could rather be interpreted as preliminary findings that could be further expanded by using a larger number of datasets.

CHAPTER 2

Literature Review

Over the years a variety of methods have been developed to analyze EEG data. They find applications in fields such as in clinical diagnosis, neuroscience, Brain-Computer Interaction (BCI), NeuroIS and more. Despite their diverse perspectives, they all use similar tools to analyze and understand the workings on the mind. A large portion of these methods have been used to analyze the periodicity of brainwaves, a very important field which offers valuable insights no matter of the field that it has been applied.

A thorough review of the existing methods has been conducted as part of this study with the goal of understand them better chose the most appropriate ones to help us with our analysis.

In the following sections we will discuss about these techniques, starting from the most traditional ones such as Fast Fourier Transform (FFT) and Wavelet Transform, and then proceeding with most advanced methods of Machine Learning. For every method we will briefly present their main characteristics, explore its application in different fields and asses some of the benefits and limitations of applying them in EEG data and their periodicity.

2.1 - Fourier Transform

The Fourier Transform (FT) is a mathematical tool which transforms signals from time domain to frequency domain (Arfken, George (1985)). The main idea is that any time-domain signal can be expressed as a sum of sinusoidal functions. Each function is defined by its frequency, amplitude and phase.

Discrete Fourier Transform (DFT) is the FT version which is applied for sampled signals which make it a very important method in signal processing including EEG data. In this case any signal can be represented as a sum of finite number of sinusoidal functions. It is expressed as:

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j\frac{2\pi}{N}kn}$$

Where:

- **X[k]**: the frequency domain representation of the signal,
- **x[n]**: the time-domain signal,
- **N**: the number of samples,
- **k**: the index of each frequency component,
- **j**: the imaginary unit.

DFT is known to be computationally expensive. Therefore, a more efficient form is used in practice, known as Fast Fourier Transform (FFT). It was developed by Cooley and Tukey in 1965 and is able to reduces the computational time extensively. This makes it suitable to be applied in real-time scenarios as well as in large datasets environments, including EEG data.

FFT can be very helpful in the EEG data analysis context specifically in Spectral analysis where it helps identifying specific frequencies and quantify their power despite the complexity of the EEG data. It is used in identification of periodic activity of EEG data. This can help in diagnosing neurological conditions, studying cognitive processes etc. FFT is used to detect and remove artifacts which are known to contaminate EEG data such as: power line interreference, eye blinking, muscle movements etc. It can support training Machine Learning algorithms and help them improve their performance.

2.1.1- FFT Applications in EEG Data

- **Epilepsy diagnosis and monitoring:** By identifying dominant signal frequencies FFT helps detecting abnormal brain patterns known to characterize epileptic seizures. This can also help localize the source zone in the brain which can further. In Larsson PG et al., (2012), FFT and other methods were used to estimate alpha frequency in EEG and to differentiate between patients with epilepsy and not.
- **Sleep studies:** FFT plays an important role in research related to sleep studies. It is known that different sleep stages such as REM or NREM are characterized by specific dominant frequencies. FFT helps to classify these stages. It is also used in the field of sleep disorders to identify pattern abnormalities related to these conditions. In Nishida et al., (2004) FFT analysis showed that the anterior cingulate cortex (ACC) showed steady theta waves during wakefulness and REM sleep but not during deep sleep.

- **Brain-Computer Interfaces (BCIs):** BCIs use EEG real-time data interpretation to control external devices. In this context FFT helps detecting the user intentions by providing the frequencies in EEG data. Some applications in this domain are controlling prosthetic limbs, interacting with communication devices by selecting letters, words through intentions or mental states etc.

2.1.2- FFT Advantages and Limitations in EEG studies

Advantages:

- **Computational Efficiency:** FFT makes possible analysis of large datasets which is common in EEG data analysis. It also allows real-time processing which is very important for real-time brain monitoring or BCIs as mentioned above.
- **Wide Applicability:** It finds a wide range of applications.

Limitations:

- **Stationarity assumption:** The main drawback of FFT is that it assumes that the signals are stationary, which means their statistical properties do not change over time. EEG data are known to be non-stationary, which makes FFT application problematic especially for long segments.
- **No time domain information:** Although it has many advantages in frequency domains, it provides no information about the time when these frequencies occur. This makes its problematic in cases where the time domain is important in the analysis.
- **Fixed frequencies:** The frequencies used in FFT calculations are determined by the method based on the length of the signal. This creates problems when a specific frequency present in the signal does not align with the frequencies used by the method. In this case its power will be misinterpreted as belonging to the closes chosen frequency. This can be problematic especially in studies that are sensitive to the frequency precision.

2.2 - Short Time Fourier Transform (STFT)

STFT is a signal processing method that was developed to address the FFT limitations in time domain information and non-stationary signals. STFT allows for a time-frequency analysis. It divides the signal in smaller segments which overlap with each other. Each segment is multiplied with a window function and FFT is calculated for each for each windowed segment. The window signal than is slid in small increments until all the original segment is covered. The magnitude results are then represented in a spectrogram where STFT magnitude are squared. This helps us understand how the signal frequencies change over time.

The results are affected by the type of the window function and its length. There are several windowing functions that can be used in STFT such as rectangular window, Hamming, gaussian etc. There is a trade-off in time and frequency resolution where wider windows provide low time and high frequency resolutions and narrow windows provide poorer frequency but high time resolutions.

As previously mentioned, STFT can be useful in EEG data by allowing for time-frequency analysis. This characteristic can be used to detect and localize transition events. It can be used for analyzing periodicity in EEG data and artifact removal. STFT has a wide field of applications including clinical diagnostics and neuroscience research.

2.2.1- STFT applications in EEG data

Similarly to FFT, STFT is commonly used in epilepsy detecting, sleep studies and sleep disorder detection, cognitive neuroscience, BCIs etc.

2.2.2 - STFT Advantages and Limitations of in EEG Studies

Advantages:

- **Time-frequency representation:** Allowing researchers to analyze how signal frequencies evolve over time.

- **Flexibility:** Allowing the choice of different window types and length, allows to adapt their characteristics to the needs of the study.

Limitations:

- **Time-Frequency Trade-Off:** The time-frequency trade-off previously described, can become a limit in studies which demand for high frequency and time resolutions.
- **Windowing effect:** Unproper choices in window type and length can highly affect the analysis and results. For example, this can lead to spectral leakage, where the power of a specific frequency is calculated or is said to leak in the adjacent frequencies while performing the calculations.
- **Computational complexity:** It can be computational complex especially for large EEG datasets.
- **Rapidly changing signals:** Its accuracy can be limited in cases where signals change rapidly in time and frequency.

2.3 - Wavelet Transform (WT)

Wavelet Transform (WT) is a mathematical method very popular in signal processing (Meyer, Yves (1992)) and EEG analysis. It offers a solution to the FFT approach where the signal is decomposed in sinusoidal functions with fixed frequencies. Instead WT uses wavelet functions which are variable form and offer time and frequency variability. This characteristic makes them useful when used with non-stationarity signals such as EEG data which characteristics tend to change over time. Its nature makes it useful to localize both global and local trends within data.

It can be applied with different wavelet forms depending on the domain and the needs of the study, making it very versatile. There are several wavelets used in neuroscience with Morlet wavelet being one of the most popular ones. It is useful to help detecting oscillatory components such as alpha, beta etc. It is also used in analyzing transient events such as ERPs. Figure 2.1 shows an application of Wavelet Transform using Morelt wavelet. We can see how the dominant frequencies in red, evolve over time.

In the EEG analysis context, it is also used for signal denoising, removing noise and artifacts while keeping the needed data. It is also used for feature extractions which results can be than used for tasks like classification or prediction.

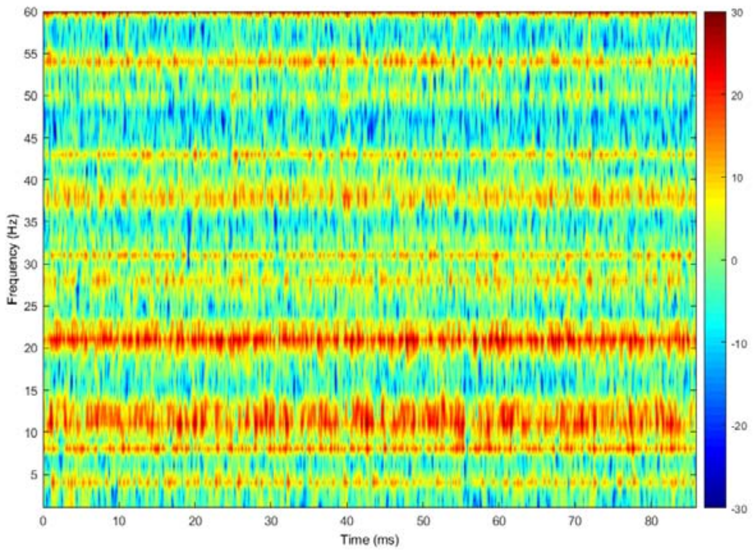


Figure 2. 1 – Wavelet Transform example

It is faster than STFT as uses fewer parameters. It also provides higher resolution in the time domain when compared to STFT.

2.3.1 - WT applications in EEG data

- **Epileptic seizure detection and prediction:** Can be more useful than FFT to detect transiting spikes or sharp waves which are characteristic of epileptic seizure. It can be used to predict a future seizure, by detecting patterns that are common to happen before, and localizing its source.
- **Sleep stages classification and disorders:** Its advantage over FFT in this domain is that it can help in analyzing the events in time-frequency domains.
- **Cognitive neuroscience:** It is used to analyze ERPs responses to cognitive of sensory events. It can help in detecting frequency bands which are occurring in encoding and recalling information from the brain.

- **Neurofeedback:** Analyzing EEG data with WT, can help professionals detect patterns that are associated with disorders such as ADHD, anxiety, depression etc. In Lee et al., (2010) WT was used to analyze EEG data from children with ADHD. The processed wavelet features were used as an input for clustering algorithm. This helped to distinguish between children with ADHD and the control.

2.3.2 - WT Advantages and Limitations in EEG studies

Advantages:

- **Time-frequency localization:** As previously mentioned, this is one of its most important advantages when compared to FFT and frequency domain methods. It helps us localize the point in time when the event happens.
- **Non-stationarity signals:** It well-studied for feature extraction and analysis in non-stationary data.
- **Flexibility in wavelet choices:** A tailored wavelet can be used for specific needs depending on the study. For example, Morelt Wavelet is very used in studies that include periodic activity.

Limitations:

- **Selection of Mother Wavelet:** It is of crucial importance to select a proper mother wavelet for the specific study. Its selection can significantly affect the analysis and the results of the study.
- **Computational Complexity:** As previously mentioned it provides more information, such as in time domain data. But this comes with a trade-off in computation complexity especially for large datasets or real-time analysis.
- **Interpretability:** Although it provides very important information, it is usually difficult to interpret and demands for knowledge in signal processing and physiological processes.

2.4 - Machine Learning in EEG analysis

Although the discussed methods are still widely used, the evolution of Machine Learning (ML) and the significant increase in the computational power during the last decades, have opened new opportunities in EEG analysis. As we are going to see through this section, traditional methods such as FFT, STFT, WT etc. are commonly used as feature extraction tools with their results being fed to ML algorithms. These algorithms will be trained on these data and will generate outputs which can be classifications, predictions etc. The EEG analysis performed by ML process is illustrated in the following picture:

Figure 2. 2 – Steps of EEG Analysis in Machine Learning

There are two major groups of ML algorithms which find application in EEG analysis context, supervised and unsupervised learning (Hosseini et al., 2020). We will continue by seeing how these algorithms have been applied in studies which include EEG analysis.

2.4.1 - Supervised Learning

The supervised learning models are trained with labeled data, where the input data is associated with output labels. By doing so these models can learn the association between these two components and be able to predict the labels in new data. The following section will demonstrate some supervised learning applications where these models have been used in EEG analysis context:

- **Logistic Regression:** In (Rajaguru et al., 2017) logistic regression was used to classify epilepsy risk levels in 20 patients based on EEG data recordings. As a first step the dimensionality of the data was reduced by applying Independent Component Analysis.

- **Linear Regression:** Dong et al., 2013 used linear regression to model visual attention. The data was collected while participants performed the Stroop task.

- **Auto-Regression:** Hu et al., (2016) developed an EEG-based authentication system using auto-regression coefficients to extract unique individual features from the participants' EEG recordings. This was achieved by analyzing ERPs generated as a response of visual stimuli to personal and non-personal photographs.

- **Support Vector Machine (SVM):** Jalilifard et al., (2016) used a single electrode recording to classify emotional states based on EEG recordings. 19 patients had to watch different video recordings and asses their emotional state. The study was focused on the emotions of fear and relaxation. The data preprocessing was performed through Wavelet Transform (WT) and STFT was used to obtain the frequency bands. SVM and KNN were used as classifiers.

- **Neural Networks:** Pfurtscheller and C. Neuper (2001) used neural networks in order to classify and differentiate between motor imagery tasks used for brain-computer interface (BIC) systems. The system is able to interpret EEG patterns associated with imagined movements of limbs.

- **Naive Bayes:** In this study (Mumtaz et al., 2018) Naive Bayesian, SVM and Logistic Regression were used as classifiers to distinguish between participants suffering from Major Depressive Disorder (MDD) and healthy controls.

- **Random Forest:** Vijaykumar et al., (2017) aimed to develop a ML approach to quantify levels of pain based on EEG recordings. The study used Continuous Wavelet Transform (CWT) to obtain time-frequency representation of EEG data. Random forest was used as a classifier for pain levels on low, medium and high categories.

- **Ensemble methods:** Hosseini et al., 2018 used ensemble methods, which is a combination of different ML algorithms, to detect epileptic seizers in real-time. This method included SVM, Multilayer Perceptron Neural Network (MLP) and K-Nearest Neighbor (KNN) algorithms.

2.4.2 - Unsupervised Learning

Unsupervised learning is the case when the algorithm builds a pattern of recognition from a data set containing only inputs with no set outputs (Hosseini et al., 2020). Asanza et al., (2016) used KNN for clustering EEG signals coming from the occipital region of the brain. Two electrodes were placed on the left and right occipital areas of the scalp. The algorithm was able to cluster 2 different frequency ranges: 5 – 9 Hz and 24 – 29 Hz.

As we noticed in many cases, several methods were used to conduct the studies. This makes it possible to take advantage of the unique characteristics that each method possesses (Hosseini et al., 2020) and provides highly accurate classification by relying on existing methods (Hosseini et al., 2013), (Dahne et al., 2014).

CHAPTER 3

Sinusoidal Regression

This chapter will start by presenting the data used in the study. Working with EEG data is usually difficult due to their nature which is usually prone to various types of noise and artifacts. Thus, knowing the source and the processing stages that these data have passed through becomes of a crucial importance. This enables us to make proper interpretation of the results of our analysis and also ensures the reliability of the study.

The following subsections will present the data source, the preprocessing procedures and finally the characteristics of the dataset.

3.1 - Data Source

The data comes from an experiment conducted by Tech3Lab and has passed all necessary ethical steps at the time of the experiment. The experiment is part of larger experiment and the data used belong to the first task. This was a N-back task, which is a continuous performance task commonly used in psychology and neuroscience (Gazzaniga et al., 2009). This type of experiment was first introduced by Wayne Kirchner in 1958 (Kirchner, W. K., 1958). The subject is presented with a sequence of stimuli and the aim is to indicate when the current stimulus matches the one from n steps earlier in the sequence.

This specific task has four levels of difficulty, specifically n=1,2,3 and 4. For n=1 the participant has to determine if the current letter is the same as the previous one. For n=2 decide if the letter shown is the same as two letters before. The same logic is followed for n=3 and n=4. As the level of difficulty increases, for n going from 1 to 4, the participant requires more cognitive load to perform the task. 25 participants took part in this task.

3.2 - Data Preprocessing

EEG data underwent several preprocessing steps to ensure its quality and usability. The following steps were essential to clean the raw data and prepare it for further analysis:

- **Filtering:** Two type of filters were applied to the data coming from this task. A notch filter to remove 60 Hz environmental noise characteristic of the electrical signal in North America. A bandpass filter which filters signal smaller than 0.5 Hz and larger than 100 Hz.
- **Rejection Criteria:** Electrodes were rejected as defective if 10% or more of the signal satisfies the following criteria: the intensity exceeds a potential difference of 100uV/200ms or if it is less than 0.5uV/100ms. Electrodes F7 and F8 were those with highest number of rejections respectively in 20 and 8 participants.

These are the only two modifications that have been performed to the raw EEG data.

3.3 - Data Characteristics and Selection

The data was recorded with a sampling rate of 256 Hz. There are 3 files for each of the 25 participants. The first file containing the EEG data for each electrode, the second the header with metadata information and the third one, the markers information. Each participant spent approximately 30 minutes on the task. The first 90 seconds of the recording were used as baseline where the participant was not involved in any task.

Different data segments were analyzed for different part of the study. For the first part (Chapter 3), the analysis is focused on data from the baseline section. Segments of different length with different starting point originating from different electrodes, were selected for this analysis. Selection criteria and detailed descriptions are provided in the relevant sections.

For the other parts of the study, we will use the data coming from Pz electrode, as it is located in the parietal lobe and is very sensitive to ERPs. It provides cleaner signal when compared to other electrodes as it is located away from areas prone to muscle artifacts. We will analyze the data coming from xTar (show target), which is what the study wants to analyze.

More precisely, model comparison (Chapter4), is performed on two types of segments. The first starting 2 seconds before the xTar event on electrode Pz and ending 1 second before it, and the second type starting 1 second before this event and ending at the event's timepoint.

For the last two parts of the study (Chapters 5 and 6), the analysis is performed on segments that include the xTar event timepoint on electrode Pz. More precisely: a 1-second segment beginning at the event's timepoint and ending one second after it, a 2-second segment starting 1 second before the event and ending 1 second after it, and a 3-second segment starting 2 seconds before the event and ending 1 second after it. Data segmentation and selection are described in detail in the relevant sections.

3.4 - Periodicity Modeling

EEG data are known for their complex nature which includes both periodic and aperiodic components. In this section we try to quantify the periodicity in these data. It is important at this point to analyze the data that is not affected by any tasks related activities. This will be achieved by analyzing the 90 seconds that was used as a baseline signal at the beginning of the recordings for each participant.

Our approach will take into consideration different segments with different signal length, starting at different points of the data for different subjects. This approach will allow us to generalize of our analysis and will offer valuable insights into the rhythmic properties of brain activity.

3.5 - Statistical Models

3.5.1 - Sinusoidal Regression

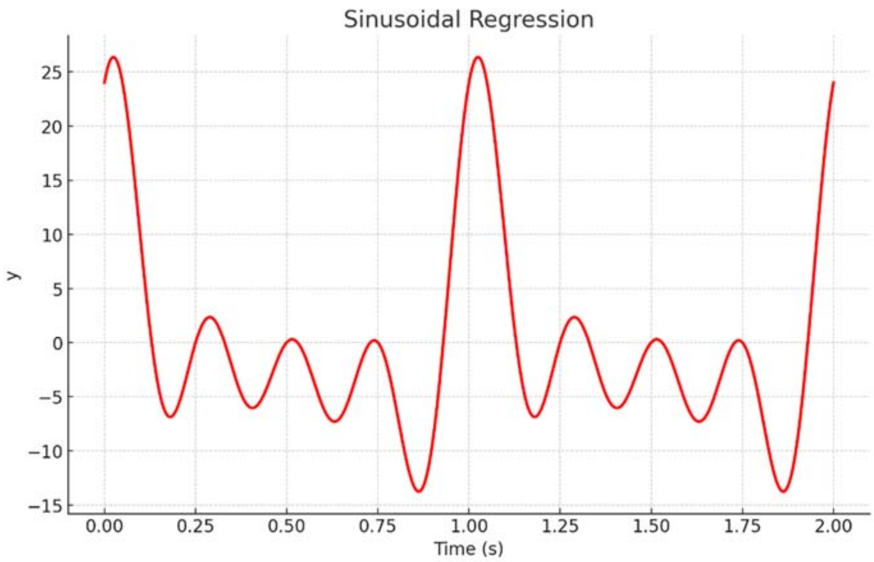
Sinusoidal regression is a form of harmonic analysis used to model periodic phenomena. It can be useful in EEG data to help us quantify its periodic component. The model that we will take into consideration is:

$$y = a + b\sin(2\pi ft) + c\cos(2\pi ft) + \epsilon$$

where:

- **y:** the dependent variable representing the observed data.
- **a:** the mean value or vertical shift of the sinusoidal function.
- **b and c:** coefficients that scale the sine and cosine components, respectively.
- **f:** the frequency of the sinusoidal wave.
- **t:** the time variable or independent variable.
- **ϵ:** the error term, representing the random noise and unexplained variability in the data.

The Figure 3.1 shows how this regression will look like for frequencies from 1 to 4, a=0, b=3 and c=6. of the application of the sinusoidal regressions with different parameters. We can clearly see how the number of oscillations within a second increases as the frequency increases.



$$y = \sum_{f=1}^{4} (3\sin(2\pi ft) + 6\cos(2\pi ft))$$

Figure 3. 1 – Sinusoidal Regression example

The first model that will be used in this part of the analysis, is a sinusoidal model with 121 parameters. The intercept, 60 sin and 60 cosine parameters for frequencies starting from 1 Hz to 60 Hz with an increment of 1 in each step (ex: 1 Hz, 2 Hz, 3 Hz, …, 60 Hz):

$$y = a + \sum_{k=1}^{60} \left(b_k \sin(2\pi kt) + c_k \cos(2\pi kt)\right) + \epsilon$$

3.5.2 - AIC Stepwise

The Akaike Information Criterion (AIC) is used extensively as a metric for model selection. When dealing with complex models with a large number of parameters. This helps to achieve a balance between the good fit and the model complexity. AIC selects a model that fits the data well and at the same time avoids over-fitting.

AIC provides also a criterion to compare models with different numbers of parameters by penalizing the inclusion of too many parameters. Lower AIC values indicate a better model.

AIC is calculated as:

$$\mathrm{AIC} = 2k - 2\ln(L)$$

Where:

- **k:** the number of parameters, in our model $k \leq 121$.
- **L:** Maximum likelihood of the model.

The process of selection can be performed in many different ways, we will use three here. In each case, amongst all the models fitted, the model with the smallest AIC is then selected:

- **Forward Selection:** Starting with no predictors. Adding one predictor at a time and calculate AIC for each case. The model with lowest value of AIC is selected in each step. Continue this process until no addition is possible.
- **Backward Selection:** Starting with the full model, with all predictors. Removing one predictor at a time and calculate AIC in each case. The model with lowest AIC is selected in each step. Continue this process until no reduction is possible.
- **Bidirectional Stepwise Selection:** Iterative combination of both methods. After adding a predictor, the method checks if any of the current predictors can be removed. The process continues until no predictors can be added or removed, based on their AIC value.

In this study we are using the bidirectional selection method.

3.5.3 - BIC Stepwise

The Bayesian Information Criterion (BIC) is another statistical measure that is used for model selection. It is similar to AIC but uses a different penalty for the number of parameters. The models are selected based on lower values of BIC. It is calculated as:

$$\mathrm{BIC} = k \ln(n) - 2 \ln(L)$$

Where:

- **k:** the number of parameters, in our model $k \leq 121$
- **n:** the number of observations,
- **L:** Maximum likelihood of the model given the data.

BIC tends to penalize an increment in sample size. Its value increases with the increment of the sample size. It tends to favor simpler models, in comparison to AIC, especially with larger datasets.

Due to the stronger penalty that this measure adds upon models with higher complexity, it tends to select models with fewer parameters when compared to those selected from AIC. Thus, the selected models in this case, are less prone to overfitting comparing to the ones selected with AIC criterion.

The selection process of the parameters follows the same procedure as described in the AIC case above. In this study we are using the bidirectional stepwise selection method.

3.6 - Methodology

This section aims to quantify the periodic component of EEG data and analyze its characteristic in different scenarios. We will apply three distinct models as described in the previous section. The first model is sinusoidal regression, while the other two are obtained through bidirectional feature selection using the AIC and BIC measures. This methodology will be applied to data collected from different subjects, various electrodes, and diverse segments with different length. This allows us to analyze periodic patterns in the data. This allows us to analyze periodic patterns in the data and ensure reliable comparison and robust conclusions.

Five different participants have been randomly selected for this step of the study, respectively subjects 19, 31, 42, 55 and 62. This selection was made because for each additional subject we had to fit a total number of 315 models. As we will discuss in the forthcoming sections, the results for the selected 5 subjects were very consistent, so we didn't see the necessity of including more subjects at this point of the analysis. For instance, we see the addition of other subjects as something to be done in future work.

We selected electrodes that are located in different brain regions. This aims to ensure generalizability in our results by verifying that our conclusions are reliable and valid across a variety of neural regions. The Table 3.1 below shows the selected electrodes and the corresponding brain regions they are associated with:

Electrode	Location	Function
C3	Left central cortex	Motor control
F4	Right frontal cortex	Executive functions and decision making
FP2	Right prefrontal cortex	Attention and higher cognitive functions
O1	Left occipital cortex	Visual processing
P4	Right parietal cortex	Sensory perception and integration

Table 3. 1 – Selected electrodes, their location and brain function

The data segments used for this part of the analysis are coming from the first 90 seconds of each recording. As we mentioned earlier, these data serve as a baseline and are not influenced by any task-related activity. Segments of different lengths were selected for each subject-electrode combination. Starting from the 40^{th} second of the signal, we created segments that end at 40.5 seconds, 41 seconds, 41.5 seconds, and so on, up to 50.5 seconds. The reason why this starting point was selected is that it resides round the middle of the baseline segment. By doing so we tried to reduce the possibility of artifacts and emotional reactions that might be present while the subject becomes emotionally and physically stabilized before starting the task. This selection provided us 21 segments of different lengths for each case, specifically segments of 0.5 seconds, of 1 second, 1.5 seconds, and continuing in 0.5-second increments up to 10.5 seconds.

3.7 - Model Interpretation

After fitting each model to every segment for each subject, we assessed the goodness of fit using Adjusted R^2. This makes possible comparing models with different number of parameters because it changes the AIC and BIC stepwise selection. The Adjusted R^2 values for every subject-electrode combination were used to produce the graphs shown below (Figure 3.2). They provide an overview of how well the models capture the underlying periodic patterns in the EEG data across different experimental conditions. This analysis allows us to evaluate visually the performance of each model across different segments, subjects, and electrodes. In total 105 segments have been used for each selected subject, for a total of 525 segments and 1575 fitted models (Each of them fitted by the three models).

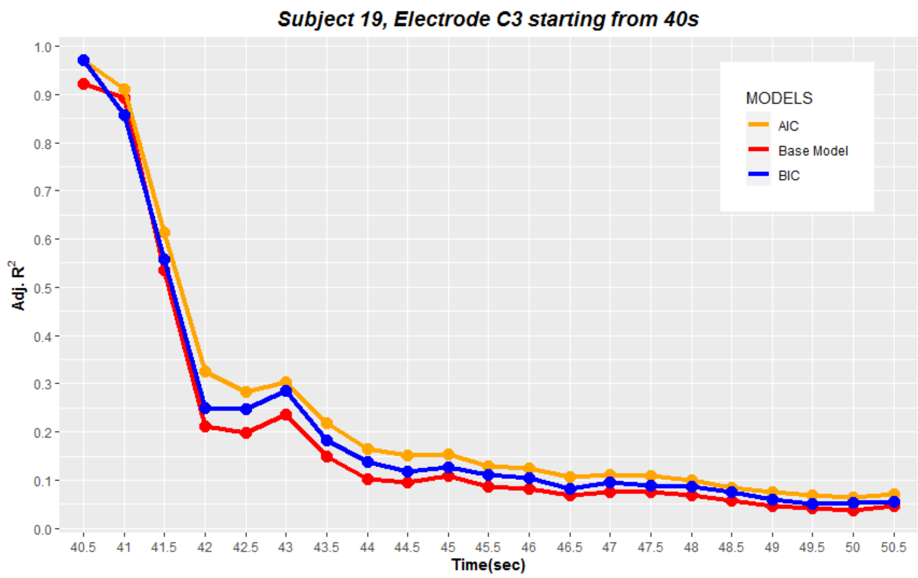


Figure 3. 2 – Sinusoidal Regression for Subject19 on C3 electrode

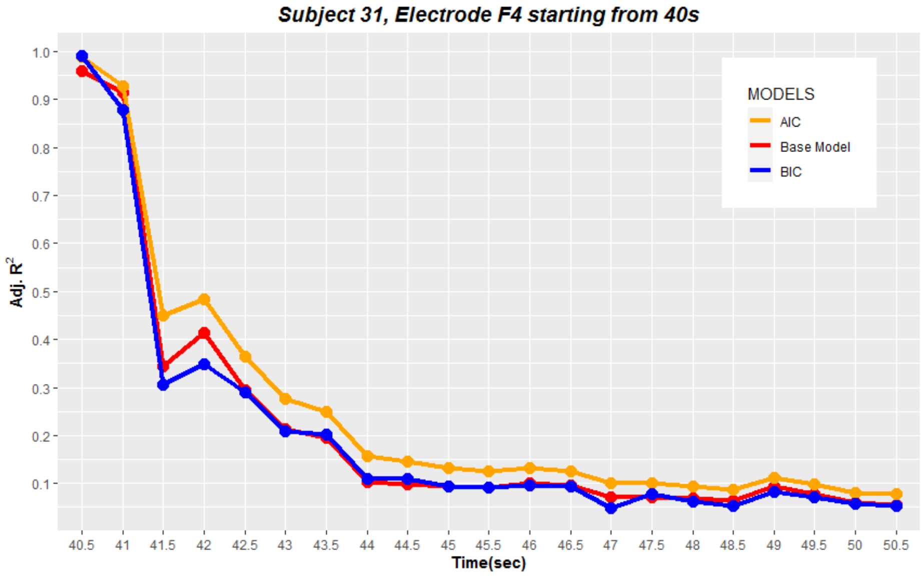


Figure 3. 3 – Sinusoidal Regression for Subject31 on F4 electrode

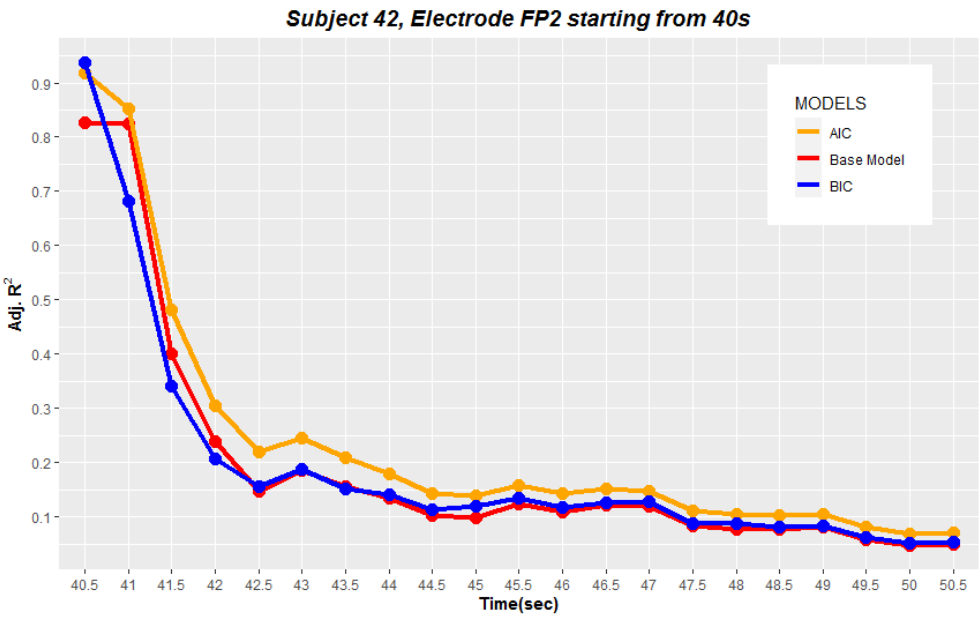


Figure 3. 4 – Sinusoidal Regression for Subject42 on FP2 electrode

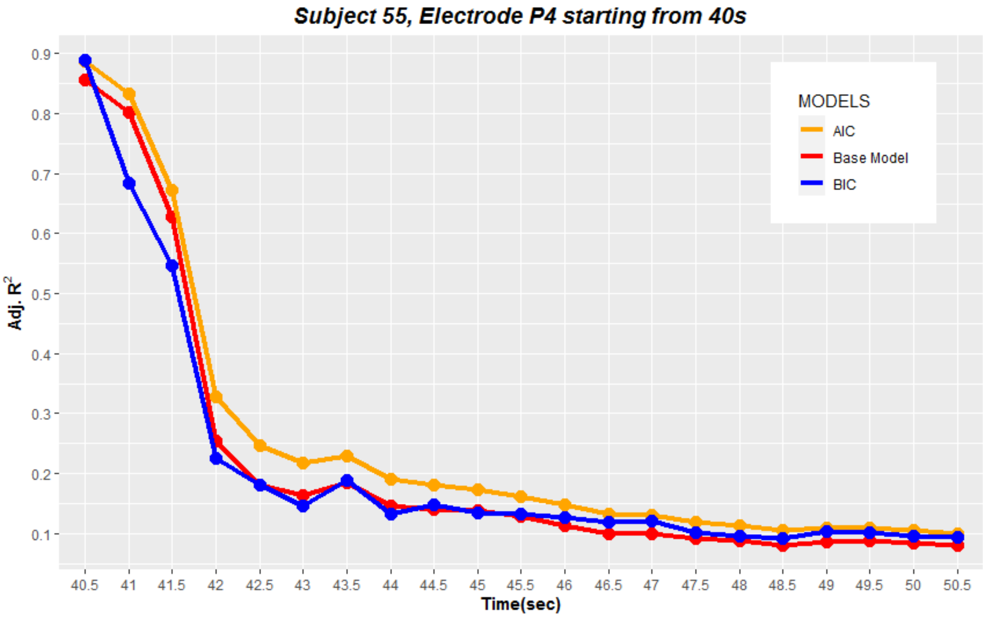


Figure 3. 5 – Sinusoidal Regression for Subject55 on P4 electrode

Only the graphs (Figure 3.2 - 3.5) for four specific subject-electrode combinations are presented here, but our analysis shows that all other combinations share similar characteristics. In these graphs, the y-axis represents the Adjusted R^2 values, while the x-axis the segment length in seconds, starting from the 40^{th} second of recording and ending at the labeled time point. Each graph contains the Adjusted R^2 values for the three models used in this section of the analysis.

For short segments, the Adjusted R^2 values indicate that the models can explain a high percentage of data variability, going above 0.9 in almost all cases. However, as the segment length increases, a decline in model performance is observed. For segments longer than 2-3 seconds, performance stabilizes at lower levels, with Adjusted R^2 values dropping below 0.1 for the longest segments.

The AIC model generally performs better than the other two. While the base model and the BIC one share similar performance levels.

Having similar results across every case suggest that these findings are generalizable among participants and electrodes. With consistent performance for signals of the same length.

3.8 - Change starting point

To continue with our generalization, the next step of our analysis will replicate the same procedure, for the same subjects and same electrodes but now staring at different time points of the 90 seconds baseline signal. Instead of relying only on the 40th second as a starting point, now we use the 50th and 60th second and will create segments with a length of 1 second up to 3 seconds, incrementing by 0.5 seconds for each segment. The graphs below (Figure 3.6) show the results for Subject 55 and O1 electrode in these new scenarios after fitting the same models as previously done.

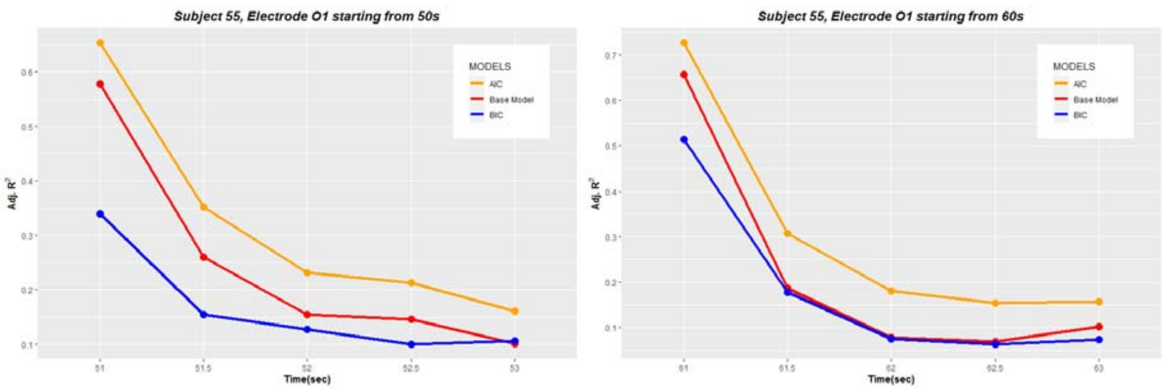


Figure 3. 6 – Sinusoidal Regression for Subject55 on O1 electrode

It is evident that the same conclusions can be drawn as in the cases studied in the previous section. Changing the starting point of the data segments, in addition to the subject and electrode, does not affect the conclusions of this part of the analysis. The consistency of these results across different starting points further strengthens the robustness of our conclusions.

3.9 - Model Performance Analysis

The decrease of sinusoidal regression model performance for this particular dataset, as the EEG signal increases can be attributed to different factors:

- **Non-Stationarity of EEG data:** This means that their statistical properties are prone to change over time. On the other hand, the sinusoidal regression assumes stationarity with periodic components that do not change over time. As the signal length increases, the stationarity assumption becomes less valid, resulting in a decrease in the model's performance.

- **Complexity, noise and artifacts:** As stated before, EEG data consist of a mix of rhythmic, arrhythmic, noise and artifacts. This means that longer EEG segments are more likely to contain noise, artifacts such as muscle movements, eye blinks etc. These can distort the signal and affect the periodic model's capacity to capture the true signal.

CHAPTER 4

Model Comparison

Based on the conclusions of the previous chapter, if we want to study and quantify the periodicity of EEG data, we have to analyze short segments that do not exceed 3 seconds. We saw that as the signal length increases, its periodic component makes up less portion of its variation. Based on these conclusions we decided that the proper segment length that we should take in consideration for periodicity analysis, should not exceed 2-3 seconds.

Our first approach for analyzing the periodic component on the ERP segment will be to analyze the periodicity of a baseline data segments, where there is no event specifically induced by the experiment. Our aim is to quantify the periodic component of a specific data segment and use the fitted values to predict the data in the next segment. The expectation being the difference between the predicted and the observed segment would then represent the expected EEG activity induced by the experiment.

More precisely, we will take into consideration two segments belonging to the baseline before the event datapoint. The first one starts 2 seconds before the event's marker and ends one second before it. The second one starts 1 second before the marker and ends at the event's timepoint. Through our analysis, we aim to quantify the periodicity in the first segment and use this to predict the second one.

This first step of the analysis will be performed on these two segments as they are not affected by any event as they belong to the baseline before the event happens. If the results are promising, we aim to use the same approach by quantifying the periodicity before the event and use it to predict the periodicity of the segment that starts at the event's timepoint and ends 1 second after that.

4.1 - Measure of performance

We will use the coefficient of determination R^2 to compare the model performance on the EEG data. R^2 represents the portion of the variance of the dependent variable, explained by the independent variables. When fitting the dataset, the range of values that R^2 can take are between 0 and 1. Where 0 means that the model explains none of the variability of the dataset and 1 means that the model is able to explain all the variability of the data. In general, higher R^2 values represent better model performance.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$$

- **SS_{RES}:** Residual Sum of Squares is calculated as in the formula above and is used to measure the variance of the dependent variable which is not explained by the independent ones. Where Y_i is the value of the dataset and $\hat{y}_i$ s its estimation from the model.
- **SS_{TOT}**: Total Sum of Squares is calculated as in the formula above and represents the variation of the dependent variable data from the sample's mean.
- **SS_{RES} / SS_{TOT}**: Is the ratio between the model and the data mean performances. It is a measure of performance of our model when compared to the data's mean performance.

There are two extreme cases of R^2 while applying a statistical model to our data. In the first one SS_{RES} is equal to SS_{TOT}. This means that our model is not performing better than the mean of the data. In this case R^2 will have a value of 0, which means that the model explains 0% of the dataset variance. The other extreme case happens when **SS_{RES}** is equal to zero. This happens when our model perfectly fits the dataset. In this case R^2 will have a value of one, which means that the model is able to perfectly explain 100% of the dataset variance.

The major drawback of this measure is that it never decreases when we add more independent variables, it rather stays the same or increases. This is the case also when we add insignificant variables. It means that the R^2 value increases by increasing the model's complexity.

Adjusted R^2, a modified version of R^2, is used to compare models that have different number of independent variables. By taking into account the number of variables, the Adjusted R^2 value increases only when the added variables improve the model's performance and decreases when the added values do not improve it.

$$Adjusted\ R^2 = 1 - \frac{(1 - R^2)(N-1)}{N-p-1}$$

- **N**: Number of datapoints in the dataset.
- **p:** Number of independent variables.
- $\mathbf{R^2}$**:** R^2 value for the model.

We will use in this section another metric: Root Mean Squared Error (RMSE) which is very popular in fields such as statistics, data science and neuroscience. RMSE is used to evaluate the performance of predictive models by measuring the difference between the predicted values coming from the model and the observed data. It is calculated as the square root of the average of the squared residuals from the predicted and observed values as in the following formula:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

Where:

- **n:** the number of observations,
- $\mathbf{y_i}$: the observed values,
- $\hat{\mathbf{y}}_i$: the predicted values.

RMSE is measured in the same units as the response variable. Lower values indicate that the predicted values are similar to the actual data. This results in a good performance of the predictive model. The opposite is true for high values of RMSE. It penalizes models with large errors, and this makes it sensitive to outliers. It is very useful when we want to compare models' performance on the same dataset.

4.2 - Selected Models

The selected models for this step of the analysis will be as follows:

4.2.1 - Sinusoidal Regression

$$y = a + b\sin(2\pi ft) + c\cos(2\pi ft) + \epsilon$$

Where:

- **y:** the dependent variable.
- **a:** the mean value or vertical shift of the sinusoidal function.
- **b and c:** coefficients that scale the sine and cosine components, respectively.
- **f:** the frequency of the sinusoidal wave.
- **t:** the time variable or independent variable.
- **ϵ:** the error term, representing the random noise and unexplained variability in the data.

In this part of the study we are trying to model the periodical part of the EEG data. We saw that there is a periodic component in the EEG data in the form of brainwaves: alpha, beta etc. Sinusoidal model allows to explicitly capture the frequencies that we want to study. Another important aspect is that though sinusoidal regression we can easily interpret the results and see easily which brainwaves were dominant in the selected data. These are some reasons why we decided to use this model in our study.

4.2.2 - Polynomial Regression

$$y = a_0 + \sum_{i=1}^{n} a_i t^i + \epsilon$$

Where:

- **y:** the dependent variable.
- **a_0:** the intercept.

- **a_i:** coefficients that scale the sine and cosine components, respectively.
- **t:** the time variable or independent variable.
- **n:** the degree of polynomial.
- **ϵ:** the error term.

Polynomial regression can be suitable for modeling EEG data as it can effectively capture non-linear and complex patterns which are known characteristic of these data. Its flexibility can be helpful to capture oscillations especially in short segments, such as the ones that we are studying. It also can capture the overall trend of the data, making it a good choice for generalization of the models.

4.2.3 - Spline Regression

Spline regression is a model which fits piecewise polynomial functions called splines to the data. It divides the data int segments and fits a function for each segment. The connecting point of these segments are known as "knots". The regression makes sure that the connection between two splines is a smooth transition. Because it fits different polynomials to the data, it is very suitable to capture complex patterns. This is an advantage when compared to the polynomial regression where the whole data it fitted with one polynomial line. The cubic polynomial spline, known as cubic spline, is the most popular one because of the balance they provide between flexibility and smoothness. The position of the knots can be chosen based on the data or by the analyst.

Figure 4.1 shows an example of a cubic spline regression with three knots. Each portion of the data divided by the red dotted lines (knots), is fitted by a cubic function. We also notice the smoothness of the function at the knots level.

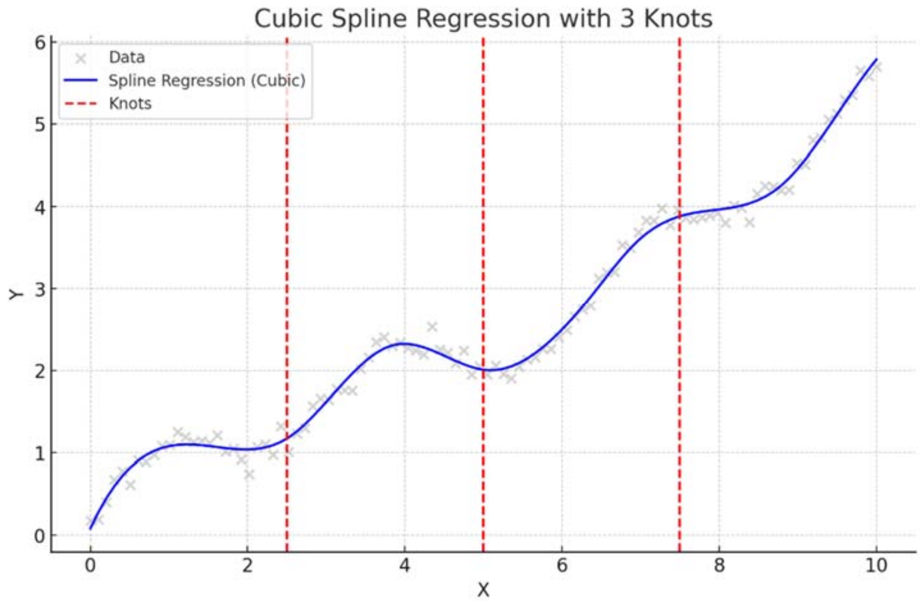


Figure 4. 1 – Cubic Spline Regression example with 3 knots

The splines' capability to model complex non-linear relationships and the smoothness at the knots make them a good fit for modeling EEG data. They also can adapt easily to the data changes over time. This can help spline regression to capture the non-stationary nature of EEG data by placing knots where there is a change in the statistical properties of the signal. For the reasons mentioned above we will use cubic splines during our analysis.

4.3 - Fitted Models

Having described the type of the models that we will be using in this part of the study

- **Sinusoidal regression with 121 parameters:**

The intercept, a sine and a cosine parameter for each frequency starting from 1 Hz to 60 Hz, with an increment of one. This is its final form:

$$y = a + \sum_{k=1}^{60} \left(b_k \sin(2\pi kt) + c_k \cos(2\pi kt)\right) + \epsilon$$

This range of frequencies is chosen to address the brainwaves type that we discussed previously. We have included in our model frequencies up to the gamma wave level, which is the brainwave with fastest oscillations, in the range of 30-60 Hz. Going beyond the range of the brainwaves that we are aiming to capture, increases the risk that the model will start to fit the noise which is inherited in EEG data.

- **BIC Stepwise selection:**

Selection performed in both directions (forward and backward) on the sinusoidal model above. This model is more restrictive than the base one and contains only statistically significant parameters.

- **Polynomial Regression with 121 parameters**:

$$y = a_0 + \sum_{i=1}^{120} a_i t^i$$

The degree of the polynomials will start from one and end at 120 with an increment of one. This model will have 121 parameters. This model will have the same number of parameters as the sinusoidal chosen above, which helps us analyse and compare the results.

- **Cubic Splines Regression with 120 knots**:

This model will have the same number of parameters as the sinusoidal chosen above, which helps our analysis and model comparison.

- **Sinusoidal Regression with 61 parameters**:

This is similar to the first sinusoidal model but has only 61 parameters: the intercept, a sine and a cosine parameter for each frequency starting from 1Hz to 30 Hz, with an increment of one.

$$y = a + \sum_{k=1}^{30} (b_k \sin(2\pi kt) + c_k \cos(2\pi kt)) + \epsilon$$

This less complex model was chosen in order to compare its results with the previous sinusoidal model. If their performance is similar, we can choose the one with fewer parameters as it can generalize better. With this model we are still able to capture the delta, theta, alpha, and beta

bands that we want to study and that are probable to be present in the data based their natural characteristics.
Since some oscillations can occur even in the range of less than 1 Hz, for example 0.5 Hz, using a larger incremental step than one, could lead to lose important information.

- **BIC Stepwise selection:**

Selection performed in both directions (forward and backward) on the sinusoidal model above. This model is more restrictive than the base one and contains only statistically significant parameters.

- **Polynomial Regression with 61 parameters:**

$$y = a_0 + \sum_{i=1}^{60} a_i t^i$$

The degree of the polynomials will start at one and end at 60 with an increment of one. This model will have 61 parameters. This model will have the same number of parameters as the sinusoidal chosen above, which helps us analyse and compare the results.

- **Cubic Splines Regression with 60 knots**:

This model will have the same number of parameters as the sinusoidal chosen above, which helps our analysis and model comparison.

- **Sinusoidal Regression with 121 parameters:**

$$\mathbf{y} = \mathbf{a} + \mathbf{b_1}\sin(2\pi t) + \mathbf{c_1}\cos(2\pi t) + \mathbf{b_2}\sin(2\pi 1.5t) + \mathbf{c_2}\cos(2\pi 1.5t) + \ldots + \mathbf{b_{30}}\sin(2\pi 30t) + \mathbf{c_{30}}\cos(2\pi 30t)$$

This model has 121 parameters, the same number of parameters as the first sinusoidal model that we chose: the intercept, a sine and a cosine parameter for each frequency from 1-30 Hz, with an increment of 0.5 Hz. The choice of smaller step size was done with the purpose to ensure that any statistically significant frequencies skipped in the other model, are captured.

that we might have missed with the first model. This model combines important frequencies while increasing the granularity of the analysis.

- **BIC Stepwise selection:**

Selection performed in both directions (forward and backward) on the sinusoidal model above. This model is more restrictive than the base one and contains only statistically significant parameters.

As we previously stated, the choice of sinusoidal regression was natural based on the nature of EEG data and the objective of the analysis at this stage. We chose different sinusoidal models with different characteristics and capacities in order account for different scenarios. The polynomial and the spline models were chosen not only for their statistical characteristic discussed above, but also to serve as a comparison reference point for the sinusoidal ones. This is the reason we chose polynomial and splines models with the same capacity as the sinusoidal ones in each step.

4.4 - Models Performance Analysis

The ten models selected in the previous section have been fitted into 80 segments of 1 second for each participant, starting 2 seconds before the 'xTar' event's timepoint and ending 1 second before it. The adjusted R^2 have been calculated for each case. In order to simplify the comparison among different models' performances, for every combination of participant and model the minimum, mean and maximum values of Adjuster R^2 are calculated. The results are shown in Table 4.1 below:

		SELECTED MODELS									
		Sin 60Hz Step 1	BIC Sin 60Hz Step 1	Polynomial Power 120	Cubic Splines 120 Knots	Sin 30Hz Step 1	BIC Sin 30Hz Step 1	Polynomial Power 60	Cubic Splines 60 Knots	Sin 30Hz Step 0.5	BIC Sin 30Hz Step 0.5
Subject 19	Min	0.772	0.640	0.133	0.755	0.665	0.551	0.109	0.647	0.659	0.698
	Mean	0.901	0.845	0.430	0.895	0.838	0.799	0.365	0.838	0.864	0.879
	Max	0.966	0.952	0.793	0.973	0.951	0.946	0.770	0.952	0.964	0.967
Subject 31	Min	0.668	0.328	0.196	0.581	0.498	0.263	0.131	0.455	0.437	0.503
	Mean	0.839	0.751	0.435	0.807	0.737	0.670	0.393	0.733	0.762	0.790
	Max	0.968	0.957	0.728	0.972	0.952	0.940	0.661	0.953	0.965	0.971
Subject 42	Min	0.749	0.655	0.184	0.789	0.690	0.598	0.109	0.683	0.721	0.737
	Mean	0.898	0.847	0.524	0.899	0.843	0.811	0.476	0.847	0.873	0.886
	Max	0.966	0.954	0.771	0.978	0.951	0.941	0.768	0.959	0.975	0.977
Subject 55	Min	0.592	0.412	0.087	0.575	0.532	0.412	0.086	0.526	0.551	0.589
	Mean	0.899	0.836	0.518	0.882	0.828	0.786	0.467	0.827	0.849	0.864
	Max	0.978	0.963	0.887	0.975	0.960	0.953	0.871	0.958	0.966	0.969
Subject 62	Min	0.831	0.784	0.074	0.893	0.772	0.745	0.071	0.783	0.864	0.881
	Mean	0.966	0.952	0.449	0.972	0.932	0.921	0.381	0.932	0.957	0.961
	Max	0.995	0.994	0.768	0.995	0.986	0.984	0.753	0.984	0.991	0.992
	Model Mean	**0.901**	**0.846**	**0.471**	**0.891**	**0.836**	**0.797**	**0.416**	**0.835**	**0.861**	**0.876**

Table 4. 1 – Adjusted R^2 values for the selected models

Comparing the models using the Adjusted R^2 means values for each subject helps us make a general and robust evaluation of their performance.

We notice that the best performing model is the sinusoidal regression with frequencies from 1 to 60 Hz, followed by the cubic splines with 120 knots and BIC of the sinusoidal model with frequencies from 1 to 30 Hz. This is true for all the cases apart from the subject 42 case where the splines model performs slightly better than the sinusoidal one. Even here the Adjusted R^2 mean values are almost at the same level, respectively 0.899 and 0.898.

On the other side, the model with worst performance is the polynomial regression with 60 parameters, followed by the polynomial with 120 parameter and the sinusoidal BIC Stepwise with frequencies from 1 to 60 Hz. This pattern is true for all the subject. It is worth mentioning that even though the BIC sinusoidal is the third worst performing model, there is a significant gap among its performance and the polynomial models. For example, in case of subject 62 for example its mean Adjusted R^2 vale is 0.921 which means that the model is able to explain 92.1% of the variation of the data in this segment, which is good performance.

The same conclusions can be drawn also by analyzing the minimal and maximal values of Adjusted R^2 for each subject. We notice that maximal values of polynomial models are in the same range as the minimal values for the other models whereas their minimal values can go up to 0.071, as in case of subject 62.

The same conclusions can be drawn also by analyzing the mean Adjusted R2 values for each model where the sinusoidal regression with frequencies from 1 to 60 Hz performs best, followed by the cubic splines with 120 knots, with respective Adjusted R2 values of 0.901 and

0.891. Based on the models' Adjusted R2 mean values the model with the worst performance is again the polynomial regression with 60 parameters with a value of 0.416, which means that the model is able to explain 41.6% of the variation of the segment.

	Subject 19		Subject 31		Subject 42		Subject 55		Subject 62	
	Model	Adj. R^2	Model	Adj. R^2	Model	Adj. R^2	Model	Adj. R^2	Model	Adj. R^2
Best perf.	Sin 60Hz - S1	0.901	Sin 60Hz - S1	0.839	Splines - 120 k.	0.899	Sin 60Hz - S1	0.899	Sin 60Hz - S1	0.966
	Splines - 120 k.	0.895	Splines - 120 k.	0.807	Sin 60Hz - S 1	0.898	Splines - 120 k.	0.882	Splines - 120 k.	0.972
	BIC Sin 30Hz - S0.5	0.879	BIC Sin 30Hz - S0.5	0.790	BIC Sin 30Hz - S0.5	0.886	BIC Sin 30Hz - S0.5	0.864	BIC Sin 30Hz - S0.5	0.961
Worst perf.	Poly - Pwr. 60	0.365	Poly - Pwr. 60	0.393	Poly - Pwr. 60	0.476	Poly - Pwr. 60	0.365	Poly - Pwr. 60	0.381
	Poly - Pwr. 120	0.430	Poly - Pwr. 120	0.435	Poly - Pwr. 120	0.524	Poly - Pwr. 120	0.518	Poly - Pwr. 120	0.449
	BIC Sin 60Hz - S1	0.799	BIC Sin 30Hz - S1	0.670	BIC Sin 30Hz - S1	0.811	BIC Sin 30Hz - S1	0.786	BIC Sin 30Hz - S1	0.921

Table 4. 2 – Summary of Adjusted R^2 values for the selected models

In general, the sinusoidal models are those with the best performance, followed by the cubic splines regression ones. As we notice from the summary Table 4.2 above, the polynomial regressions are the ones with worse performance. For all the subject the polynomial regression which goes up to the power of 60 is the one with the worse performance, with adjusted R^2 value of 0.071 in the case of Subject 62. This means that this model is able to explain only 7,1% of the data's variability.

4.5 - Prediction using Adjusted R^2

After having fitted our models on all the [-2;-1] data segments with respect to the 'xtar' event's timepoint, the next step will be to use these values to predict the data segment starting one second before the event timepoint and ending at the event timepoint. This approach will allow us to measure the predictive performance of the selected models in the scenario where the signal is not affected by the event.

Table 4.3 below shows the minimal, maximal and mean values of Adjusted R^2 for each subject. We can easily notice that in almost all the cases the values are negative. This fact indicates poor predictive capacity for all our models. Analyzing these results based on the formula used to calculate the Adjusted R^2, we see that negative results derive from that the sum of squares from the models (SS_{RES}) larger than the sum of squares of the sample's mean (SS_{TOT}). This leads to negative values for both R^2 and Adjusted R^2 meaning that the models have less predictive capacity than the sample's mean.

		SELECTED MODELS									
		Sin 60Hz Step 1	BIC Sin 60Hz Step 1	Polynomial Power 120	Cubic Splines 120 Knots	Sin 30Hz Step 1	BIC Sin 30Hz Step 1	Polynomial Power 60	Cubic Splines 60 Knots	Sin 30Hz Step 0.5	BIC Sin 30Hz Step 0.5
Subject 19	Min	-8.012	-4.461	-2.043	-8.042	-5.153	-4.381	-1.715	-5.172	-7.755	-6.132
	Mean	-2.944	-1.305	-0.756	-2.967	-1.652	-1.202	-0.635	-1.665	-2.821	-2.106
	Max	-1.088	-0.189	-0.156	-1.132	-0.406	-0.162	0.036	-0.439	-1.035	-0.632
Subject 31	Min	-5.877	-3.116	-2.403	-5.878	-3.644	-2.817	-2.192	-3.665	-5.652	-4.553
	Mean	-3.001	-1.224	-0.873	-3.047	-1.697	-1.129	-0.747	-1.700	-2.895	-2.127
	Max	-1.330	-0.086	-0.104	-1.323	-0.543	-0.137	-0.049	-0.539	-1.255	-0.727
Subject 42	Min	-10.024	-5.996	-2.112	-10.105	-6.526	-5.699	-2.052	-6.621	-9.797	-8.088
	Mean	-3.003	-1.359	-0.900	-3.027	-1.698	-1.253	-0.799	-1.713	-2.893	-2.138
	Max	-1.247	-0.279	-0.067	-1.230	-0.514	-0.246	0.004	-0.521	-1.183	-0.646
Subject 55	Min	-5.224	-2.848	-1.653	-5.226	-3.243	-2.532	-1.659	-3.215	-5.004	-4.000
	Mean	-2.714	-1.139	-0.761	-2.739	-1.510	-1.051	-0.641	-1.512	-2.613	-1.872
	Max	-0.105	0.448	0.456	-0.097	0.275	0.487	0.539	0.286	-0.048	0.195
Subject 62	Min	-10.916	-6.935	-3.187	-10.939	-7.192	-6.439	-2.385	-7.157	-10.545	-8.437
	Mean	-3.942	-2.238	-1.026	-3.967	-2.375	-2.003	-0.835	-2.384	-3.808	-2.931
	Max	-1.368	-0.526	-0.166	-1.373	-0.621	-0.446	-0.050	-0.609	-1.292	-0.850
	Model Mean	**-3.121**	**-1.453**	**-0.863**	**-3.150**	**-1.786**	**-1.328**	**-0.731**	**-1.795**	**-3.006**	**-2.235**

Table 4. 3 – Adjusted R^2 values on [-1; 0] segment

Table 4.4 below presents minimum, maximum and mean RMSE values for each combination of subject and selected model under the same scenario.

		SELECTED MODELS									
		Sin 60Hz Step 1	BIC Sin 60Hz Step 1	Polynomial Power 120	Cubic Splines 120 Knots	Sin 30Hz Step 1	BIC Sin 30Hz Step 1	Polynomial Power 60	Cubic Splines 60 Knots	Sin 30Hz Step 0.5	BIC Sin 30Hz Step 0.5
Subject 19	Min	5.531	5.292	5.045	5.592	5.374	5.032	4.564	5.401	5.551	5.531
	Mean	7.554	7.397	6.586	7.575	7.446	7.347	6.438	7.463	7.544	7.534
	Max	11.417	11.345	10.133	11.436	11.338	11.313	9.961	11.355	11.418	11.397
Subject 31	Min	2.603	2.372	2.420	2.637	2.500	2.372	2.429	2.532	2.566	2.565
	Mean	4.815	4.647	4.280	4.844	4.757	4.616	4.184	4.759	4.824	4.814
	Max	8.030	7.988	7.219	8.046	7.939	7.762	7.228	7.912	7.984	7.985
Subject 42	Min	4.866	4.605	4.141	4.906	4.763	4.605	4.044	4.814	4.869	4.871
	Mean	6.869	6.732	6.146	6.889	6.778	6.677	6.051	6.797	6.873	6.859
	Max	13.119	13.083	11.902	13.120	13.057	12.952	11.735	13.065	13.110	13.093
Subject 55	Min	4.531	3.800	3.921	4.547	4.328	3.800	3.872	4.304	4.534	4.460
	Mean	8.189	7.993	7.358	8.215	8.094	7.956	7.185	8.096	8.193	8.173
	Max	12.720	12.604	12.527	12.721	12.615	12.604	12.550	12.646	12.713	12.706
Subject 62	Min	6.423	6.397	4.898	6.428	6.385	6.290	4.654	6.329	6.408	6.406
	Mean	10.006	9.951	8.384	10.033	9.938	9.898	8.101	9.958	10.017	10.013
	Max	18.001	17.967	14.638	18.016	17.970	17.960	13.381	17.940	18.010	18.007
	Model Mean	**7.487**	**7.344**	**6.551**	**7.511**	**7.403**	**7.299**	**6.392**	**7.415**	**7.490**	**7.478**

Table 4. 4 – RMSE values on [-1; 0] segment

RMSE values share the same measuring unit of the dependent variable which makes their interpretability context dependent. Being a relative measure of performance, RMSE interpretation should take in consideration the statistical characteristics of the data, such as mean and standard deviation.

Analyzing our data segments, we noticed that their values tend to have values close to zero, which is an expected characteristic of EEG data. However, the RMSE values across all subject, as we notice from Table 4.4, are significantly higher. The minimal value of 2.372 belongs to subject 31 in case of BIC sinusoidal models (columns two and six) and maximal value of

18.016 to subject 62 the cubic splines model with 120 knots. These results prove once again what we noticed when analyzed the adjusted R^2 values, where the predictive models were performing worse than the mean of the actual data.

We can use also the standard deviation to relatively measure the model's performance. A commonly used rule of thumb is that RMSE values should be significantly lower than the standard deviation of the data. In many cases a threshold of 0.5 the value of standard deviation ($0.5 * \sigma$) is used to indicate a good performing model. On contrary, a RMSE value close to or higher than the standard deviation indicates a poorly performing model. In our case, all RMSE values were higher than the data's standard deviation, reinforcing the conclusions drawn from our previous analysis about the model's performance.

4.6 - Graphical interpretation

We tried to explain part of the [-1; 0] segments' variability using the periodic component of the [-2; -1] interval, aiming to reduce the "periodic noise" when there were no signals induced. However, based on the results that we obtained, it is clear that none of the models served to our purpose. Thus, at this point it is not beneficial to compare the different models' performances among each.

We can graphically suggest the same conclusion based on Figure 4.2 below. The black line represents the data for the [-1;0] segment with respect to xTar event number 21 of subject 19. While the blue line represents the fitted values obtained from the BIC sinusoidal model with frequencies from 1 to 60 Hz on the [-2; -1] segment of the same event and subject. Upon analyzing the graph (Figure 4.2), we notice that the periodic model fails to capture the variability of the actual data. This can particularly be noticed around the 0,2 and 0.7 second points, where the actual data and the predicted values diverge significantly, moving on opposite directions.

These results support what we faced in the previous analysis where the model was seen to perform worse than the mean of the data. We are using this case for illustrative purposes, but the same results have been noticed for all the subjects and all models.

These results can derive from the inherent nature of EEG data, which are considered to be highly variable. These signals are affected by various physiological factors, such as neural activity and cardiac rhythms, as well environmental factors. They are also considered to be non-stationary, which means that the statistical characteristics tend to change over time. For example, a segment of EEG data can have dominant frequencies and amplitudes which are different from the ones of the subsequent segment. As a result, models that perform well on a particular segment might fail to generalize in other segments, resulting in low predictive capability.

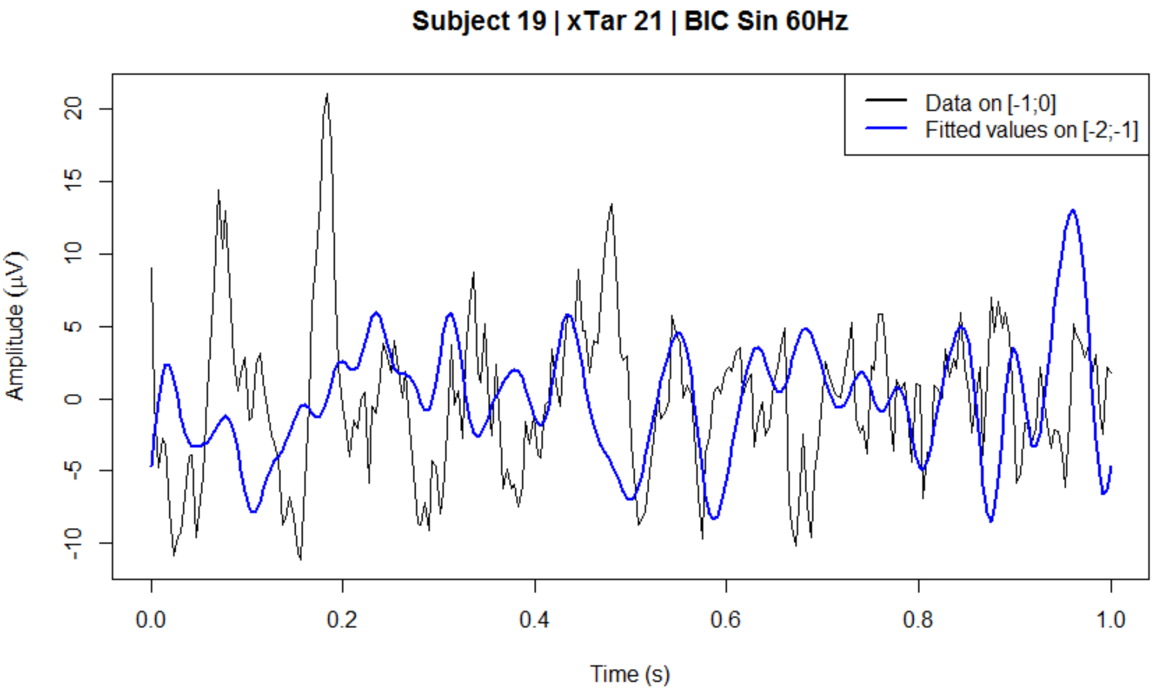


Figure 4. 2 – ERP signal and BIC fitted values

Based on the results of our analysis and the nature of EEG data we can conclude that if we want to quantify the periodical component of the data while the event is happening (ERP), it would be advisable to analyze segments which contain the event rather than trying to predict it based on other segments of the data. By doing so, it is more likely that the signal contains the true periodic component present during the event and is less affected by high variability and non-stationarity nature.

CHAPTER 5

FFT Filter

Based on the results of the analysis in the previous chapter, we concluded that our attempt to quantify the periodic component of an EEG data segment based on the periodicity of a preceding adjacent segment was not successful. For this we used two segments of 1 second length. Despite our analysis included different types of statistical models with different capacities, the results were always unsatisfactory. Through quantitative and graphical analysis, we concluded that these results came as a result of the nature of the EEG data. Their non-stationary nature, with statistical properties changing over time, was the main factor contributing to these unsatisfactory results.

Therefore, the next step of our analysis will attempt to quantify the periodic component directly within the data segments that contain the potential reaction, which are central to this study. In this chapter, as we will see in the following sections, we will use FFT in order to localize the periodic components of the data. An iterative procedure will be followed which is based on the Fast Fourier Transform (FFT) and its inverse method Inverse FFT (IFFT). This procedure will be applied across different scenarios, to data segments of different lengths.

5.1 - Methodology: FFT Filter

The data segments where this method will be applied will remain unchanged. We will use the data coming from Pz electrode of the same five subjects, specifically will focus on the data related to the xTar event, with 80 trials for each participant.

We discussed in literature review that FFT is able to deconstruct a time-domain segment of data as a sum of a finite number of sinusoidal functions with different frequencies, amplitudes and phases. That is why it is often stated that FFT makes possible to perfectly represent the frequency-domain of a time-domain signal (Mike X Cohen, 2014). This means that by applying the inverse function we can pass from the frequency domain back to the time domain without any loss of information. This method is known as Inverse Fast Fourier Transform (IFFT).

This is the FFT property that we will use to construct our FFT filter procedure as follows:

1. **Perform FFT** on the chosen signal to obtain the frequency-domain representation.
2. **Localize the n frequencies** with highest power. Where n is chosen by us.
3. **Set their power to zero** to remove their contribution to the signal.
4. **Perform IFFT** to go back to time-domain.

Going back to the purpose of our study where we stated that by removing the "periodic noise" or the periodic component from the segments with induced potential, we aim to reduce the noise of the signal and increase the signal-to-nose ratio (SNR). In the FFT context, the sinusoidal functions with the most powerful frequencies are those that contribute most significantly to the periodical component of the data. By removing their contribution to the potential-induced segment, we can increase the SNR of a single trial and potentially reducing the number of trials in the ERP that are needed to achieve the same level of SNR. This improvement can help us improve the ecological validity of the experiment, that is what we are trying to achieve in this study.

In this context, the step number three of the FFT filter described above helps us remove the most powerful periodic components from the segments we are studying.

There are several questions that needs to be considered at this point:

1. How to measure the signal-to-noise ratio (SNR)?
2. How many sinusoidal functions should be to remove through FFT filter, or how many frequency coefficients do we need to set to zero at step 3?
3. How to make sure that we are not losing any important information by applying FFT filter?

5.2 - Signal to Noise Ratio (SNR)

Signal-to-noise ratio (SNR) is a measure used in several fields, including signal processing, that compares the level of desired signal to the level of background noise (Sherman and Butler, 2007). A value larger than one indicates that the signal level is greater than the level of the noise. To measure the SNR, we need to be able to dissociate signal from noise but in case of EEG data this is not possible as the signal is a mix of signal and noise (Mike X Cohen, 2014).

In this case we have to rely on estimations. Several methods can be used, but we will estimate SNR as the ratio of the signal's mean to its standard deviation:

$$\boldsymbol{SNR} = \frac{\mu}{\boldsymbol{\sigma}}$$

Where:

- **μ:** the mean of the signal.
- $\boldsymbol{\sigma}$**:** the standard deviation of the signal.

This measurement will help us compare the results obtained from different scenarios to the SNR of the ERP calculated using the original data.

5.3 - Number of Frequency Components to Remove

As a first step, we will apply the FFT filter to the [0; 1] segment, which starts at the xTar event timepoint and has a length of 1 second. We will consider three scenarios: removing 2 most powerful frequencies, removing 4 and 6 frequencies. This approach allows us to analyze the effect that the number of removed frequencies has on the data. The ERP calculated as a mean of the filtered segments will be than compared to the original ERP through their SNR values and the ability of the filtered ERP to preserve the characteristics of the original ERP, such as its general shape and its important components(ex: N100, P300 etc.).

As the mean of both approaches is equal, the difference in SNR will come as a result of the standard deviation values. That's why we will use the standard deviation as a comparison criterion. Table 5.1 shows the standard deviation values for Subject 19, when we average the first 10, 20 up to 80 trials. Note that for illustrative purposes, we will demonstrate mainly the results for subject 19. However, our analysis showed that the results were consistent among other subjects as well.

The first column represents the SD values coming from the original data and the second one those coming from the FFT filter approach.

	SUBJECT 19 - FFT	
Trials	Data	[0;1]
10	2.7022	1.69
20	2.2756	1.311
30	2.0386	1.0967
40	1.8013	0.9767
50	1.665	0.8678
60	1.6278	0.8006
70	1.5697	0.7154
80	1.5376	0.6414

Table 5. 1 – SD values for Subject 19 on [0; 1] removing 4 freq.

We notice a significant reduction of SD when comparing to the values for the same from the original data. The same level of SD for the average of 80 trials from the original data is obtained by averaging between 10 and 20 trials of the data with the periodicity removed.

As mentioned earlier, our aim is not only to reduce the number of trials need to obtain the same level of SNR, reflected as standard deviation (SD) in our case, but also to make sure that the mean of ERP coming from the FFT filtered method is able to preserve the shape of the original ERP.

The graph below (Figure 5.1) compares the 80 trials average coming from the original data (in black) with the 80 trials average coming from the same segments without their periodic component (in red). The x-axis shows the time in seconds while the y-axis the amplitude.

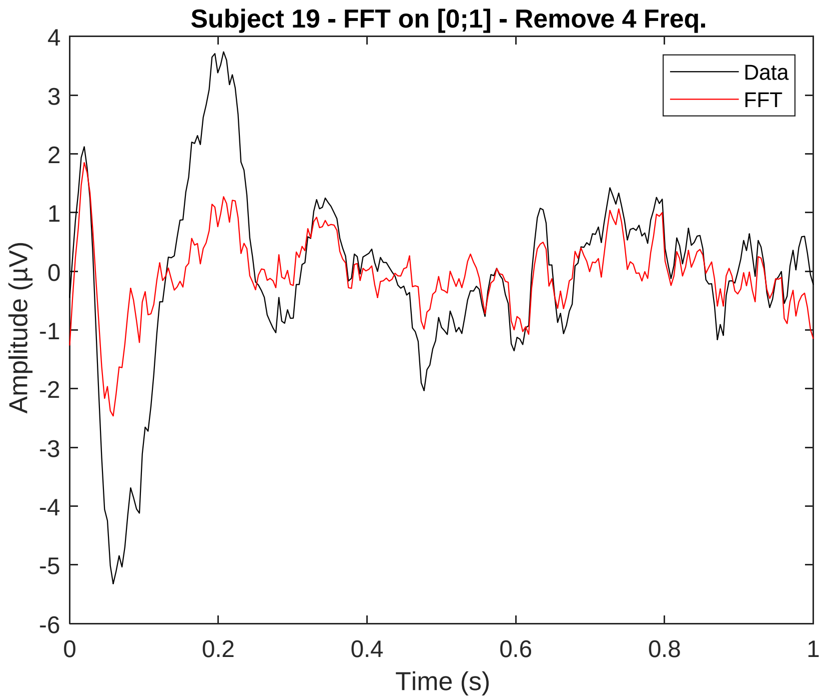


Figure 5. 1 – ERPs for Subject 19 on [0; 1] removing 4 freq.

The first thing that we notice when comparing these two graphs is that the shape of the one coming from the residuals is significantly flatter than the original one. The same conclusion can be derived also by comparing the SD levels of the two averaged segments, where the original data has a SD level of 1.53 compared to 0.64 of the residuals.

Although the method seems to reduce the SD, which theoretically represents the noise component of the data segments, the graph shows that the shape and characteristics of the original data are not preserved. This is more noticeable between 0.05 and 0.3 seconds, where originally the response potential was recorded. The most significant differences are noticed at the first negative peak round 0.05 seconds and the second positive peak round 0.2 seconds. Beyond 0.3 seconds, both graphs seem to share the similar characteristics.

The ERP portion that is most affected by this method is also the most important one, where the main ERP components reside. Affecting these components can significantly influence also the next steps of the analysis or worse, make the results unreliable. Instead of detecting and removing the underlying periodicity from the ERP, the method seems to focus more on the ERP components.

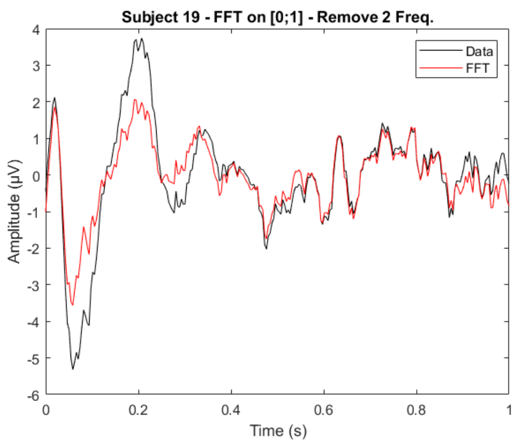
	SUBJECT 19 - FFT			
Trials	Data	2 freq.	4 freq.	6 freq.
10	2.7022	1.9675	1.6900	1.4180
20	2.2756	1.7447	1.3110	1.0743
30	2.0386	1.4600	1.0967	0.8718
40	1.8013	1.2956	0.9767	0.7749
50	1.665	1.2130	0.8678	0.6731
60	1.6278	1.1244	0.8006	0.6273
70	1.5697	1.0508	0.7154	0.5534
80	1.5376	0.9770	0.6414	0.4922

Table 5. 2 – SD values for Subject 19 on [0; 1]

The next step of the analysis involves the removal of 2 and 6 frequencies, following the same procedure.

Note that in the following sections, we will use the terms "2-frequencies scenario" and "6-frequencies scenario" to the cases where we apply the FFT filter by removing 2 most powerful sinusoidal functions and 6 most powerful sinusoidal functions respectively.

Table 5.2 shows the SD values for the other the three scenarios. We notice that the same SD level from the ERP of 80 trials of the original data, is achieved between 10 and 20 averaged trials in 4 frequencies scenario and in the 6 frequencies scenario the same level is achieved for less than 10 trials. These are the graphs (Figure 5.2) that compare the ERP obtained in these two scenarios, comparing the one coming from the original data to the ones coming from FFT filter.



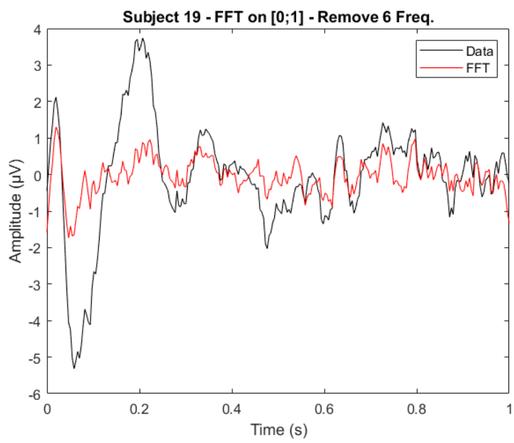


Figure 5. 2 – ERPs for Subject 19 on [0; 1], removing 2 and 6 freq.

As it was expected the “6-frequences scenario” has a flatter filtered ERP because larger portion of the signal was removed. In case of the “2-frequences scenario”, the filtered ERP appears to have similar shape when compared to the original ERP.

Similar results were observed for the other subjects. The “2-frequenies scenario” is able to preserve the general shape and the components of original ERP while achieving a similar level of SD with fewer averaged trials compared to the original ERP.

For the 1 second signal the FFT filter that removes the 2 most powerful frequencies seems to perform better than the other two scenarios. It achieves satisfactory levels of SNR, reflected in SD in our case, while preserving the shape and the components of the original ERP.

5.4 - Extended Signal Length

To account for the underlying periodicity of the data, we decided to increase the length of the analyzed signal. In the first scenario, we analyze a 2 second segment that starts 1 second before the xTar marker and ends 1 second after the event timepoint. While in the second scenario a 3 second segment is analyzed, starting 2 seconds before the marker and ending 1 second after that. This additional portion of the signal added in both scenarios comes from the data before the xTar event and is not affected by other occurring events.

We followed the same procedure as in the case of [0;1] signal and obtained the following SD values for subject 19 (Table 5.3). It is important to mention that after applying the FFT filter to the [-2;1] and [-1;1] segments, the SD of their ERPs is calculated for their [0;1] portion. This ensures that the values are comparable to the SD values of the original ERP and those of the [0;1] scenario.

Trials	Data	SUBJECT 19 - FFT								
		[0;1]			[-1;1]			[-2;1]		
		2 freq.	4 freq.	6 freq.	2 freq.	4 freq.	6 freq.	2 freq.	4 freq.	6 freq.
10	2.7022	1.9675	1.6900	1.4180	2.2919	2.0606	1.8124	2.3599	2.1237	1.8881
20	2.2756	1.7447	1.3110	1.0743	1.9032	1.6292	1.4161	1.9012	1.7597	1.5598
30	2.0386	1.4600	1.0967	0.8718	1.7033	1.4531	1.2604	1.7788	1.6057	1.4150
40	1.8013	1.2956	0.9767	0.7749	1.5110	1.2957	1.1243	1.5677	1.4353	1.2808
50	1.665	1.2130	0.8678	0.6731	1.3783	1.1572	1.0020	1.4282	1.3172	1.1560
60	1.6278	1.1244	0.8006	0.6273	1.3274	1.0904	0.9564	1.4154	1.2691	1.1009
70	1.5697	1.0508	0.7154	0.5534	1.2729	1.0220	0.8820	1.3667	1.2159	1.0511
80	1.5376	0.9770	0.6414	0.4922	1.2182	0.9707	0.8426	1.3305	1.1619	0.9983

Table 5. 3 – All SD values for Subject 19 all scenarios

Table 5.3 adds more information to Table 5.2 that we previously analyzed. The highlighted cells represent the approximate number of averaged trials needed to achieve a similar level of SD as in the original ERP with 80 averaged trials. We notice that in all these cases the number of needed trials varies from 10 to 40, indicating that fewer trials are needed to achieve similar level of SD. We will proceed by analyzing the ERP graphs (Figure 5.3) for these new scenarios:

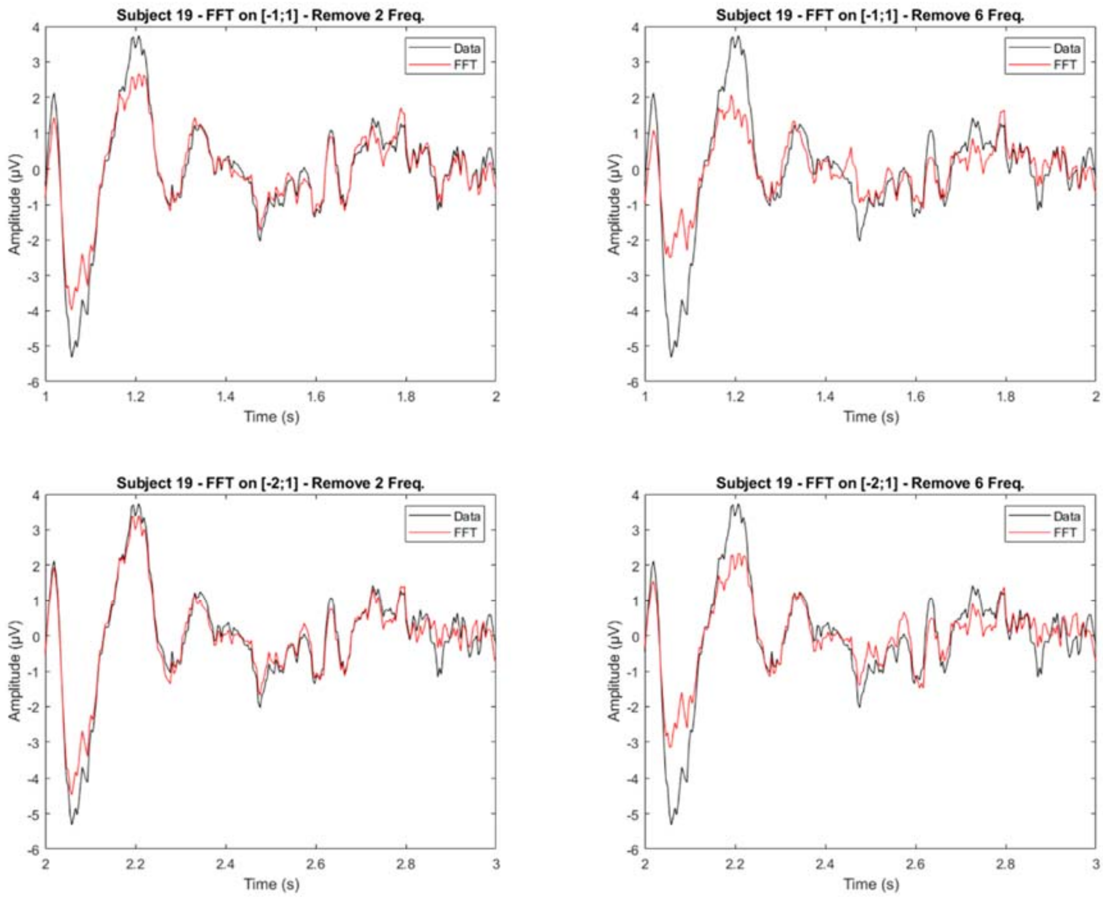


Figure 5. 3 – ERPs for Subject 19 on [-1; 1] and [-2; 1], removing 2 and 6 freq.

The black line in both graphs (Figure 5.3) represent the average of 80 trials coming from the unmodified data, while the red one represents the 80-trial average coming from the same data after removing the periodic component detected by FFT filter respectively on [-1;1] and [-2;1] segments. Even though the FTT filter was performed on these segments, here we are showing only their [0;1] second portion which is starting at the marker's datapoint. The x-axis values are also maintained in the original form to serve as a reference point for these graphs' origin.

The third graph (Figure 5.3) coming from [-2;1] segment with 2 frequencies removed, seems to be almost identical to the original ERP. While the one coming from [-1;1] segment with 6 frequencies removed is flatter than the others. In general, we observed that by increasing the length of the signal, we are able to improve the quality of the signal when compared to the original ERP, while requiring fewer trials to achieve the similar level of SNR, or SD in our case. When comparing the scenarios based on their signal length, we find that for the same number of removed frequencies, the models coming from the [-2;1] scenario are able to preserve better the characteristics of the original ERP while reducing the number of trials needed to achieve the same level of SNR. In general, we can state that the longer the signal and fewer frequencies removed, the more closely the filtered ERP resembles to the original one, which is both logical and intuitive.

This might also be an indicator that when incorporating portions of the segment with no signal is induced, the model is able to capture the underlying periodicity that is present before and during the event. We saw that this aspect was missing in the case of [0;1] segment while removing 4 or 6 frequencies as the FFT filter was removing the ERP features instead of the periodic component. However, are still in the early stages and further research needs to be performed in this direction as we will suggest at the end of the study.

The scenario coming from [-2;1] segment with 4 frequencies removed will be the focus of the analysis in the following chapter. This was chosen as it represents a better balance of both signal length and number of removed frequencies.

CHAPTER 6

ERP Comparison

This section of the analysis builds on the results of the previous chapter. By removing the four most powerful frequencies from the [-2; 1] segments relative to the xTar event timepoint from each trial and computing their average, we achieved lower SNR values compared to the average of the original data. In addition, following this procedure, we were able to preserve the shape of the original ERP with its dominant components despite the FFT filtering.

In the following part of the analysis our objective is to determine if these similarities can be proved on the empirical level. This will be done by quantifying the similarity between the averages of the filtered ERP and the original data through the Kullback-Liebler (KL) Divergence. First, we begin by introducing this divergence, explaining how it is calculated, how to interpret the results and its different applications. Its application will allow us to investigate whether the functions that come from the traditional ERP approach and the FFT filtering converge towards the same function as the number of trials goes towards their maximum. Finally, we will conclude this analysis by examining which of these functions converges the fastest.

6.1 - Kullback-Leibler Divergence

Kullback-Leibler (KL) Divergence, also refer to as relative entropy (Csiszar, 2007), is a very important statistical distance used widely in information theory and statistics. It was introduced by Solomon Kullback and Richard Leibler in 1951 and measures the divergence between two probability distributions.

For two probability distributions P and Q defined on the same probability space, the KL measures the divergence from Q to P and is calculated as:

$$D_{KL}(P\|Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)}$$

Where:

- **P(x):** the target distribution,
- **Q(x):** the approximation distribution.

It quantifies the information loss when distribution Q is used to approximate the true distribution Q. It can be seen as simply measuring how different two distributions are.

Its value is always non-negative. The divergence is equal to zero if and only if the distributions are P = Q almost everywhere. In other words, when there is no information loss when Q is used to approximate P. Low values indicate low divergence, which can be interpreted as high similarity between distributions and small information loss. The opposite is true for large values of KL Divergence.

It is a relative measure, not an absolute one. This means that the value itself should be interpreted in context of the study. It can be compared to a chosen baseline or to compare among other KLD values.

KL divergence is asymmetric which means:

$$D_{KL}(P\|Q) \neq D_{KL}(Q\|P)$$

This is an important property to be taken into consideration when doing the analysis and interpreting the results as reversing the order of the distributions may lead to different results.

The KL Divergence finds several applications in Data Science such as: model selection and comparison, Information Gian in decision trees which indicates the best feature to split on, used as loss function in classification problems etc. It finds several uses in neuroscience and EEG data in cases such as: comparing EEG signals coming from different states, features extraction and classification, anomaly detection by comparing distribution of EEG signal during normal activity and during an anomaly like epileptic seizure etc.

6.2 - Convergence towards the same function

Before measuring KL Divergence, we need to pass through a preprocessing step which will be performed on the original data that will be used in this step. This is the baseline transformation process and is commonly used in EEG analysis, especially in studies about Event-Related

Potential (ERP). This procedure consists in subtracting the mean of a predefined pre-stimulus segment from each data point of that particular trial. In our case we measure the mean on each segment starting 200ms before the event time point and ending at the event xTar time point.

EEG signal usually contains low-frequency noise or artifacts that are not related to brain activity. By performing bassline transformation these effects are minimized, leading to a clearer observation of the response to the stimulus. By normalizing trials to their baseline, we can reduce the variability among trials which produces cleaner averages and more reliable ERP. It this context the baseline transformation helps to improve SNR.

To determine if the results from BT method and the FFT filter converge towards the same function as the number of trials increases towards 80, which is the maximum number of trials available for xTar event, we calculated the KLD for the first 10 trials, first 20, and so on up to 80 trials. The results are illustrated in the graph below (Figure 6.1). The y-axis represents the KLD value while the x axis the number of averaged trials in each step. The colored lines indicate the divergences for each subject, while the bold black like represents their average.

It is evident that for all subjects, as the number of trials increase the KL Divergence decreases. This means that the results coming from these two methods become more similar as the number of averaged trials increases.

Some subjects, such as subject 62, have an initial KLD value higher compared to other subjects, but we notice a rapid decrease as the number of trials increases, until reaching the final value of 0.053. In other cases, such as Subject 42, the KLD decline is more gradual.

The thick black line in the graph represents the mean KLD across all subjects. It shares the same trend as the individual subjects. This demonstrates that as the number of trials goes towards its maximum, the BT and FFT filtering functions become increasingly similar. It reaches the last point at the value of 0.089. Based on the theoretical knowledge where the KLD value of zero is achieved if and only if the two functions are identical almost everywhere, this low value suggests significant similarity between the results of these two approaches.

Being a relative measure, as mentioned before, its value needs to be interpreted in the context of the study. Usually a threshold is chosen with values between 0.1 and 0.2 considered low values indicating similarities between the distributions that are being compared.

Based on the general trend of each subject and also on the fact that even if there is a slight increase in KLD, it is followed by larger decrease, as in the case of Subject 55 at the 20 trials average level. And also seeing the KLD values for 80 trials averaging on all subject, with Subject 55 having the maximum value of 0.11, we can say that in general the two approaches lead to similar functions.

It is evident that even when there are slight increases in KLD values, such as the one observed at the 20-trial average level for Subject 55, they are followed by more significant decreases as the number of averaged trials increases. This is true for all subjects, where the KLD values at 80 trials are generally low. The maximal value of 0.11 belonging to Subject 55, which can be considered as relatively low. All these factors indicate that the Baseline Transform and FFT Filter approaches tend to converge towards similar functions as the number of goes towards their maximum.

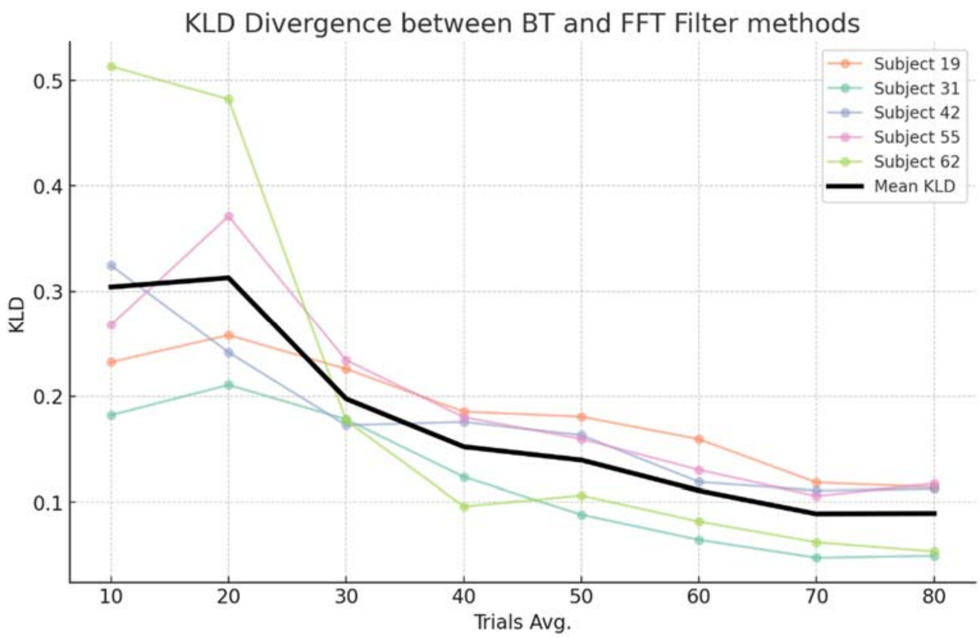


Figure 6. 1 – KLD values for each subject and their mean

6.3 - Which function converges the fastest

As previously stated, the main objective of this study is to identify a method that achieves an ERP with a signal-to-noise ratio (SNR), comparable to the obtained from averaging all trials of raw EEG data, with a reduced number of trails. This reduction will help improving the ecological validity of the ERP experiments by creating a more naturalistic experimental setting.

To address this, we compared two preprocessing approaches, Baseline Transform (BT) and FFT filtering, by evaluating their performance across different numbers of averaged trials (10, 20, …,80). The effectiveness of each method was assessed by measuring the Kullback-Leibler Divergence (KLD) between the ERP generated from each trial step and the ERP generated from the full 80-trial average of the same method. A lower KLD indicates greater similarity of that particular step to the 80-trial average, indicating a more effective method of reducing the number of trials.

The graphs below (Figure 6.2) display the Kullback-Leibler Divergence (KLD) between the ERPs obtained with different numbers of trial averages (10, 20, ..., 80) and the ERPs obtained with 80 trial averages of the same method. Each graph represents the KLD values for each subject coming from the BT and FFT filter methods. The x-axis shows the number of trials averaged in each step, while the y-axis the KLD values.

We notice that for all the subjects, the KLD values decrease as the number of averaged trials increases. This indicates that the ERPs coming from both methods become more similar to the 80-trial average for increased number of averaged trials. This is expected as we know that increased number of trials reduces noise and leads to more reliable ERPs.

In the first four graphs (Subjects 19, 31, 42 and 55) we notice that the FFT filter line is flatter when compared to the BT one. This means that in general the trials average coming from FFT filter in each step is more similar to its 80 trials average than BT averages are to their 80-trial average. This difference is more evident for smaller number of trials. While the number of trials increases, round 60 trials in all cases, we notice that KLD values coming from these two approaches tend to have similar values.

For subject 62 we notice that both KLD values start relatively high compared to the other cases, but they both decrease rapidly at a similar rate while the number of averaged trials increases. The BT method starts lower, indicating better performance and greater similarity to the final ERP. However, between 40 and 50 trials the FFT filter slightly outperforms the BT one.

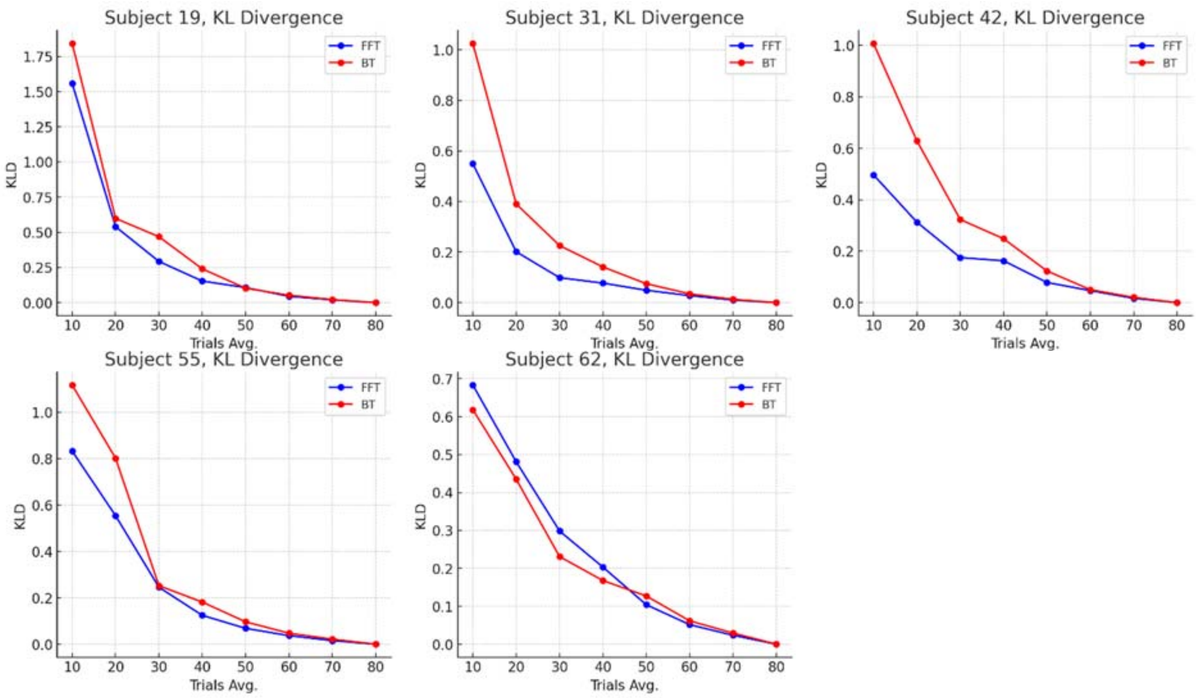


Figure 6. 2 – KLD for all subjects

The graph below (Figure 6.3) provides an overall view of the Kullback-Leibler Divergence (KLD) across all subjects. The blue line indicates the mean KLD values for all subjects coming from the FFT filter method. While the red line represents the KLD mean values of the Baseline Transform (BT) method. The shaded areas show the maximum and minimum KLD values for each level of averaged trials across all subjects. The red area indicating the FFT filter vales and the blue one the values coming from BT method.

The mean KLD values for both methods share the same characteristic of the individual components. They decrease as the number of averaged trials increases.

We notice that the KLD values for the FFT filter are consistently lower than the BT ones for each level of averaged trials. This is more evident for the first 40 trials where the gap is larger. This suggests that on average, the FFT filtering method achieves greater similarity to the 80-trial average with fewer trials, compared to the BT method.

The FFT method has also a slightly narrower shaded area compared to BT. This indicates more consistent performance across subjects.

It is interesting to notice that starting from the level of 20 averaged trials, the maximal KLD values for the FFT filter method, the upper side of the shaded blue area, has similar values to the mean of the BT method. On the other side, the best performance of the BT method shares similar values with the mean of the FFT filter method. Graphically this is represented by the lower edge of the shaded red area and the blue line. This means that even in the worst cases, the FFT filter performs as well as the average case of BT. FFT filter provides more consistently good results across different subjects and trial numbers. This consistency is important for ensuring reliable ERP quality with fewer trials.

For the studied dataset, FFT tends to reach lower KLD values with fewer trials which suggests that it allows for a faster and more efficient convergence toward the 80-trial average ERP when compared to the results of the BT method.

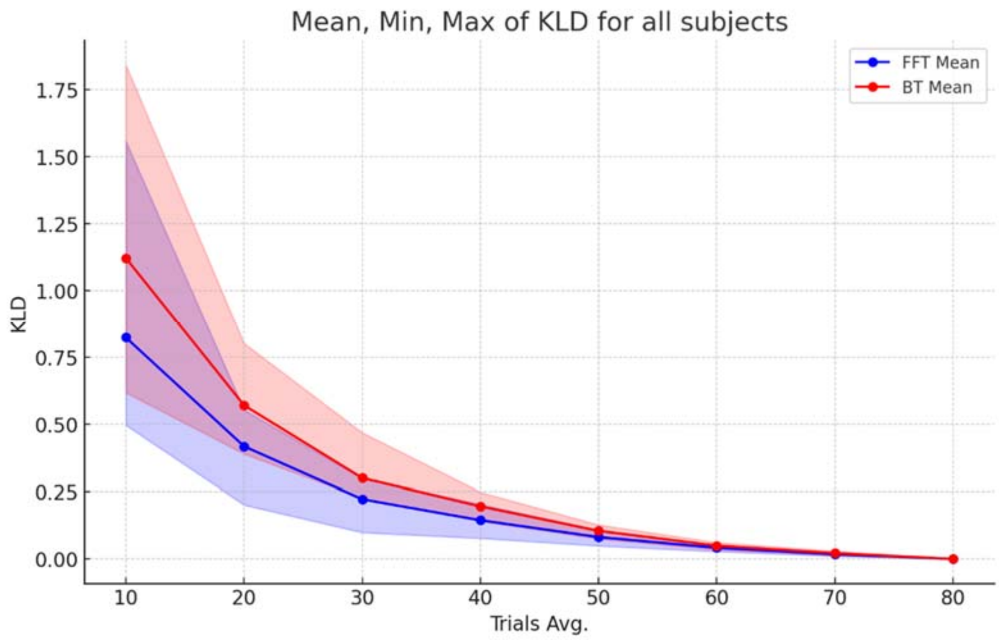


Figure 6. 3 – Min, **Max** and **Mean KLD for all subjects**

CHAPTER 7

Conclusion

The primary goal of this study was to provide guidelines and develop statistical methodologies for reducing the number of trials needed in ERP analysis while keeping similar levels of SNR. This aimed to help enhancing the ecological validity of ERP research. This approach aimed to quantify and remove the periodic part of trials that are used to calculate the ERP.

We began to analyze the periodicity in baseline data, which is not affected by any task related potential. For this task we fitted a sinusoidal regression along with its AIC and BIC derivatives. These models were fitted in data coming from different subjects, recorded on different electrodes, on segments with varying length and different stating points. The results indicated that the model's performance was similar and was not affected by all the above-mentioned criteria, apart from the signal length. The models were able to perform well for segment of 2-3 seconds ad after that the performance decreased significantly.

In the next step we applied our analysis in two adjacent segments of one second. More precisely the segments [-2; -1] and [-1; 0] with respect to the xTar event timepoint from the electrode Pz. We tried to quantify the periodic part of the [-1; 0] segment using the fitted values of several non-linear models on [-2; -1]. The poor results of this step proved what is known in theory that the statistical characteristics of EEG data tend to change over time.

This led us to shift our approach towards analyzing the segments that contain the induced potential of the event that we were studying. At this stage we applied a set of steps based on the methods of FFT and its inverse function. We applied this method to deconstruct the frequencies of the data segments and remove their dominant periodic function. In this part of the analysis we used 9 scenarios. We used 3 segments of different length, more specifically the segments [-2; 1]. [-1; 1] and [0; 1] in relation to the xTar event of electrode Pz. By applying the FFT filter method, we removed 2, 4 and 6 most powerful frequencies from each of these segments. The ERP was calculated for these new scenarios and the SNR values were compared to the ones of the original data. We discovered that by conducting the analysis for the [0; 1] segment we were able to improve the SNR, but it affected the ERP shape by removing its important features. Therefore, we increased the signal length to the left of the event timepoint in order to capture underlying periodicity that was present before and while the event took place. We were able to obtain similar levels of SNR for the ne ERPs while being able to preserve the components of the original ERP, components that are significant to further studies. By comparing the results coming from these 9 scenarios (3 signal length and 3 frequency removal cases), we decided that using the [-2; 1] segment while removing the 4 most powerful frequencies from the trials that studied, we were able to preserve the properties of the ERP while reducing significantly the noise, by having larger values for SNR.

The last step was performed with the purpose to quantify the similarities of the results that were obtained in the previous step. This was done by comparing the ERPs coming from the Baseline Transformed (BT) method and the FFT filter. Here we tried to answer two key questions:

1. **Do ERPs obtained through these two approaches converge towards the same function as the number of trials goes towards the maximum?**
   We measured the divergence using KL Divergence and observed that as number of trials increased, the values of their divergence decreased until going toward values close to 0.1. This result indicates high similarity between the two ERPs coming from the two methods. This once again showed what we deducted from the analysis in the previous step when we did graphic analysis.
2. **Which approach converges faster towards its 80-trial average?**
   This analysis showed that the FFT filter method outperformed the BT approach. Showing lower divergence towards its 80-trial average across different averaged trial counts. Faster convergence suggested that the FFT filter method is able to preserve the shape of the ERP while using fewer trials.

These results do not provide conclusive proof. They offer some indication that the FFT filtering approach has the potential to achieve better results than the traditional method. The ability to achieve similar levels of SNR with fewer trials, while preserving the ERP significant components, suggests that the FFT filter might contribute in enhancing the ecological validity of the experiments. However, these suggestions are only the first step and require further investigation.

It is important to mention the limitations of this study, especially the low representativity coming as a result of a small sample size from a single study. For future studies, it is important that the approach developed here could be generalized. Which means applying it in other datasets coming from different experiments, applied to different electrodes and tasks.

Additionally, exploring the possibility of other statistical models and approaches to quantify the periodic component of EEG data.

BIBLIOGRAPHY

Haas LF (January 2003). "Hans Berger (1873–1941), Richard Caton (1842–1926), and electroencephalography". Journal of Neurology, Neurosurgery, and Psychiatry. 74 (1): 9. doi:10.1136/jnnp.74.1.9

Riedl, R., and Léger, P.-M. (2016). Fundamentals of NeuroIS: Information Systems and the Brain, Berlin, Heidelberg: Springer.

Moran A (August 2, 2004). Sport and Exercise Psychology: A Critical Introduction. Hove, East Sussex: Routledge. p. 104. ISBN 9781134704101

Abhang Priyanka A., Gawali Bharti W. et Mehrotra Supresh C. (2016). Technological Basics of EEG Recording and Operation of Apparatus», Introduction to EEG- and Speech-Based Emotion Recognition, 1st edition, Massachussetts, Academic Press.

Dimoka, A., Pavlou, P. A., Davis, F. D. (2007). NEURO-IS: The potential of cognitive neuroscience for information systems research. Proceedings of the 28th International Conference on Information Systems, 1-20

Luck SJ (2005). An Introduction to the Event-Related Potential Technique. The MIT Press. ISBN 978-0-262-12277-1

van Dinteren R, Arns M, Jongsma ML, Kessels RP. P300 development across the lifespan: a systematic review and meta-analysis. PLoS One. 2014 Feb 13;9(2):e87347. doi: 10.1371/journal.pone.0087347

Kutas M, Federmeier KD (December 2000). "Electrophysiology reveals semantic memory use in language comprehension". Trends in Cognitive Sciences. 4 (12): 463–470. doi:10.1016/s1364-6613(00)01560-6

Kutas M, Federmeier KD (2011). "Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP)". Annual Review of Psychology. 62: 621–647. doi:10.1146/annurev.psych.093008.131123

Fredette, Marc & L. LeMoyne, Elise & Léger, Pierre-Majorique & Courtemanche, François & Senecal, Sylvain. (2015). Research Directions for Methodological Improvement of the Statistical Analysis of Electroencephalography Data Collected in NeuroIS. 10. 10.1007/978-3-319-18702-0_27.

Gratton, G.: Biosignal Processing. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G.G. (eds.) Handbook of Psychophysiology, 3rd edn., pp. 834–858. Cambridge University Press, Cambridge (2007)

Kieffer, Suzanne (2017-06-30). "ECOVAL: Ecological Validity of Cues and Representative Design in User Experience Evaluations". AIS Transactions on Human-Computer Interaction. 9 (2): 149–172. doi:10.17705/1thci.00093

Arfken, George (1985), Mathematical Methods for Physicists (3rd ed.), Academic Press, ISBN 9780120598205

Larsson PG, Eeg-Olofsson O, Lantz G. Alpha frequency estimation in patients with epilepsy. Clin EEG Neurosci. 2012 Apr;43(2):97-104. doi: 10.1177/1550059411433611

Masaki Nishida, Nobuhide Hirai, Fumikazu Miwakeichi, Taketoshi Maehara, Kensuke Kawai, Hiroyuki Shimizu, Sunao Uchida, Theta oscillation in the human anterior cingulate cortex during all-night sleep: an electrocorticographic study, Neuroscience Research, Volume 50, Issue 3, 2004, Pages 331-341, ISSN 0168-0102, https://doi.org/10.1016/j.neures.2004.08.004

Meyer, Yves (1992), Wavelets and Operators, Cambridge, UK: Cambridge University Press, ISBN 0-521-42000-8

Seung Hyun Lee; Berdakh Abibullaev; Won-Seok Kang; Yunhee Shin; Jinung An; "Analysis of Attention Deficit Hyperactivity Disorder in EEG Using Wavelet Transform and Self Organizing Maps", ICCAS 2010, 2010

H. Rajaguru and S. K. Prabhakar. Non linear ica and logistic regression for lassification of epilepsy from eeg signals. In Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of, volume 1, pages 577–580. IEEE, 2017.

Q. Dong, B. Hu, J. Zhang, X. Li, and M. Ratcliffe. A study on visual attention modeling—a linear regression method based on eeg. In Neural Networks (IJCNN), The 2013 International Joint Conference on, pages 1–6. IEEE, 2013.

J. Hu and Z. Mu. Eeg authentication system based on auto-regression coefficients. In Intelligent Systems and Control (ISCO), 2016 10th International Conference on, pages 1–5. IEEE, 2016.

A. Jalilifard, E. B. Pizzolato, and M. K. Islam. Emotion classification using single-channel scalp-eeg recording. In Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the, pages 845–849. IEEE, 2016.

G. Pfurtscheller and C. Neuper, “Motor Imagery and Direct Brain – Computer Communication,” vol. 89, no. 7, pp. 1123–1134, 2001

W. Mumtaz, S. S. A. Ali, M. A. M. Yasin, and A. S. Malik. A machine learning framework involving eeg-based functional connectivity to diagnose major depressive disorder (mdd). Medical & biological engineering & computing, 56(2):233–246, 2018.

V. Vijayakumar, M. Case, S. Shirinpour, and B. He. Quantifying and characterizing tonic thermal pain across subjects from eeg data using random forest models. IEEE Transactions on Biomedical Engineering, 64(12):2988–2996, 2017.

M.-P. Hosseini, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh. Random ensemble learning for eeg classification. Artificial intelligence in medicine, 84:146–158, 2018.

M.-P. Hosseini, S. Lu, K. Kamaraj, A. Slowikowski, and H. C. Venkatesh. Deep learning architectures. In Deep Learning: Concepts and Architectures, pages 1–24. Springer, 2020.

V. Asanza, K. Ochoa, C. Sacarelo, C. Salazar, F. Loayza, C. Vaca, and E. Peláez. Clustering of eeg occipital signals using k-means. In Ecuador Technical Chapters Meeting (ETCM), IEEE, pages 1–5. IEEE, 2016.

M. P. Hosseini, H. Soltanian-Zadeh, and S. Akhlaghpoor. Three cuts method for identification of copd. Acta Medica Iranica, pages 771–778, 2013.

S. Dähne, F. Bießmann, F. C. Meinecke, J. Mehnert, S. Fazli, and K.-R. Mtüller. Multimodal integration of electrophysiological and hemodynamic signals. In Brain-Computer Interface (BCI), 2014 International Winter Workshop on, pages 1–4. IEEE, 2014.

Gazzaniga, Michael S.; Ivry, Richard B.; Mangun, George R. (2009). Cognitive Neuroscience: The Biology of the Mind (2nd ed.).

Kirchner, W. K. (1958). "Age differences in short-term retention of rapidly changing information". Journal of Experimental Psychology. 55 (4): 352–358. doi:10.1037/h0043688

Mike X Cohen (2014). Analyzing Neural Time Series Data: Theory and Practice https://doi.org/10.7551/mitpress/9609.001.0001 ISBN (electronic): 780262319553 Publisher: The MIT Press Published: 2014

Charles Sherman; John Butler (2007). Transducers and Arrays for Underwater Sound. Springer Science & Business Media. p. 276. ISBN 9780387331393

Csiszar, I (February 1975). "I-Divergence Geometry of Probability Distributions and Minimization Problems". Ann. Probab. 3 (1): 146–158. doi:10.1214/aop/1176996454.