



Signals and Systems Project

INSTRUCTOR: PROF. HAMID AGHAJAN

SHARIF UNIVERSITY OF TECHNOLOGY

**Analysis of Phase Locking Value during Olfactory Stimulation
as a Biomarker for Alzheimer's Disease in EEG Signals**

Mahdi Tabatabaei

Spring 2023

Contents

1	Introduction	2
1.1	Neurodegenerative Diseases	2
1.2	Olfactory Dysfunction	2
1.3	Goal of the Project	2
2	Electroencephalography (EEG)	3
2.1	What is EEG?	3
2.2	Alzheimer's Disease	5
2.3	Frequency Bands of EEG	5
2.4	Sampling frequency	7
3	EEG Signal Processing	8
3.1	Task Definition	8
3.2	Data Description	8
3.3	Pre-Processing	9
3.4	Phase Locking Value (PLV)	23
4	Results	27
4.1	Values	27
4.2	Distributions	27
4.3	Statistical Significance	27
4.4	Phase Difference	27
4.5	Heatmaps	27
5	*Bonus	40
5.1	Mild Cognitive Impairment (MCI)	40
5.1.1	Additional Information	40
5.1.2	MCI Data Processing	42
5.2	Phase-Amplitude Coupling (PAC)	51
5.2.1	Metrics	51
5.2.2	Implementation	52
6	Conclusion	56

1 Introduction

1.1 Neurodegenerative Diseases

Neurodegenerative diseases, including Alzheimer's Disease (AD) and Mild Cognitive Impairment (MCI), pose significant challenges to individuals, families, and healthcare systems worldwide. These conditions are characterized by progressive deterioration of cognitive functions, leading to severe impairment and loss of independence. Understanding the importance of these diseases and the urgency to find reliable biomarkers for their early detection and accurate diagnosis is critical for effective intervention and disease management and has become a priority in the field of neurodegenerative research.

Alzheimer's disease, the most common form of dementia, affects millions of individuals globally, and its prevalence is expected to rise with the aging population. MCI, often considered a transitional stage between normal aging and AD, is characterized by subtle cognitive decline that does not severely impact daily functioning. Scientists recognize that the combination of a person's genes and environment contributes to their risk of developing a neurodegenerative disease. For example, someone might have a gene that makes them more susceptible to Parkinson's disease, but their environmental exposures can affect whether, when, and how severely they are affected. [3]

1.2 Olfactory Dysfunction

The sense of smell is today one of the focuses of interest in aging and neurodegenerative disease research. In several neurodegenerative diseases, such as Parkinson's disease and Alzheimer's disease, the olfactory dysfunction is one of the initial symptoms appearing years before motor symptoms and cognitive decline which manifests as a decreased ability to detect, identify, or differentiate odors and thus, being considered a clinical marker of these diseases' early stages and a marker of disease progression and cognitive decline. [4]

One of the primary reasons olfactory dysfunction is prominent in neurodegenerative diseases is the presence of pathological changes in the olfactory system. In AD, for example, amyloid plaques and neurofibrillary tangles, the hallmark pathological features of the disease, are found not only in brain regions associated with memory and cognition but also in areas involved in olfaction, such as the olfactory bulb and olfactory cortex.

1.3 Goal of the Project

Understanding the significance of olfactory dysfunction in neurodegenerative diseases is important as it can serve as a potential biomarker for early detection and help unravel underlying disease mechanisms. The study of olfactory dysfunction in neurodegenerative diseases is an active area of research. Researchers are investigating the potential of olfactory testing as a diagnostic tool and exploring the mechanisms underlying olfactory dysfunction. They are also examining the role of olfactory dysfunction in disease progression and exploring therapeutic interventions targeting the olfactory system.

In this project, we want to identify early biomarkers for related brain disorders through olfactory stimulus.

2 Electroencephalography (EEG)

2.1 What is EEG?

There are different tools for collecting data from the brain. One of the methods of capturing brain signals is called Electroencephalography (EEG). These signals are changes in voltage level caused by changes in brain signals captured by some electrodes. These voltages are microVolt-level, so they can be sensitive to small noises.

One of the EEG advantages compared to other methods is its high temporal accuracy (i.e. high sampling frequency) while it suffers from low spatial accuracy. Another benefit of EEG devices is their smaller size compared to other devices like fMRI (functional Magnetic Resonance Imaging). While fMRI devices occupy the whole room, you can use EEG via portable devices.

EEG headsets are devices built to save EEG signals. These headsets could contain many electrodes. One internationally recognized electrode placement method is the **10-20 system**. This method was developed to maintain standardized testing methods ensuring that a subject's study outcomes (clinical or research) could be compiled, reproduced, and effectively analyzed and compared using scientific methods. It is called 10-20 because the distance between adjacent electrodes is 10% or 20% of the skull's total front-back or right-left distance.

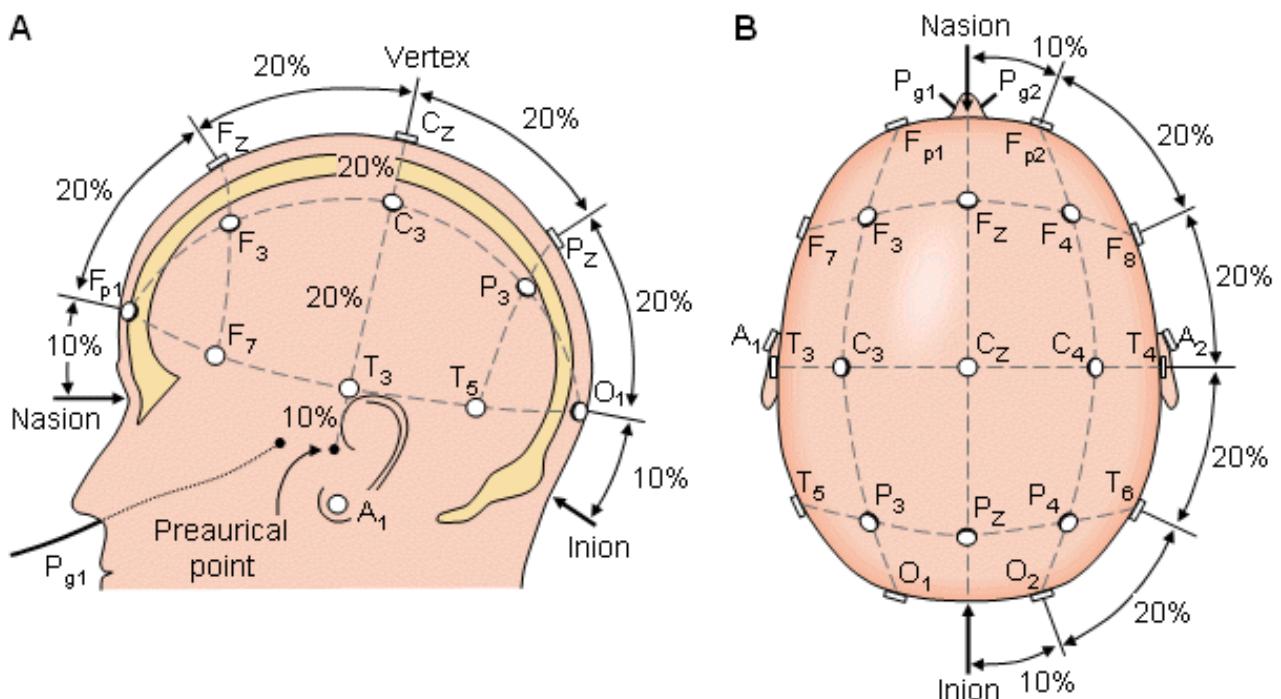


Figure 1: EEG 10-20 Electrode Placement System

Based on the picture above, What does each electrode's name stand for? Explain the naming method used in the 10-20 EEG system.

Answer**Letters in 10-20 System**

The letter of the electrode stands for the general brain region that the electrode covers. From front to back, the electrode letter labeling is as follows:

- Fp (pre-frontal or frontal pole)
- F (frontal)
- C (central line of the brain)
- P (parietal)
- T (temporal)
- O (occipital)

Electrodes lying between these lines combine multiple letters, ordered from front to back. In addition, the letters M and A are sometimes used to refer to the mastoids or earlobes respectively. Typically, these locations are included to serve as a (offline) reference for signal analysis.

Numbers in 10-20 System

The number of the electrode gives information about the distance from the electrode to the midline of the brain. At the midline, the electrodes are labeled with a 'z' to represent zero. The electrode numbers increase as you move away from the midline. Odd numbers represent electrodes on the left hemisphere and even numbers represent electrodes on the right hemisphere.

2.2 Alzheimer's Disease

Alzheimer's Disease (AD) is a progressive and irreversible neurological disorder that affects the brain, primarily causing problems with memory, thinking, and behavior. It is the most common cause of dementia, a general term for a decline in cognitive ability severe enough to interfere with daily life.

The exact cause of Alzheimer's disease is not yet fully understood, but it is believed to involve a combination of genetic, lifestyle, and environmental factors. The staging of the AD is associated with the accumulation of Amyloid- beta ($A\beta$) proteins in the brain. These depositions cause synaptic and neuronal loss, which leads to major cognitive dysfunction in the advanced levels of the disease.

While EEG is not currently used as a primary treatment for Alzheimer's disease, it can be a valuable tool in the diagnosis and monitoring of the disease. EEG can help in the diagnosis of Alzheimer's by detecting abnormal patterns of brain activity that are characteristic of the disease. In individuals with AD, EEG often shows changes such as a reduction in certain brainwave frequencies and an increase in others. These patterns can aid in differentiating Alzheimer's from other types of dementia or cognitive disorders.

2.3 Frequency Bands of EEG

In the frequency domain, EEG signals are divided into 5 bands. [2]

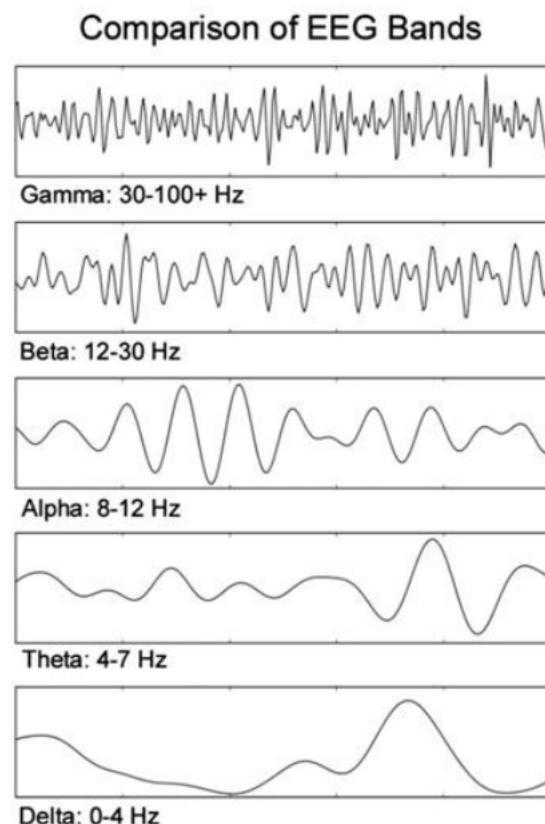


Figure 2: EEG Frequency Bands

Determine the activities each frequency band is associated with.

Answer

In EEG analysis, different frequency bands are associated with specific neural activities and cognitive processes. Here are the commonly recognized frequency bands and their corresponding activities:

- **Delta (0.5 - 4 Hz):** Delta waves are typically observed in deep sleep and unconscious states. They are associated with slow-wave sleep, brain development in infants, and some abnormal brain conditions.
- **Theta (4 - 8 Hz):** Theta waves are commonly observed during drowsiness, relaxation, and light sleep. They are also associated with creative thinking, daydreaming, and memory retrieval.
- **Alpha (8 - 12 Hz):** Alpha waves are prominent when an individual is awake but in a relaxed state with eyes closed. They are associated with a relaxed and calm mental state, as well as meditation and visualization.
- **Beta (12 - 30 Hz):** Beta waves are dominant during active and focused cognitive processing, such as attention, concentration, problem-solving, and decision-making. They are also associated with states of arousal and mental alertness.
- **Gamma (30 - 100+ Hz):** Gamma waves are the fastest and highest-frequency brain waves. They are associated with higher-order cognitive functions, such as perception, attention, memory encoding and retrieval, and conscious awareness. Gamma oscillations are often observed during tasks requiring integration of information from different brain regions.

It's important to note that these frequency bands are general guidelines, and there can be variations in the specific frequency ranges depending on the context and individual differences. Additionally, the interpretation of frequency bands can vary across different research studies and applications. Advanced EEG analysis techniques, such as event-related spectral perturbation (ERSP) or time-frequency analysis, can provide more detailed insights into the dynamics of brain activity across different frequency bands during specific tasks or conditions.

2.4 Sampling frequency

Based on frequency bands and Nyquist criterion, which sampling frequencies are preferred for EEG signals?

Answer

For EEG signals, the Nyquist criterion states that the sampling frequency should be at least twice the highest frequency of interest in the signal in order to accurately capture the information. Based on the frequency bands typically observed in EEG signals, the following sampling frequencies are preferred:

- **Delta Band (0.5 - 4 Hz):** A minimum sampling frequency of 8 Hz is preferred.
- **Theta Band (4 - 8 Hz):** A minimum sampling frequency of 16 Hz is preferred.
- **Alpha Band (8 - 12 Hz):** A minimum sampling frequency of 24 Hz is preferred.
- **Beta Band (12 - 30 Hz):** A minimum sampling frequency of 60 Hz is preferred.
- **Gamma Band (30 - 100+ Hz):** A minimum sampling frequency of 200 Hz or higher is preferred.

It's important to note that these are minimum preferred sampling frequencies based on the Nyquist criterion. In practice, higher sampling frequencies are often used to capture additional high-frequency components and to minimize the impact of anti-aliasing filters. Additionally, the specific requirements may vary depending on the application and research goals.

For EEG signals, the preferred sampling frequencies depend on the frequency bands of interest and the Nyquist criterion. Recommended sampling frequency would be greater than 200 Hz.

3 EEG Signal Processing

In this section, firstly you would get familiar with the task and the structure of the data.

3.1 Task Definition

[7] To identify the effect of olfactory dysfunction among different brain health states, the following task was performed to collect the data. The same sequence of stimuli was presented to all participants. The stimulation sequence was composed of two different odors, one occurring frequently (standard) with a probability of 0.75 and the other presented rarely (deviant) with a probability of 0.25. Each trial consisted of a 2s stimulus presentation followed by 8s of rest (pure water vapor). The odors were delivered to the participants using a laboratory olfactometer. The experiment involved 120 trials in which 90 frequent and 30 rare stimulation cycles were presented in a predetermined, randomized order. Lemon essence was used as the frequent odorant and rose essence was used as the rare odorant. These odors were selected to avoid trigeminal system activation as the olfactory and trigeminal systems are interconnected and may interact with each other during exposure to certain stimuli [8]. The duration of odor presentation was set at 2s to enable regular breathing cycles for the participants.

3.2 Data Description

[7] The dataset consists of three files as follows:

- **AD.mat**: Contains data for Alzheimer's disease patients.
- **Normal.mat**: Contains data for healthy elderly participants.
- **MCI.mat**: Contains data for mild cognitive impairment patients. (Described in part 5.1)

The structure of the files is the same. Each file is organized as a structure array, in which each row contains information of one participant and the three columns correspond to the “epoch”, “odor” and “noisy” fields as described in Table 1.

Field	Description
epoch	This is a 3D array structured as $4 \times 600 \times \text{Num_trials}$. The first dimension indicates EEG channels respectively from the first column as Fp1, Fz, Cz, and Pz. The second dimension contains EEG samples from 1 s pre stimulus to 2 s post stimulus, which at a 200 Hz sampling rate amounts to 600 samples. The last dimension shows the number of trials. This could be different for each participant as some trials were deleted during preprocessing.
odor	This is a 2D binary array shaped as $\text{Num_trial} \times 1$. This array shows the odorant type (lemon/rose) the participant was exposed to in each trial. The value = 1 indicates the rose odor and the value = 0 indicates the lemon odor.
noisy	This is a 2D array with the size $1 \times \text{Num_noisy}$. This array indicates noisy trials identified based on comparing the instantaneous and average trial amplitudes. These noisy trials can be ignored in processing and were included for the dataset completeness.

Table 1: Description of each structure array (.mat file) in the dataset.

3.3 Pre-Processing

Using a standard pipeline in EEG signal preprocessing is crucial for ensuring consistency, reproducibility, and objectivity in research. It reduces bias, enhances the reliability of results, and provides established best practices for addressing common challenges. A popular and widely used pipeline for EEG signal preprocessing is Makoto's pipeline ([Makoto's preprocessing pipeline - SCCN](#)).

The collected raw data from all participants were preprocessed following the full pipeline of Makoto with the use of EEGLAB and posted as a dataset, as described in the following steps:

1. Apply 1 Hz high pass filter to remove baseline drifts.
2. Apply relevant notch filter to remove the 50 Hz line noise.
3. Reject bad channels as a critical step before average referencing with the use of `clean_rawdata()` EEGLAB plugin.
4. Interpolate the removed channels.
5. Re-reference the data to the average of all channels to obtain a good estimate of reference-independent potentials.
6. Apply `clean_rawdata()` for cleaning the data by running artifact subspace reconstruction(ASR).
7. Re-reference the data to the average again to compensate for any potential changes in the data caused by the previous step.

8. Run independent component analysis (ICA) to identify EEG sources as well as the sources associated with noise and artifacts.
9. Fit single and bilateral (if available) current dipoles.
10. Further clean the data by source (dipole) selection using `ICLabel()` plugin in EEGLAB.

In the `Dataset/Preprocess` folder you can find the raw data for 2 subjects with the corresponding additional information provided. In this section you are required to preprocess these data and save your final preprocessed cleaned data.

However, there is no need to fully implement the Makoto's pipeline and a simplified version of this is as follows; follow the instructions below and provide the required results in each step:

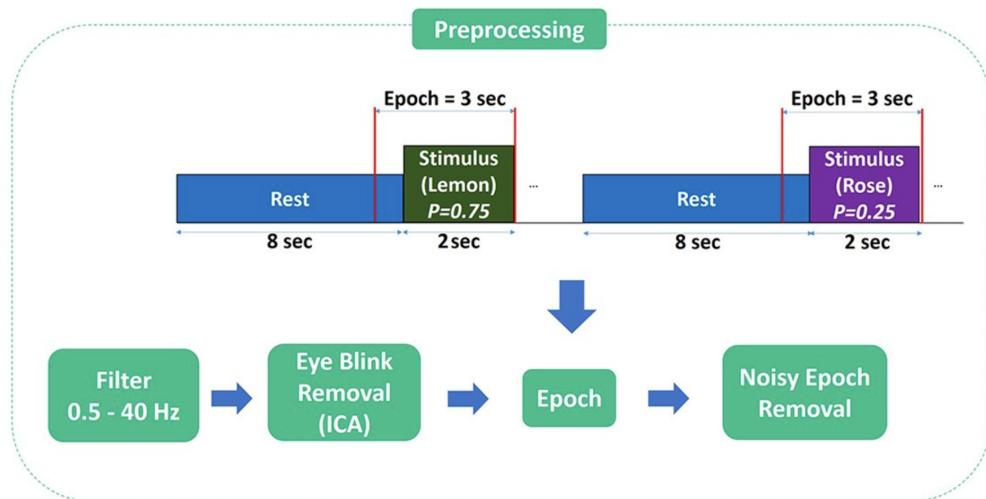
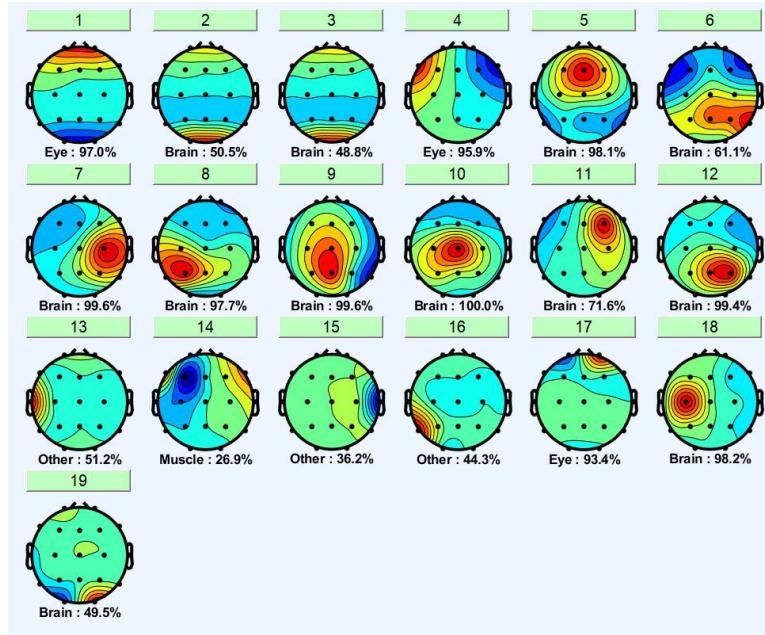


Figure 3: Task and Preprocessing Steps [5]

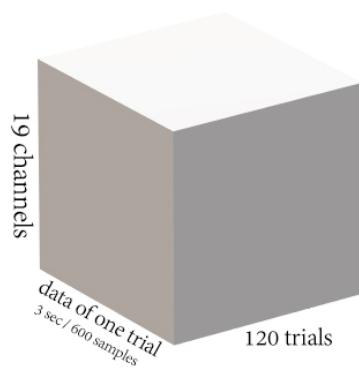
- **Step 1:** To preprocess using EEGLAB, first re-reference data to the mean of the channels. Then use a bandpass filter to filter 0.5 - 40.5 Hz frequencies. As we have filtered to 40.5 Hz, there is no need to apply a 49.9 - 50.1 Hz notch filter to remove the line noise (However, keep this step in mind as this is a crucial step in EEG signal preprocessing!). Using FFT function or EEGLAB, plot the frequency spectrum of Fz channel data. (Just to note, your data will be saved at EEG struct in MATLAB workspace.)
- **Step 2:** In this part you would remove the artifacts of the signal. Artifacts include blinking, eye movement, muscle movement, heart rate and etc. For this, load your data at EEGLAB. Now load `Standard-10-20-Cap19.loc` file from `edit-channel` locations menu that contains locations of channels. Then run ICA (Independent Component Analysis) algorithm from `tools-decompose` data by ICA menu. Please note that this part would probably takes more time. Then you will have the a figure like [Figure 4](#) by running `tools → classify components using ICLabel-label components`. By clicking on

**Figure 4:** An example of ICA components

each component, you can see some details about it as well. Present a figure from one of the brain components with its details.

Now remove all non-brain components. For this purpose, from `Tools-remove components from data` enter the number of components that must be removed.

- **Step 3:** Epoch the data of each subject. Epoch is a 3D matrix of the shape $\{\text{Num_Channels} \times \text{Samples} \times \text{Num_Trials}\}$. In fact, all data must be reshaped as the following figure suggests:

**Figure 5:** Epoch

For epoching the data, starting point of the experiment is required. This is provided in

the `help` file for each subject. Please note that you must epoch the data by considering this time as the start. Also, the data after 120 trials should be neglected as well.

- **Step 4:** In this step, you need to remove noisy trials. There are two ways to achieve this:

1. Observe data at EEGLAB and remove any trial that seems noisy. (PREFERRED!)
2. Using power spectrum of each trial, remove trials that their standard deviation of their power spectrum is bigger than 3.5 .Create a 3D matrix by each trial's power spectrum for each channel using `pspectrum` in MATLAB. You can use the following commands to find noisy trials:

```
vr = sum(nanstd(p,[],2).^2,2);  
noisy_trials = find(abs(zscore(vr))>3.5);
```

In these commands, `p` is a matrix of frequency spectrums of all trials of a channel. `noisy_trials` contains the number of noisy trials of that channel. These commands must run for each channel individually and the resultant noisy trials must be accumulated over all channel. Then remove all `noisy_trials` from your epoch.

- **Step 5:** In the final step, only subsample the data corresponding to the `Fp1`, `Fz`, `Cz` & `Pz` channels. You can find the channels' orders in the `Channels.jpg`.

Do these 5 steps for each subject and save the final data through an `struct` with the same format as described in [Table 1](#). Also, consider the order of `odor` being the same as the ones used for normal participants.

Step 1

*Before we start to explain, it is important to clarify that the steps will be described for Subject1, while the output will be shown for both Subject1 and Subject2.

First, we import the .mat file into EEGLAB and then remove the 20th channel from the data. Although there were some challenges during this process, I successfully resolved them by utilizing the following MATLAB Function. The code allowed me to convert the data into a MATLAB variable, remove the 20th channel. Then I imported the modified data into EEGLAB.

```
function subject = load_and_remove_20th_Channel(path)
    % Load the data from the specified path
    data = load(path);

    % Convert the data table to an array and transpose it
    subject = table2array(data.subject)';

    % Remove the 20th channel from the data
    subject = subject([1:19], :);
end
```

Now that we have imported the data into EEGLAB, we proceeded with the referencing of the data to the mean of the channels. Subsequently, a [bandpass filter](#) was applied to the data. Our next step involves plotting the frequency spectrum of the **Fz** channel data, which corresponds to channel No.5. We will achieve this by utilizing the FFT (Fast Fourier Transform) algorithm and the following MATLAB function:

```
function [f, P1] = fftFunc(X, Fs)
    L = length(X); % Length of the input data
    Y = fft(X); % Perform the FFT
    P2 = abs(Y/L); % Two-sided spectrum
    P1 = P2(1:L/2 + 1); % One-sided spectrum
    P1(2:end-1) = 2*P1(2:end-1); % Multiply by 2 (except the
        % DC component and Nyquist frequency)
    f = (0:(L/2))*Fs/L; % Frequency axis
end
```

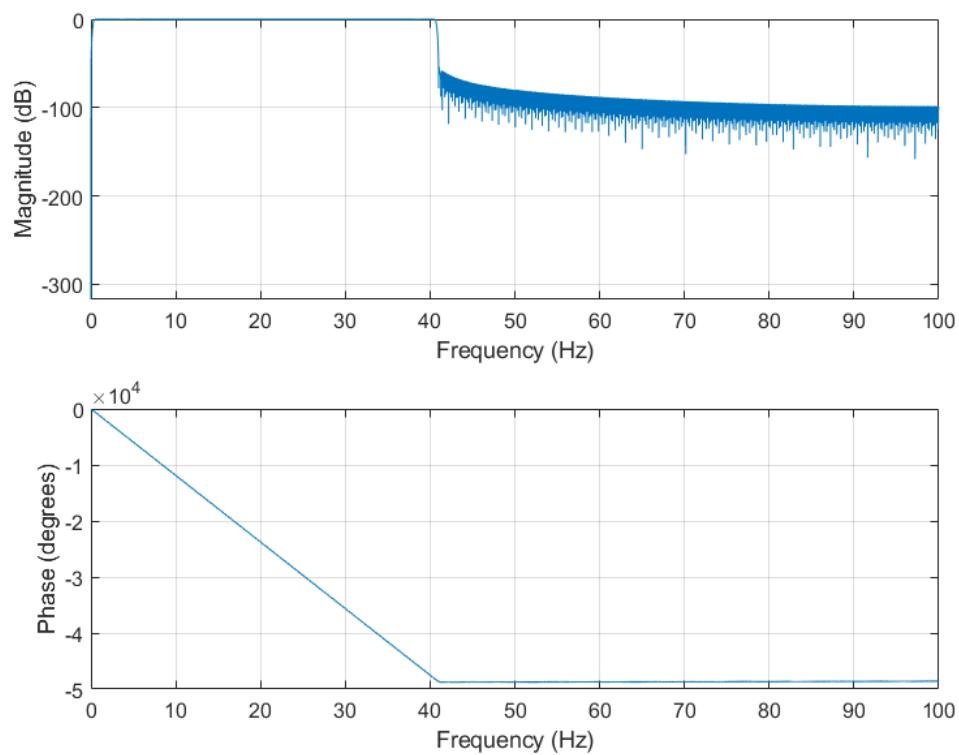


Figure 6: Bandpass Filter(0.5Hz - 40.5Hz)

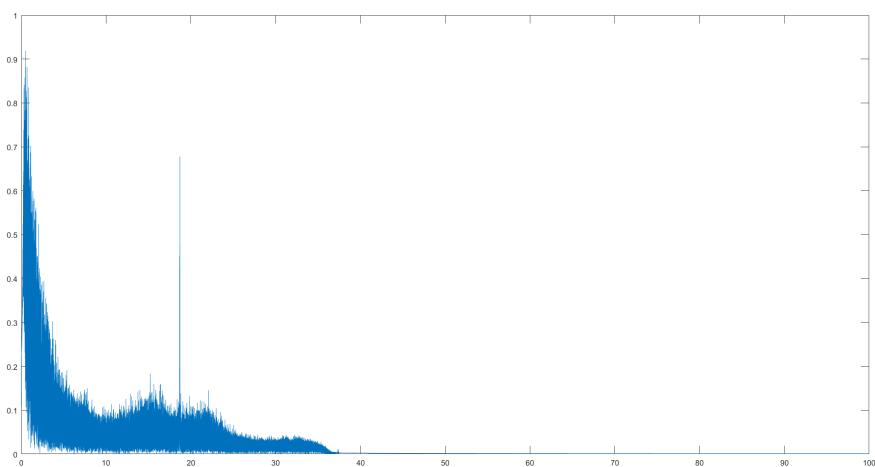


Figure 7: frequency spectrum of channel **Fz** of Subject1

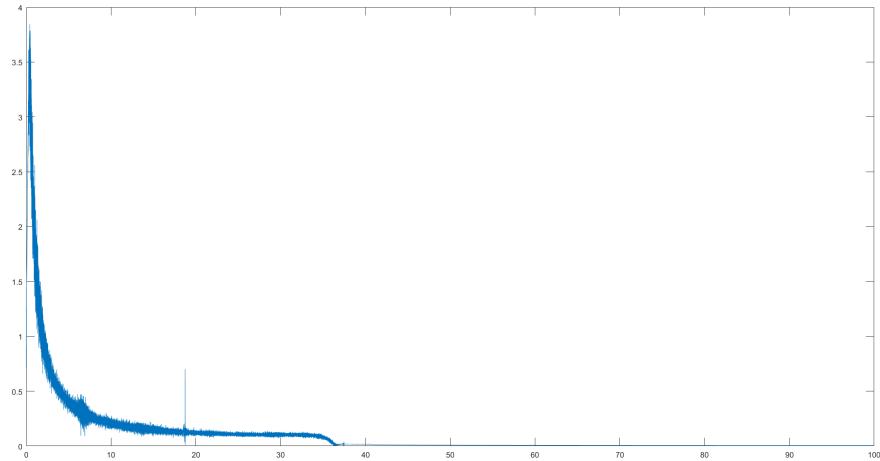


Figure 8: frequency spectrum of channel **Fz** of Subject2

Step 2

In this step, we begin by setting the location of the channels. The 2-D and 3-D locations of the channels can be found in figure 11. Next, we apply the ICA (Independent Component Analysis) algorithm and classify the components using labels. By clicking on each component, we can access detailed information about it. As observed, the rate of decrease in [figure 12](#) and [figure 14](#) with respect to frequency is inversely proportional to $1/f$.

Afterwards, we proceed to remove non-brain components using EEGLAB. This step helps to eliminate artifacts and isolate the brain-related signals of interest, enhancing the quality of the data for further analysis.

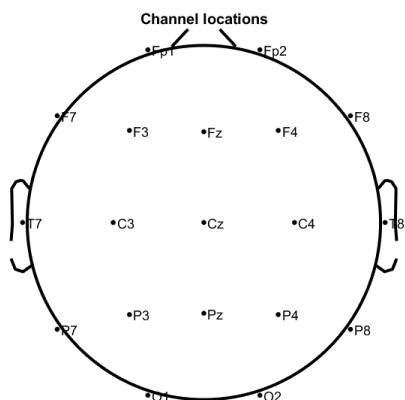


Figure 9: 2-D Channel Location

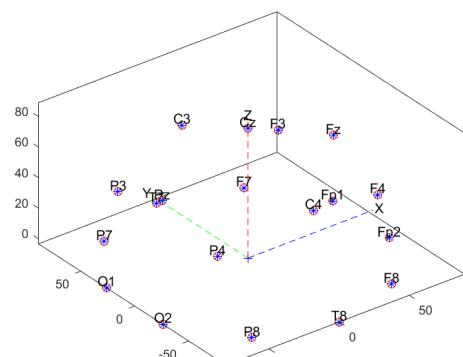


Figure 10: 3-D Channel Location

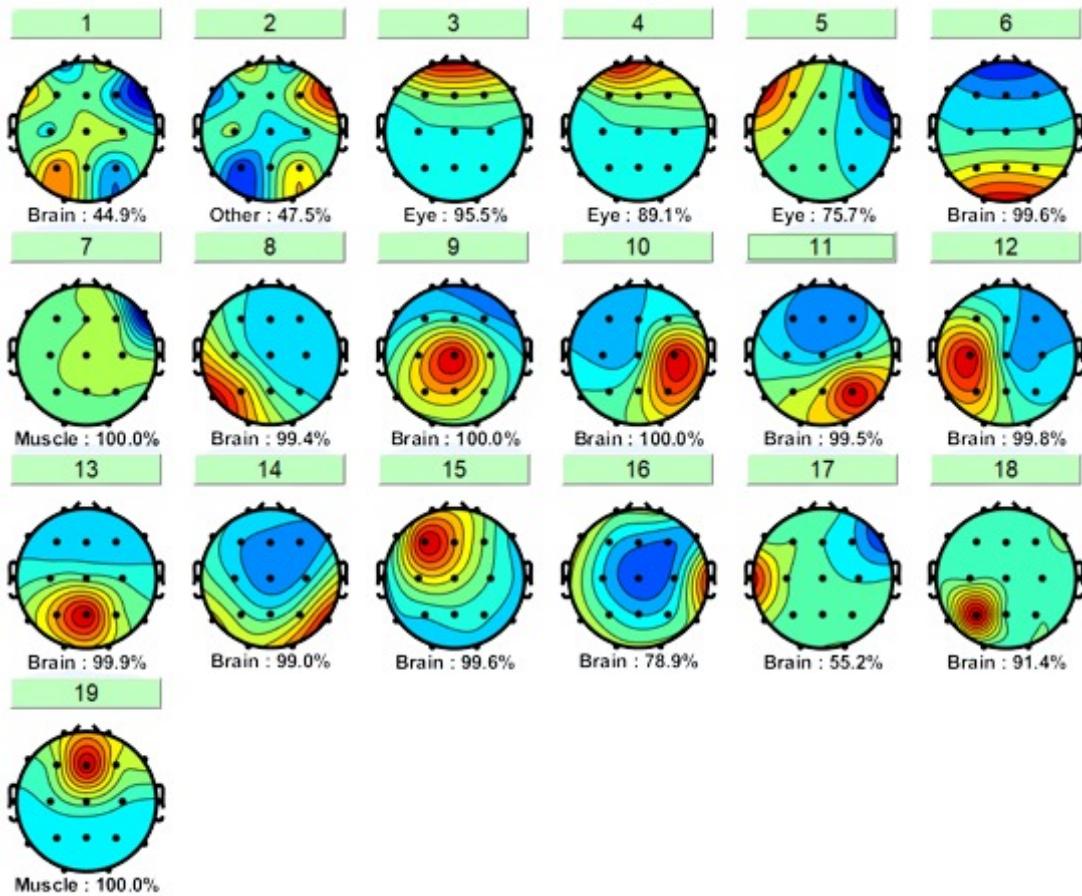


Figure 11: ICA Components of Subject1

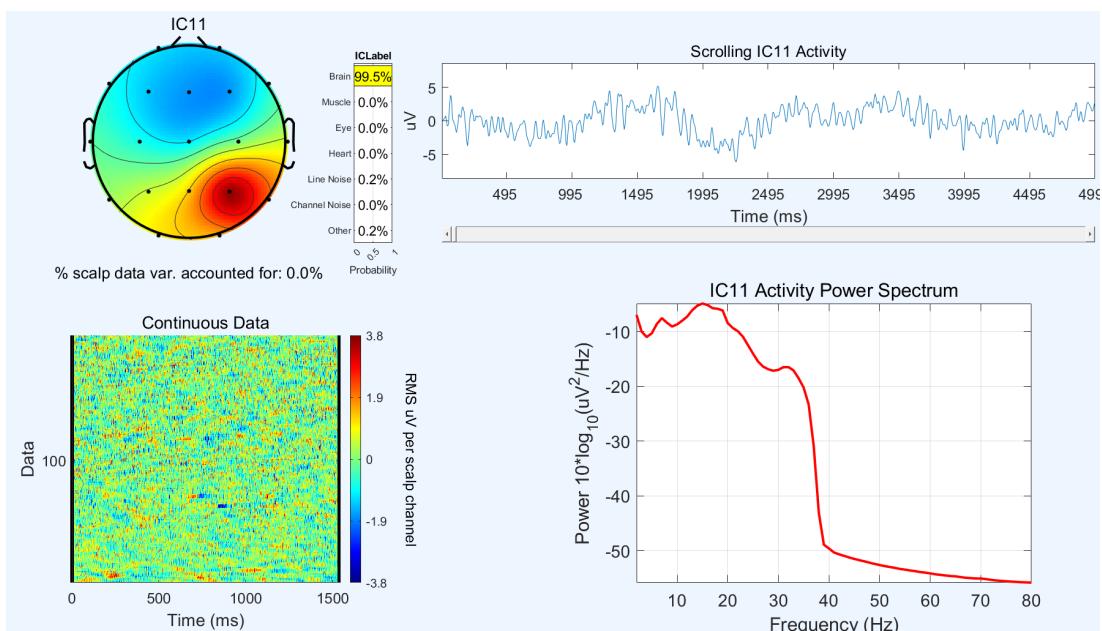
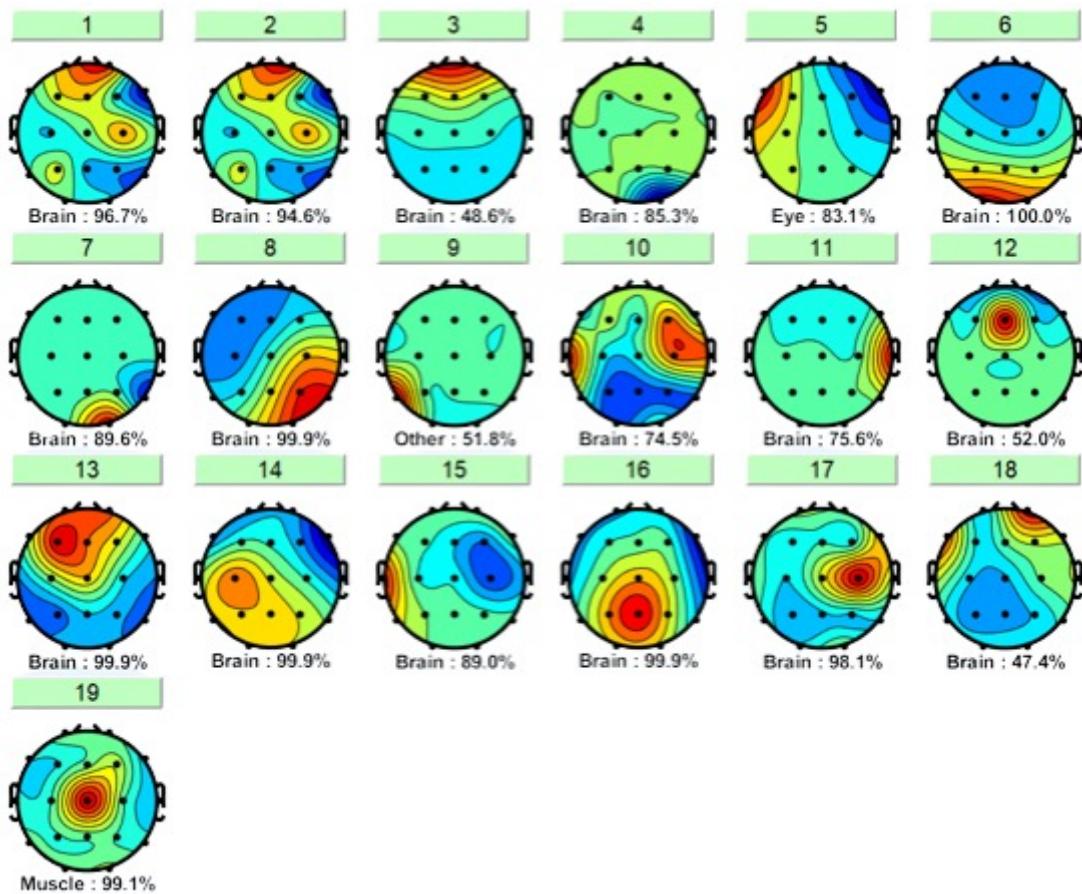
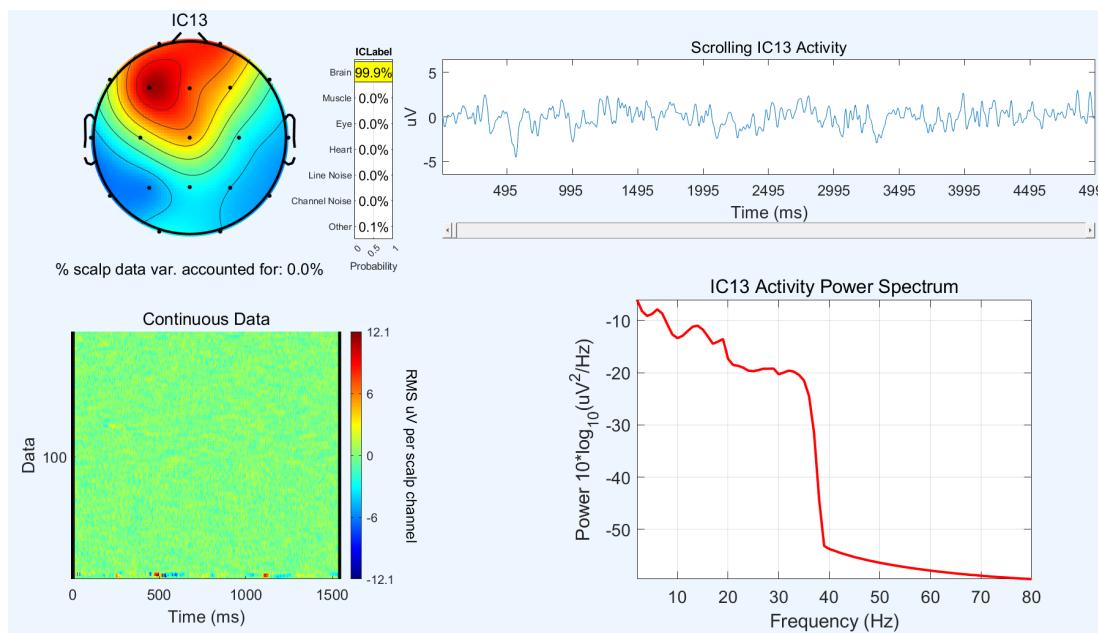


Figure 12: IC11 of Subject1's ICA Components

**Figure 13:** ICA Components of Subject2**Figure 14:** IC13 of Subject2's ICA Components

Step 3

To epoch the data, I used following Matlab Function:

```

function epoches = epochData(data, initTime, epochDuration,
    samplingRate, numTrials)

    epochSamples = epochDuration * samplingRate; % Number of
        samples per epoch (600)

    numChannels = size(data, 1); % Number of channels (19)

    epoches = zeros(numChannels, epochSamples, numTrials); %
        Initialize the epoch data matrix

    initSample = initTime * samplingRate; % Initial sample
        index

    % Iterate over channels, trials, and samples to extract
        epochs
    for i = 1:numChannels
        for j = 1:numTrials
            % Calculate the start and end indices of the epoch
            startSample = initSample + (samplingRate * 10 * j) -
                epochSamples + 1;
            endSample = initSample + (samplingRate * 10 * j);

            % Extract the data for the current epoch
            epoches(i, :, j) = data_to_epoch(i, startSample:
                endSample);
        end
    end
end

```

We should note that the starting time of each subject is not zero, and we need to begin epoching the data from the specified initial time. It is also possible to remove the data before the starting time in the first step itself.

After that we have to import data in EEGLAB again and do the rest of steps.

We could have used the built-in event-related epoching functionality in EEGLAB, which is designed to handle such tasks. However, since we don't have access to specific event markers in the data, we needed to perform the epoching process manually without relying on the built-in functionality provided by EEGLAB.

Step 4

This step is crucial as it involves identifying and removing noisy trials from the data. Detecting and removing noisy trials is essential to ensure the quality of the EEG data analysis. Noisy trials can arise due to various factors such as blinking, body movement, and other sources of artifacts.

We also use EEGLAB to remove noisy trials.

I also attempted to use second algorithm to automatically detect and remove noisy trials. However, I found that this algorithm did not yield satisfactory results and was not efficient in identifying all the noisy trials accurately. As a result, I resorted to relying on my own observation and expertise to detect and remove the noisy trials.

In the analysis of Subject 1, we identified a total of 4 noisy trials, while in the analysis of Subject 2, we detected 36 noisy trials. The trial numbers of these noisy trials are stored in a data struct, which will be created in the next step of our analysis.

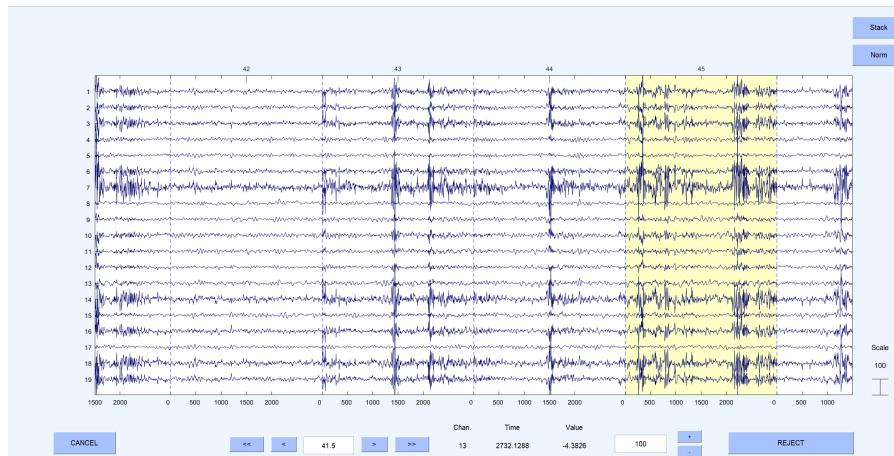


Figure 15: Some Noisy Trial of Subject1(Trial No.45)

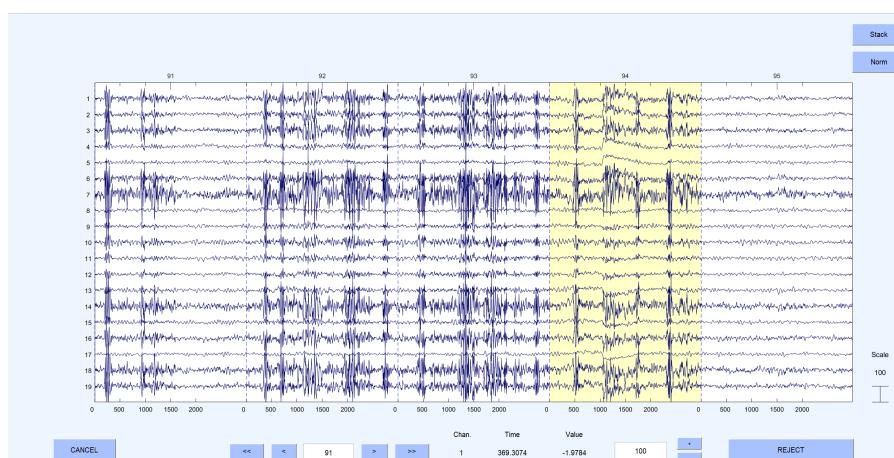


Figure 16: Some Noisy Trial of Subject1(Trial No.94)

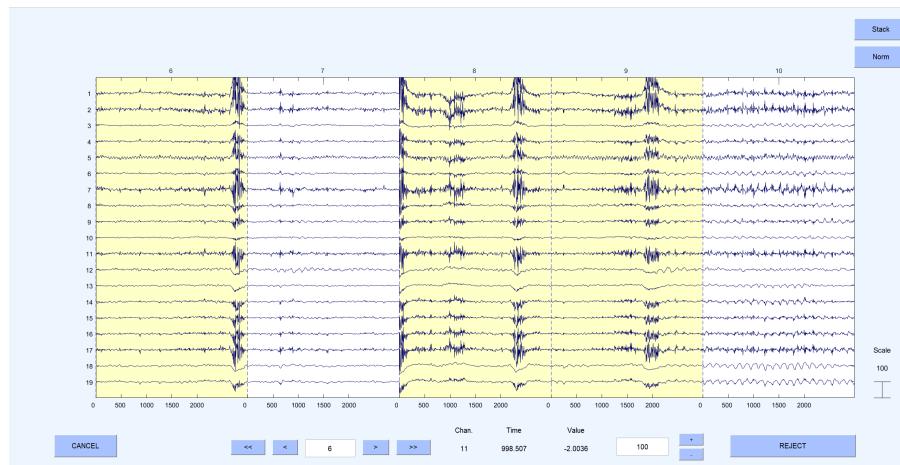


Figure 17: Some Noisy Trial of Subject2(Trial No.6,8,9)

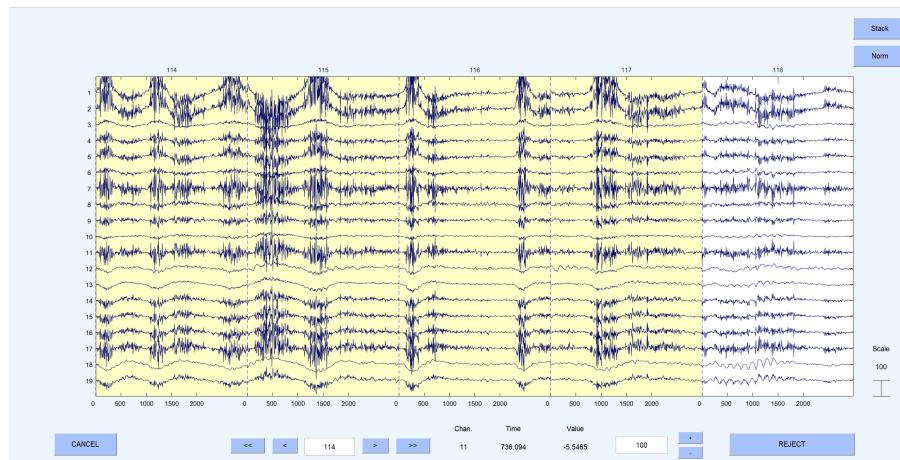


Figure 18: Some Noisy Trial of Subject2(Trial No.114-117)

Step 5

In the final step of the preprocessing stage, it is necessary to remove all channels except for FP1, Fz, Cz, and Pz. These specific channels correspond to channel numbers 1, 5, 10, and 15, respectively. To accomplish this, I utilized a MATLAB function that effectively removes all other channels from the dataset, retaining only the desired channels for further analysis.

```
function removeChannels(epochedData, filePath)
    % Input:
    % - epochedData: The epoched data in the format [
    %     NumChannels x NumSamples x NumTrials]
    % - filePath: The file path to save the processed data

    channelsIdx = ismember(1:19, [1, 5, 10, 15]); % Indices
    % of the desired channels

    selectedEpochData = epochedData(channelsIdx, :, :); % Select the desired channels

    save(filePath, 'selectedEpochData'); % Save the
    % processed data
end
```

By selectively removing unwanted channels, we can streamline the data and focus solely on the relevant electrode locations of interest. This step is crucial for optimizing subsequent analysis steps and ensuring that our results are specific to the targeted electrode sites.

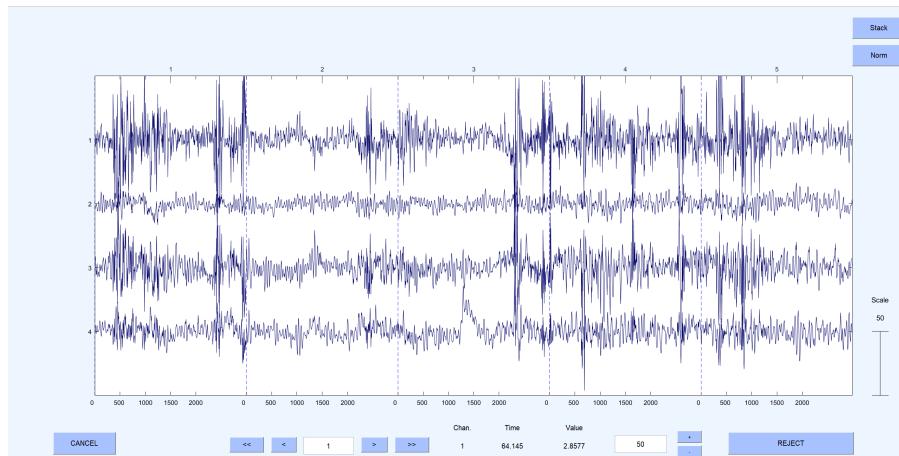


Figure 19: Output of Subject 1 with 4 channels

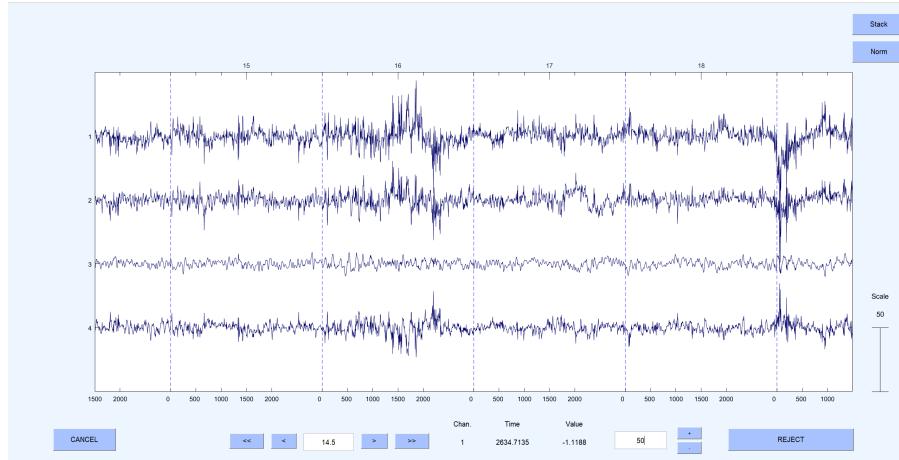


Figure 20: Output of Subject 2 with 4 channels

End Of Steps

In the End, We need to store data like [Table 1](#) in a struct. In order to do it, we use following matlab function:

```

function myStruct = makeStruct(selectedEpochData ,
noisyTrials , filePath)
% Input:
% - selectedEpochData: The selected epoch data matrix
% - noisyTrials: The indices of the noisy trials
% - filePath: The file path to save the struct

% Create a struct and assign the matrices as fields
myStruct = struct();

% Adding the clean epoch data to the struct
myStruct.cleanEpochData = selectedEpochData;

% Adding the odor from normal data to the struct
odor = load("Dataset\Normal.mat").normal(2).odor;
myStruct.odor = odor; % Num_trials x 1

% Adding the noisy trials to the struct
myStruct.noisy = noisyTrials.'; % Num_noisy x 1

% Save the struct to the specified file path
save(filePath , 'myStruct');
end

```

The struct files can be found in the "Data" folder of the uploaded zipped file.

3.4 Phase Locking Value (PLV)

Phase Locking Value (PLV) is a metric used to quantify the degree of phase synchronization or phase consistency between two oscillatory signals. It assesses the relationship between the phases of two signals at a specific frequency range. PLV is commonly used in the analysis of neural signals, including electroencephalography (EEG) and magnetoencephalography (MEG), to investigate the synchronization of oscillatory activity between different brain regions or across different frequency bands within a single region.. It provides insights into the functional connectivity and coordination of neural activity.

PLV ranges from 0 to 1, where a value of 1 indicates perfect phase synchronization, while a value close to 0 represents a lack of synchronization. High PLV values suggest that the phases of the two signals are consistently aligned or coupled, indicating strong synchronization. This synchronization can reflect functional interactions between brain regions or coordinated activity within a network. In contrast, low PLV values indicate weaker or desynchronized activity, suggesting less functional coupling between the signals.

- What does phase synchronization indicate from a functional point of view? Discuss its importance with valid references.
- Formulate the definition of PLV and briefly discuss the mathematical tools needed to calculate it.
- Implement a function which finds the PLV between two channels in a specific frequency range. This function is going to be needed in the [section 4](#). (NOTE: You are allowed to define this function with any required input arguments.)

Answer: phase synchronization

Phase synchronization refers to the phenomenon in which two or more oscillatory systems exhibit a consistent relationship between their phase angles. It is a fundamental concept in the study of complex systems, especially in the fields of neuroscience, physics, and engineering. From a functional point of view, phase synchronization carries important implications for information processing, communication, and coordination within and between different parts of a system.

The importance of phase synchronization lies in its ability to facilitate efficient and reliable communication and coordination between different elements of a complex system. In the context of neuroscience, phase synchronization has been extensively studied in relation to brain activity and cognitive processes. Research has shown that synchronized neural oscillations, particularly in the gamma frequency range (30-100 Hz), are associated with various cognitive functions such as attention, memory, perception, and consciousness.

For example, studies have demonstrated that phase synchronization between different brain regions is crucial for coordinating information processing during tasks that require the integration of multiple sensory inputs or the coordination of different cognitive processes. In these cases, synchronized oscillations enable the precise timing of neural activity, allowing for efficient communication and coordination between brain regions.

Furthermore, phase synchronization has been linked to the binding problem in neuroscience, which refers to the challenge of integrating different features of sensory stimuli into a unified percept. It has been proposed that phase synchronization between oscillatory populations in different brain regions may serve as a mechanism for binding together the neural representations of different features, leading to coherent perception.

Beyond neuroscience, phase synchronization also plays a significant role in other fields. In physics, for example, it is relevant in the study of coupled oscillators, where synchronization phenomena can arise in various systems, including lasers, mechanical oscillators, and electrical circuits. Phase synchronization in such systems allows for precise control and coordination of their behavior, enabling applications in areas such as communication, signal processing, and synchronization-based computing.

Overall, phase synchronization is a crucial phenomenon that enables efficient communication and coordination in complex systems. Its importance spans various domains, from neuroscience to physics, and understanding its underlying mechanisms can provide valuable insights into the functioning of diverse systems.

References:

1. Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9(10), 474-480.
2. Varela, F., Lachaux, J., Rodriguez, E., Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2(4), 229-239.
3. Buzsáki, G., Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679), 1926-1929.
4. Pikovsky, A., Rosenblum, M., Kurths, J. (2001). *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge University Press.

Answer: Formulate PLV

Phase Lock Value (PLV) is a measure used to quantify the phase synchronization or coupling between two signals or channels in the frequency domain. It assesses the consistency and strength of the phase relationship between the signals at different frequencies.

Mathematically, PLV is defined as the average of the absolute value of the complex-valued phase difference between the two signals, computed across a range of frequencies. It is calculated as follows:

$$PLV = \left| \frac{1}{N} \sum_{n=1}^N e^{i(\phi_1 - \phi_2)} \right|$$

- N is the number of frequency bins or samples over which the PLV is calculated.
- ϕ_1 is the phase of signal 1 at a specific frequency.
- ϕ_2 is the phase of signal 2 at a specific frequency.

By averaging the absolute values of the complex-valued phase differences across frequencies and normalizing by the number of frequencies, PLV provides a scalar value that represents the degree of phase synchronization between the signals. A value of 1 indicates perfect phase lock or synchronization, while a value close to 0 suggests a lack of phase coupling.

Answer: Calculate PLV Function

```
function plv = calculatePLVInRange(signal1, signal2, fs,
freqRange)
    % Perform Fourier transform on the signals
    fft1 = fft(signal1);
    fft2 = fft(signal2);

    % Frequency axis
    f = (0:length(signal1)-1)*(fs/length(signal1));

    % Indices of frequencies within the specified range
    freqIndices = find(f >= freqRange(1) & f <= freqRange(2));
    ;

    % Extract the phase components of the signals within the
    % frequency range
    phase1 = angle(fft1(freqIndices));
    phase2 = angle(fft2(freqIndices));

    % Calculate the Phase Locking Value (PLV)
    plv = abs(mean(exp(1i * (phase1 - phase2))));

end
```

4 Results

In this section, you need to present the required results to assess the difference of Phase Locking Values (PLV) among two groups, namely AD and Normal in the slow gamma frequency range, which is 35 to 40 Hz.

To fairly compare your results in this part, you do not need to use your preprocessing data from section 3.3 and the preprocessed data of 15 healthy (normal) (age = 69.27 ± 6.65 , female = 53.33%) individuals and 13 AD patients (age = 75.31 ± 9.90 , female = 61.54%) are available through `Dataset/Normal.mat` and `Dataset/AD.mat`.

4.1 Values

Find the PLV for all participants of both groups on both frequent and rare odors between the `Fz` and `Cz` channels using the function you implemented in section 3.4 .

4.2 Distributions

Draw the box plots of PLVs you found in the previous part among two groups and two odors. Also, fit a gaussian distribution on these PLVs and present your results. You need to specify the corresponding p-values to evaluate the statistical significance of your findings.

4.3 Statistical Significance

Based on the p-values you founded in the previous part, discuss whether we could state that the "PLV is significantly different among AD and Normal subjects in the slow gamma frequency range".

4.4 Phase Difference

Draw a polar histogram of the phase difference between `Fz` and `Cz` channels during frequent odor trials for a random subject in each group and compare the results. Also, plot the mean value of this quantity among all the subjects of each group and discuss the results.

4.5 Heatmaps

Now you need to plot a heatmap which has the PLVs between each pair of the channels. Find whether PLV between other channel pairs are significantly different among two groups in the slow gamma frequency range and test your results. (NOTE: You need to provide p-values for your hypothesis if you found any significantly different channel pairs apart from (`Fz,Cz`).)

Answer: Values

Firstly, we observe that Fz and Cz are the second and third channels in our EEG data. Our objective is to develop a MATLAB function that takes EEG data, sampling frequency, frequency range, the number of individuals, and channel indices as inputs. Within this function, we will examine the epoch data and trials of each individual. In our dataset, there is an array called 'odor' which allows us to distinguish between rare and frequent odors, namely lemon and rose odors respectively. Upon separation, we need to utilize our `calculatePLV` function to perform this analysis for each trial. Finally, we will average the resulting PLVs and assign them to each individual.

```

function [plvFreq, plvRare] = plv2Channels(data, fs,
    freqRange, numIndividuals, channel1, channel2)
    % Initialize output variables
    plvRare = zeros(1, numIndividuals);
    plvFreq = zeros(1, numIndividuals);

    % Iterate over each individual in the dataset
    for i = 1:numIndividuals
        % Extract the epochs for the rare and frequent conditions
        % for the current individual
        rareEpoch = data(i).epoch(:, :, data(i).odor == 1);
        frequentEpoch = data(i).epoch(:, :, data(i).odor == 0);

        % Extract the data for the selected channels from the
        % rare and frequent epochs
        rareFz = squeeze(rareEpoch(channel1,:,:));
        rareCz = squeeze(rareEpoch(channel2,:,:));

        freqFz = squeeze(frequentEpoch(channel1,:,:));
        freqCz = squeeze(frequentEpoch(channel2,:,:));

        % Calculate the number of rare and frequent trials
        numRareTrials = size(rareFz,2);
        numFreqTrials = size(freqFz,2);

        % Iterate over the rare trials
        for j = 1:numRareTrials
            % Calculate the PLV between the two channels for the
            % current rare trial
            plvRare(i) = plvRare(i) + calculatePLVInRange(rareFz
                (:,j), rareCz(:,j), fs, freqRange);
        end
        % Average the PLV values for the rare trials
        plvRare(i) = plvRare(i)/numRareTrials;
    end
end

```

Answer: Values

```
% Iterate over the frequent trials
for j = 1:numFreqTrials
    % Calculate the PLV between the two channels for the
    % current frequent trial
    plvFreq(i) = plvFreq(i) + calculatePLVInRange(freqFz
        (:,j), freqCz(:,j), fs, freqRange);
end
% Average the PLV values for the frequent trials
plvFreq(i) = plvFreq(i)/numFreqTrials;
end
end
```

The desired output should consist of four vectors, as depicted in the following table:

	No-F	No-R	AD-F	AD-R
1	0.57	0.50	0.70	0.73
2	0.99	1	0.30	0.27
3	0.94	0.97	0.55	0.49
4	1	1	0.83	0.86
5	0.99	0.99	0.79	0.77
6	0.92	0.95	0.86	0.81
7	0.91	0.93	0.97	0.96
8	0.75	0.79	0.71	0.73
9	0.93	0.92	0.75	0.82
10	1	0.99	0.60	0.59
11	0.90	0.89	0.63	0.59
12	0.70	0.71	0.28	0.29
13	0.79	0.74	0.32	0.31
14	0.28	0.29		
15	0.36	0.36		

Table 2: PLV Values

Answer: Distribution

First, We draw box plots of PLVs with `boxplot` built-in function. Then, we use `histfit` to plot a bar plot and fit a gaussian distribution on it.

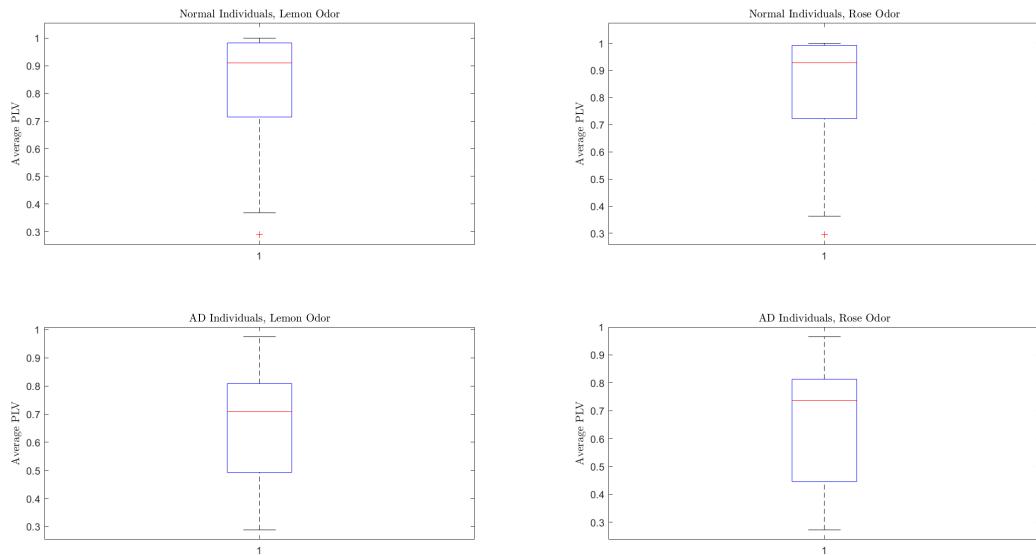


Figure 21: Box Plots of PLVs

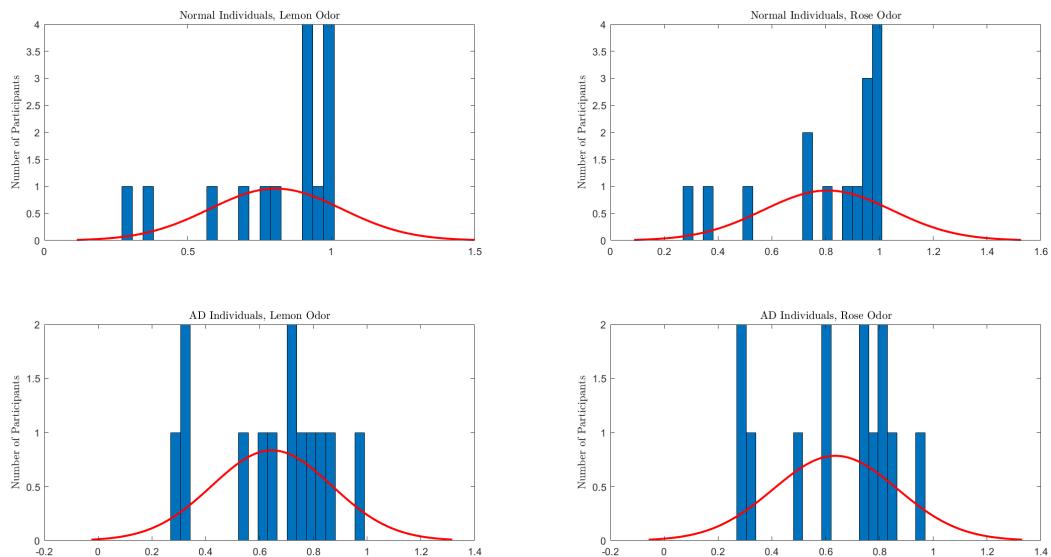


Figure 22: Fit Gaussian Distribution

Answer: P-Values

In statistics, the p-value is a measure of evidence against a null hypothesis. It represents the probability of obtaining a test statistic as extreme as, or more extreme than, the observed value, assuming that the null hypothesis is true. In simpler terms, the p-value helps determine if the results of a statistical test are statistically significant or due to chance.

The p-value for a two-sample t-test can be calculated using the following steps:

1. State the null hypothesis (H_0) and the alternative hypothesis (H_a).

- H_0 : The means of the two populations are equal.
- H_a : The means of the two populations are not equal.

2. Collect your data and calculate the following statistics:

- Sample means (\bar{x}_1 and \bar{x}_2) for the two groups.
- Sample standard deviations (s_1 and s_2) for the two groups.
- Sample sizes (n_1 and n_2) for the two groups.

3. Calculate the test statistic (t-value) using the formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

4. Determine the degrees of freedom (df) for the t-distribution using the formula:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2/(n_1 - 1) + \left(\frac{s_2^2}{n_2}\right)^2/(n_2 - 1)}$$

5. Use the calculated t-value and degrees of freedom to find the p-value. This can be done by referring to a t-distribution table or using statistical software.

6. Interpret the p-value based on your chosen significance level (α). If the p-value is less than the chosen α (commonly 0.05), you would reject the null hypothesis and conclude that there is evidence of a significant difference between the means of the two groups. If the p-value is greater than α , you would fail to reject the null hypothesis.

We use `ttest2()` function to find p-value between 2 data. P-Value of normal frequent PLVS and AD frequent PLVS is one P-Value and another is P-Value of normal rare PLVS and AD rare PLVS.

P-Value (Normal Frequent PLVS, AD Frequent PLVS) = 0.0738

P-Value (Normal Rare PLVS, AD Rare PLVS) = 0.0670

Answer: P-Values

We are now ready to calculate the p-values. We need to determine two separate p-values, one for the rare smell of rose and the other for the frequent smell of lemon. These p-values will be obtained by comparing two groups: "AD" and "Normal". We can use Table 2 to find \bar{X} and X .

Frequent:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad \rightarrow \quad \begin{cases} \bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_{i,N} = \frac{1}{15} \sum_{i=1}^{15} X_{i,N} = 0.8064 \\ \bar{X}_{AD} = \frac{1}{N} \sum_{i=1}^N X_{i,AD} = \frac{1}{13} \sum_{i=1}^{13} X_{i,AD} = 0.6458 \end{cases}$$

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2 \quad \rightarrow \quad \begin{cases} s_N^2 = \frac{1}{14} \sum_{i=1}^{15} (X_{i,N} - \bar{X})^2 \rightarrow s_N = 0.2309 \\ s_{AD}^2 = \frac{1}{12} \sum_{i=1}^{13} (X_{i,AD} - \bar{X})^2 \rightarrow s_{AD} = 0.2235 \end{cases}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{0.8064 - 0.6458}{\sqrt{\frac{0.23^2}{15} + \frac{0.22^2}{13}}} \rightarrow t = 1.8545$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2/(n_1-1) + \left(\frac{s_2^2}{n_2}\right)^2/(n_2-1)} = \frac{\left(\frac{0.23^2}{15} + \frac{0.22^2}{13}\right)^2}{\left(\frac{0.23^2}{15}\right)^2/(15-1) + \left(\frac{0.22^2}{13}\right)^2/(13-1)} \simeq 25.72 \simeq 26$$

Rare:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad \rightarrow \quad \begin{cases} \bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_{i,N} = \frac{1}{15} \sum_{i=1}^{15} X_{i,N} = 0.8071 \\ \bar{X}_{AD} = \frac{1}{N} \sum_{i=1}^N X_{i,AD} = \frac{1}{13} \sum_{i=1}^{13} X_{i,AD} = 0.6362 \end{cases}$$

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2 \quad \rightarrow \quad \begin{cases} s_N^2 = \frac{1}{14} \sum_{i=1}^{15} (X_{i,N} - \bar{X})^2 \rightarrow s_N = 0.2396 \\ s_{AD}^2 = \frac{1}{12} \sum_{i=1}^{13} (X_{i,AD} - \bar{X})^2 \rightarrow s_{AD} = 0.2313 \end{cases}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{0.8071 - 0.6362}{\sqrt{\frac{0.24^2}{15} + \frac{0.23^2}{13}}} \rightarrow t = 1.9216$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right)^2/(n_1-1) + \left(\frac{s_2^2}{n_2}\right)^2/(n_2-1)} = \frac{\left(\frac{0.24^2}{15} + \frac{0.23^2}{13}\right)^2}{\left(\frac{0.24^2}{15}\right)^2/(15-1) + \left(\frac{0.23^2}{13}\right)^2/(13-1)} \simeq 25.74 \simeq 26$$

Answer: Statistical Significance

Now we consider $\alpha = 0.05$. Our null hypothesis H_0 assumes equality of the means ($\mu_N = \mu_{AD}$), while the alternative hypothesis H_a assumes inequality between them ($\mu_N \neq \mu_{AD}$). If the P-value is greater than α , we reject H_0 ; otherwise, if the P-value is less than or equal to α , we accept H_0 . I used [this site](#) to find P-Value.

Frequent:

P-Value = $P(|t| > 1.85) \simeq 0.075711$, which is greater than α . Therefore, we reject H_0 . This implies that for the frequent smell (lemon), the mean of normal individuals and AD individuals are not equal.

Rare:

P-Value = $P(|t| > 1.92) \simeq 0.065892$, which is greater than α . Hence, we reject H_0 . This indicates that for the rare smell (rose), the mean of normal individuals and AD individuals are not equal.

We can set alpha equal to 0.1 in order to ensure that its value is less than alpha. Therefore, with 90% confidence, we can conclude that these two means are equal.

We should note that a lower P-value corresponds to a greater level of confidence. I expected that due to the larger sample size in the lemon experiment, our p-value would be smaller. However, unfortunately, this did not happen here.

Note that, our calculated P-value and MATLAB's P-value were almost equal.

Answer: Phase Difference

Through the utilization of a polar histogram, we can visually examine how phase differences are distributed and concentrated across the angular axis for both groups (normal and AD patients) and odors (lemon and rose). A narrower distribution or a peak closer to zero suggests a stronger and more consistent phase locking between the channels, as the phase differences are concentrated around zero. Conversely, a wider distribution or a peak further from zero indicates a more varied and weaker phase locking pattern.

To find phase difference between 2 channels Fz and Cz during frequent odor trials for a random subject, We use Following MATLAB function:

```
function [averagePhaseFreq, averagePhaseRare] =
    avgPhase2Channels(data, channel1, channel2, subjectIndex)
    % Initialize average phase variables
    averagePhaseRare = 0;
    averagePhaseFreq = 0;

    % Extract rare and frequent epochs for the specified
    % subject index
    rareEpoch = data(subjectIndex).epoch(:, :, data(
        subjectIndex).odor == 1);
    frequentEpoch = data(subjectIndex).epoch(:, :, data(
        subjectIndex).odor == 0);

    % Calculate average phase difference for rare trials
    numRareTrials = size(rareEpoch, 3);
    for i = 1:numRareTrials
        % Retrieve epoch data for the specified channels
        trialRare = rareEpoch(:, :, i);

        % Calculate phase difference between the two channels
        phaseData1 = angle(trialRare(channel1, :));
        phaseData2 = angle(trialRare(channel2, :));

        phase = rad2deg(phaseData1 - phaseData2);

        % Accumulate phase differences
        averagePhaseRare = averagePhaseRare + phase;
    end

    % Calculate average phase difference for rare trials
    averagePhaseRare = averagePhaseRare / numRareTrials;
```

Answer: Phase Difference

```
% Calculate average phase difference for frequent trials
numFreqTrials = size(frequentEpoch ,3);
for i = 1:numFreqTrials
    % Retrieve epoch data for the specified channels
    trialFreq = frequentEpoch(:,:,i);

    % Calculate phase difference between the two channels
    phaseData1 = angle(trialFreq(channel1, :));
    phaseData2 = angle(trialFreq(channel2, :));

    phase = rad2deg(phaseData1 - phaseData2);

    % Accumulate phase differences
    averagePhaseFreq = averagePhaseFreq + phase;
end

% Calculate average phase difference for frequent trials
averagePhaseFreq = averagePhaseFreq / numFreqTrials;
end
```

After that, we have to plot polar histogram.

that is to note we could avoid separating data to freq and rare in this step. But I prefer to it.

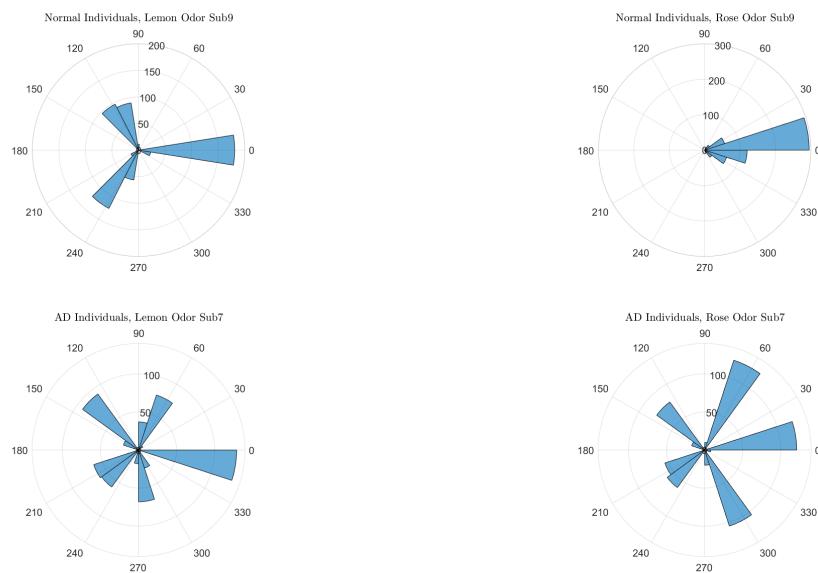


Figure 23: Polar Histogram of a Random Subject Phase Difference of 2 channels

Answer: Phase Difference

Now, We have to calculate mean value of phase difference. in order to it, We average data for all individuals like following MATLAB function.

```
function [meanPhaseFreq, meanPhaseRare] = meanPhaseDiff(data,
    channel1, channel2, numIndividuals)
    meanPhaseFreq = 0; % Initialize mean phase difference
    for channel1 and channel2
        meanPhaseRare = 0; % Initialize mean phase difference
        for channel2 and channel1

            % Iterate over the specified number of individuals
            for i = 1:numIndividuals
                [temp1, temp2] = avgPhase2Channels(data, channel1,
                    channel2, i); % Calculate average phase
                    difference
                meanPhaseFreq = meanPhaseFreq + temp1; % Accumulate
                    phase differences for channel1 and channel2
                meanPhaseRare = meanPhaseRare + temp2; % Accumulate
                    phase differences for channel2 and channel1
            end

            meanPhaseFreq = meanPhaseFreq / numIndividuals; %
                Compute mean phase difference for channel1 and
                channel2
            meanPhaseRare = meanPhaseRare / numIndividuals; %
                Compute mean phase difference for channel2 and
                channel1
        end
    end
```

Phase difference is a concept commonly used in signal processing and waveform analysis to compare the relative positions of two waveforms or signals. It quantifies the angular difference between two waveforms at a particular point in time.

The polar histogram reveals a noteworthy trend where the results are closer to zero for normal participants as compared to AD patients. This observation suggests that normal participants may display a higher level of consistency or stronger phase locking in their response to the odors. This visual pattern aligns with the concept that AD patients may exhibit differences in neural synchronization or connectivity when compared to individuals without the condition. However observation is not enough and we need statistical test to become sure.

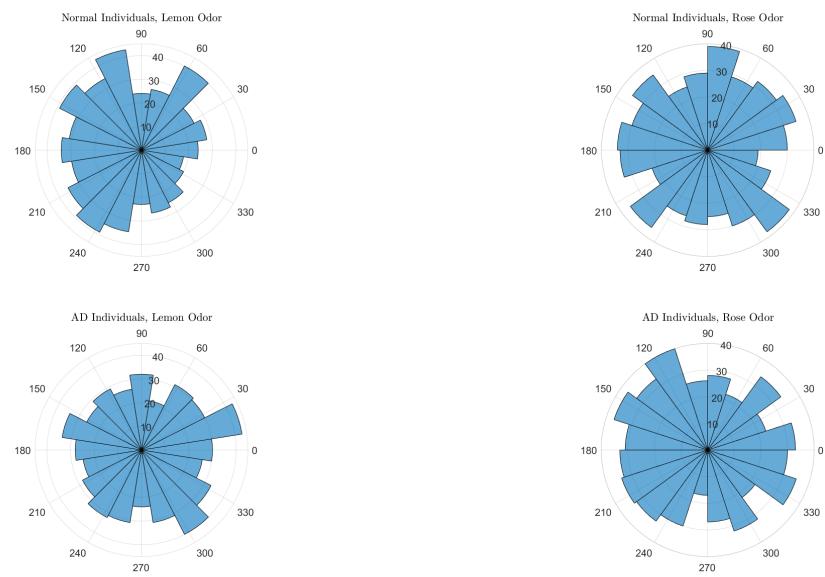


Figure 24: Polar Histogram of Mean Phase Difference

Answer: Heatmaps

We can enhance the code by leveraging MATLAB's vectorization capabilities, which eliminate the need for a nested loop. First, we'll apply a PLV calculation function to calculate the Phase Locking Value (PLV) between each pair of channels. The resulting PLVs will be stored for each individual, and then we can compute the mean PLV using a built-in MATLAB function. Please refer to the MATLAB code below that demonstrates this approach:

```

for i = 1:numChannels
    for j = 1:numChannels
        [temp1, temp2] = plv2Channels(normal, fs, freqRange,
            normalIndividuals, i, j);
        plvNormalFreqMatrix(i,j) = mean(temp1);
        plvNormalRareMatrix(i,j) = mean(temp2);
        [temp3, temp4] = plv2Channels(AD, fs, freqRange,
            ADIndividuals, i, j);
        plvADFreqMAtrix(i,j) = mean(temp3);
        plvADRareMatrix(i,j) = mean(temp4);
        [~, pValFreq(i,j)] = ttest2(temp1, temp3);
        [~, pValRare(i,j)] = ttest2(temp2, temp4);
    end
end

```

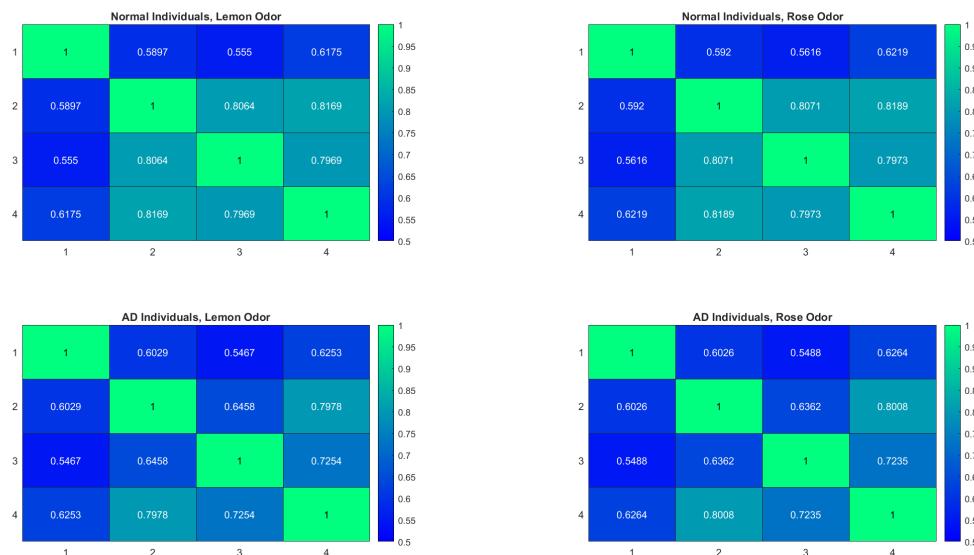


Figure 25: PLV Heatmap of all channel's pair

Answer: P-Values

In the previous nested loops, we calculated and stored P-values. We have two separate tables for the P-values, one for frequent smells and the other for rare smells.

P-Val	CH1	CH2	CH3	CH4
CH1	0	0.8941	0.9314	0.9423
CH2	0.8941	0	0.0738	0.7834
CH3	0.9314	0.0738	0	0.3647
CH4	0.9423	0.7834	0.3647	0

Table 3: Frequent Smell P-Value for all Channels

P-Val	CH1	CH2	CH3	CH4
CH1	0	0.9201	0.8974	0.9678
CH2	0.8941	0	0.0670	0.8012
CH3	0.8974	0.0670	0	0.3733
CH4	0.9678	0.8012	0.3733	0

Table 4: Rare Smell P-Value for all Channels

In this test, a result that is closer to 1 and greater indicates that the test is more statistically significant or extreme compared to other tests. A value closer to 1 suggests a stronger deviation from the null hypothesis or a greater effect size. It signifies that the observed data is less likely to occur by chance alone and provides stronger evidence in favor of the alternative hypothesis.

By observing [Table3](#) and [Table4](#) indicates a statistically significant difference in PLVs between channels 1 and 4 in both rare and frequent experiments.

5 *Bonus

5.1 Mild Cognitive Impairment (MCI)

Mild Cognitive Impairment (MCI) is the stage between the expected decline in memory and thinking that happens with age and the more serious decline of dementia. MCI may include problems with memory, language or judgment. People with MCI may be aware that their memory or mental function has slipped. [1]

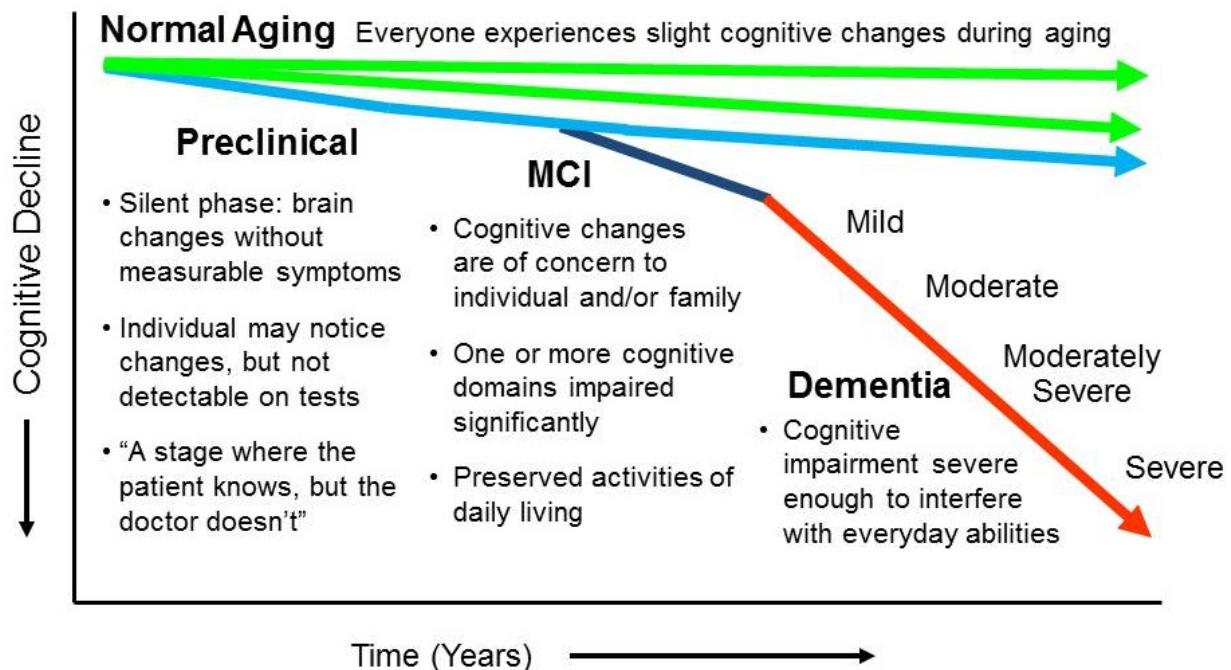


Figure 26: Normal Aging to Demantia Process

5.1.1 Additional Information

Describe the relationship between MCI and AD. Explain whether MCI would always result in AD and briefly investigate the causes of MCI.

Answer

Relationship between MCI and AD: The relationship between Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD) is complex. MCI is a condition characterized by mild cognitive decline that is noticeable but does not significantly impair daily functioning. AD, on the other hand, is a progressive neurodegenerative disorder characterized by severe cognitive decline and functional impairment.

Does MCI always result in AD? It is important to note that not all individuals with MCI will inevitably develop AD. MCI is considered a prodromal stage or an intermediate state between normal cognitive aging and AD. While some individuals with MCI may progress to develop AD over time, others may remain stable or even revert to normal cognitive functioning. Studies suggest that the conversion rate from MCI to AD varies widely, and factors such as age, genetics, underlying brain pathology, and comorbidities can influence this progression.

Causes of MCI: The causes of MCI are multifactorial and can vary among individuals. Some potential causes and risk factors for MCI include:

1. Age-related brain changes: Normal aging involves some degree of cognitive decline, and MCI can be a manifestation of age-related brain changes.
2. Neurodegenerative diseases: MCI can result from underlying neurodegenerative conditions, including AD, Parkinson's disease, Lewy body dementia, or vascular dementia.
3. Vascular factors: Conditions that affect blood flow to the brain, such as hypertension, diabetes, or strokes, can contribute to the development of MCI.
4. Genetic and environmental factors: Certain genetic variations, such as the APOE $\epsilon 4$ allele, have been associated with an increased risk of developing both MCI and AD. Additionally, lifestyle factors, including education, physical exercise, social engagement, and cognitive stimulation, may impact the risk of MCI.
5. Other medical conditions: Medical conditions such as depression, anxiety, sleep disorders, and chronic diseases can contribute to cognitive impairment and the development of MCI.

Overall, the relationship between MCI and AD is complex, and while MCI can be an early indicator or precursor to AD, not all individuals with MCI will progress to develop AD. The causes of MCI are multifaceted and involve a combination of age-related changes, underlying neurodegenerative conditions, vascular factors, genetic and environmental factors, and other medical conditions.

5.1.2 MCI Data Processing

In the provided dataset, you can find `MCI.mat` file. This dataset contains preprocessed cleaned EEG recording of the same task described in sections 3.1 and 3.2 for 7 MCI patients.

Based on the significantly different coupled channels you found for differentiation between AD and Normal groups, find the Phase-Locking-Value (PLV) for the MCI subjects and provide the required results by comparing all the the 3 states (Normal, MCI, AD). Your findings must include the significance testing by providing the corresponding p-values.

Answer: MCI Data Processing

Based on heatmaps figure25 we choose 3 pairs that has most significant difference. those pairs are channels(1,2), channels(1,3), channels(1,4).

We use previous MATLAB functions to calculate PLV of channels of MCI and plot boxplot and histfit in gamma range.

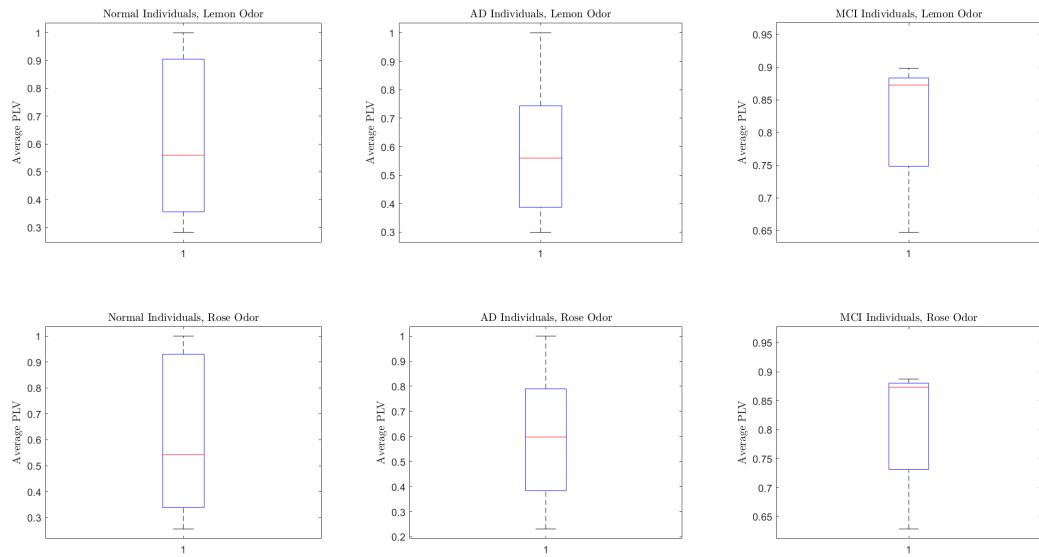
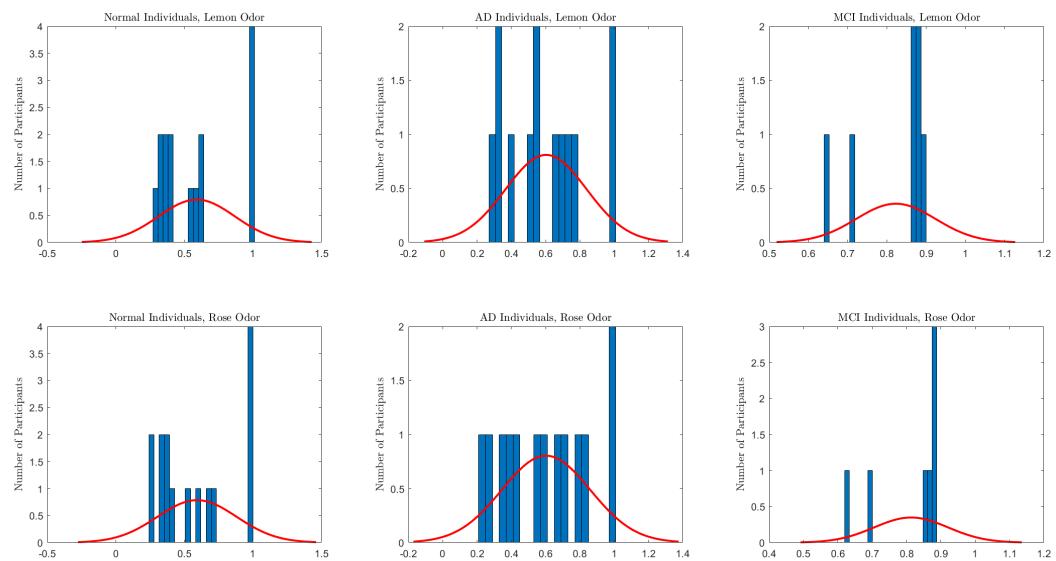
We define following MATLAB function to do it for us:

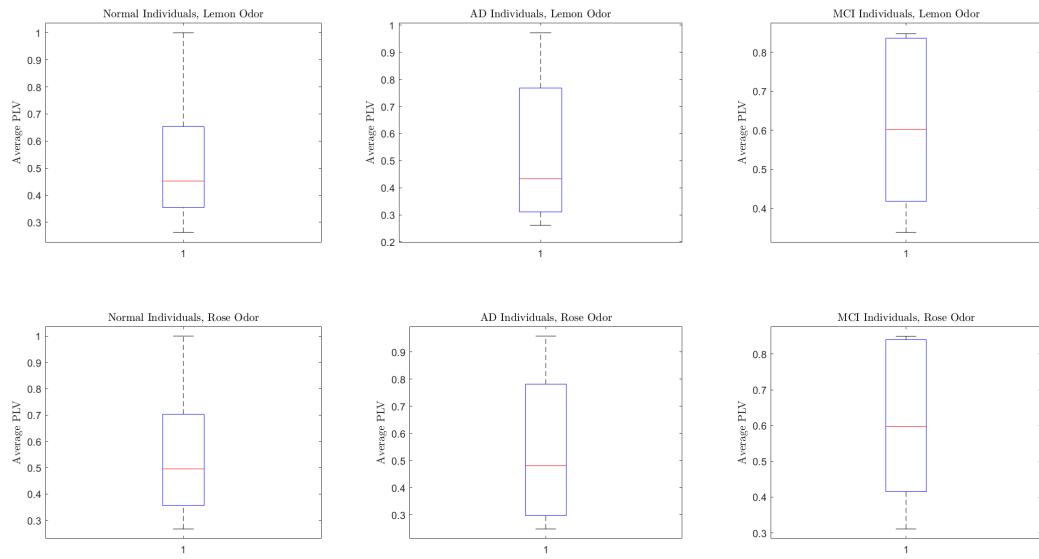
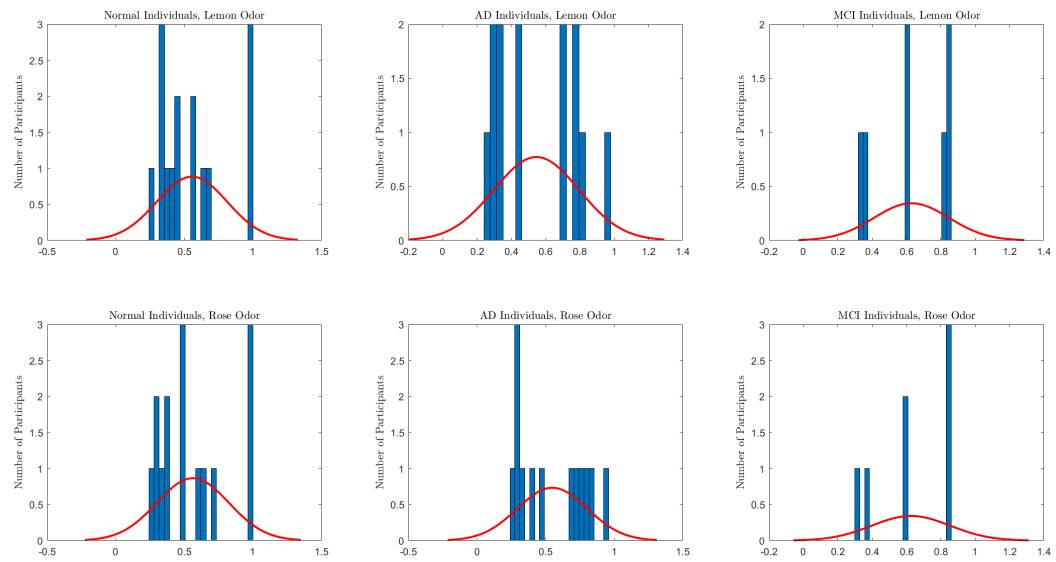
```
function plotting(plvNormalFreq, plvNormalRare, plvADFreq,
plvADRARE, plvMCIFreq, plvMCIRare)
% Box Plot
figure()

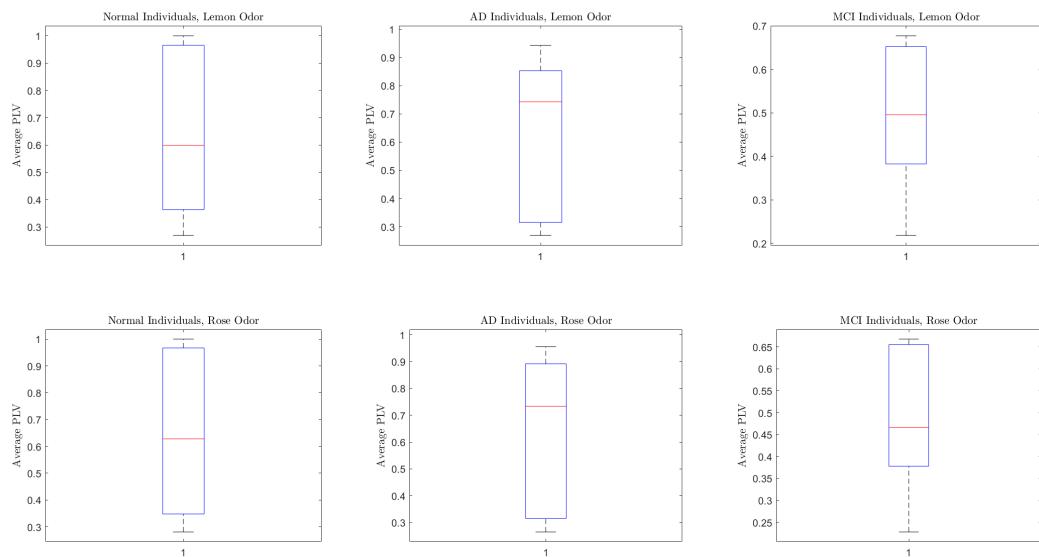
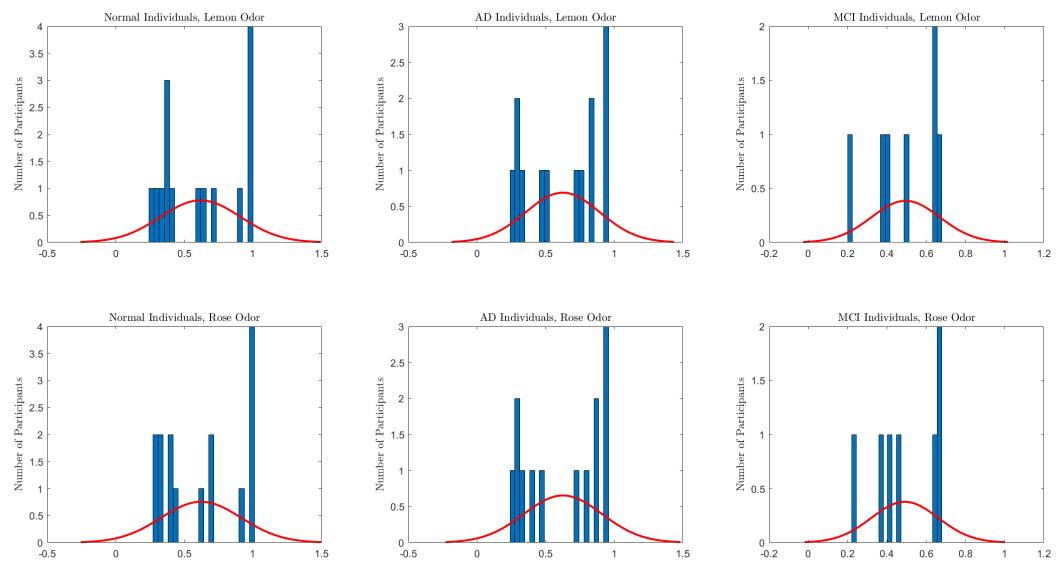
subplot(2,3,1)
boxplot(plvNormalFreq);
ylabel('Average PLV', 'Interpreter','latex')
title('Normal Individuals, Lemon Odor', 'Interpreter',
' latex')
...up to 6

% Histfit
figure()

subplot(2,3,1)
histfit(plvNormalFreq, 20, "normal");
ylabel('Number of Participants', 'Interpreter','latex')
title('Normal Individuals, Lemon Odor', 'Interpreter',
' latex')
...up to 6
end
```

**Figure 27:** Box Plot of Coupled Channels 1,2 PLV**Figure 28:** Histogram of Coupled Channels 1,2 PLV

**Figure 29:** Box Plot of Coupled Channels 1,3 PLV**Figure 30:** Histogram of Coupled Channels 1,2 PLV

**Figure 31:** Box Plot of Coupled Channels 1,4 PLV**Figure 32:** Histogram of Coupled Channels 1,2 PLV

Answer: P-Value

We have three sets of data, and we cannot use the `ttest2()` function because it is specifically designed for two-sample t-tests. In order to compare multiple groups and obtain the p-value, we can utilize the ANOVA (Analysis of Variance) method. ANOVA allows us to assess the statistical significance of differences between multiple groups.

ANOVA (Analysis of Variance) is a statistical method used to compare the means of three or more groups. It helps determine if there are significant differences between the group means and assesses whether these differences are due to random variation or actual group effects.

The basic idea behind ANOVA is to partition the total variation observed in the data into two components: variation between groups and variation within groups. By comparing the amount of variation between groups to the amount of variation within groups, ANOVA calculates an F-statistic that measures the ratio of these two components.

If the calculated F-statistic is large enough and surpasses a critical value, it indicates that the variation between groups is significantly greater than the variation within groups. This implies that there are likely true differences in the means of the groups, and the null hypothesis of equal group means can be rejected.

However, ANOVA assumes that the data within each group follows a normal distribution and that the variances within each group are equal. Violations of these assumptions can affect the accuracy of the results.

We use following Matlab function to find anova tavble and P-Value betwenn 3 groups of data that have non-equal size:

```
function [p, tb] = anova_data(data1, data2, data3)
    % Pad the shorter data sets with NaN values
    max_length = max([length(data1), length(data2), length(
        data3)]);
    data1(end+1:max_length) = NaN;
    data2(end+1:max_length) = NaN;
    data3(end+1:max_length) = NaN;

    % Use anova1 function to compare the three data sets
    [p, tb] = anova1([data1', data2', data3']);
end
```

In our analysis, we are required to report a total of 6 P-values. This is due to the fact that we have examined three coupled channels of data, with each channel corresponding to two different types of smells: rare and frequent.

Answer: channel 1,2

Frequent:

$$H_0 : \mu_{Normal} = \mu_{AD} = \mu_{MCI}$$

P-Value = 0.0949 → It means we are 90% sure that H_0 will be accepted and their mean are equal.

Rare:

$$H_0 : \mu_{Normal} = \mu_{AD} = \mu_{MCI}$$

P-Value = 0.1436 → It means we are 85% sure that H_0 will be accepted and their mean are equal.

Answer: channel 1,3

Frequent:

$$H_0 : \mu_{Normal} = \mu_{AD} = \mu_{MCI}$$

P-Value = 0.7580 → It means we are just 20% sure that H_0 will be accepted and their mean are equal.

Rare:

$$H_0 : \mu_{Normal} = \mu_{AD} = \mu_{MCI}$$

P-Value = 0.8025 → It means we are just 15% sure that H_0 will be accepted and their mean are equal.

Answer: channel 1,4

Frequent:

$$H_0 : \mu_{Normal} = \mu_{AD} = \mu_{MCI}$$

P-Value = 0.5378 → It means we are 45% sure that H_0 will be accepted and their mean are equal.

Rare:

$$H_0 : \mu_{Normal} = \mu_{AD} = \mu_{MCI}$$

P-Value = 0.5165 → It means we are 45% sure that H_0 will be accepted and their mean are equal.

Answer: P-Values

Note that, box plots with notch on arguments and tables that has been output of ANOVA function are in Folder of Images in the uploaded file.

Answer: Phase Difference

We use previous MATLAB functions to calculate phase difference. once, between to random subjects and once mean of all subjects. I won't show them for 1 subject and the results are in **Images Folder** in uploaded file.

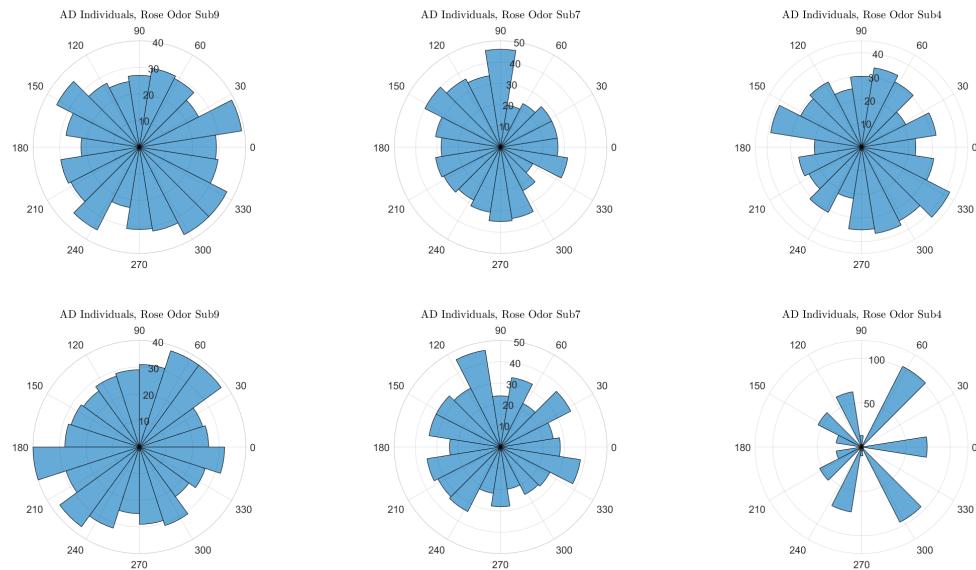


Figure 33: Polar Histogram of mean values of Coupled Channels 1,2 PLV

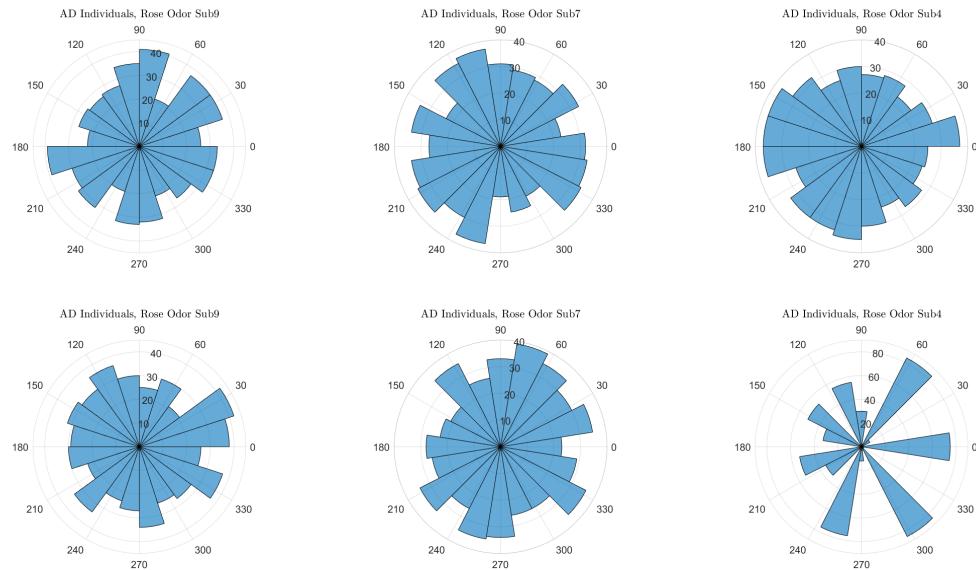


Figure 34: Polar Histogram of mean values of Coupled Channels 1,3 PLV

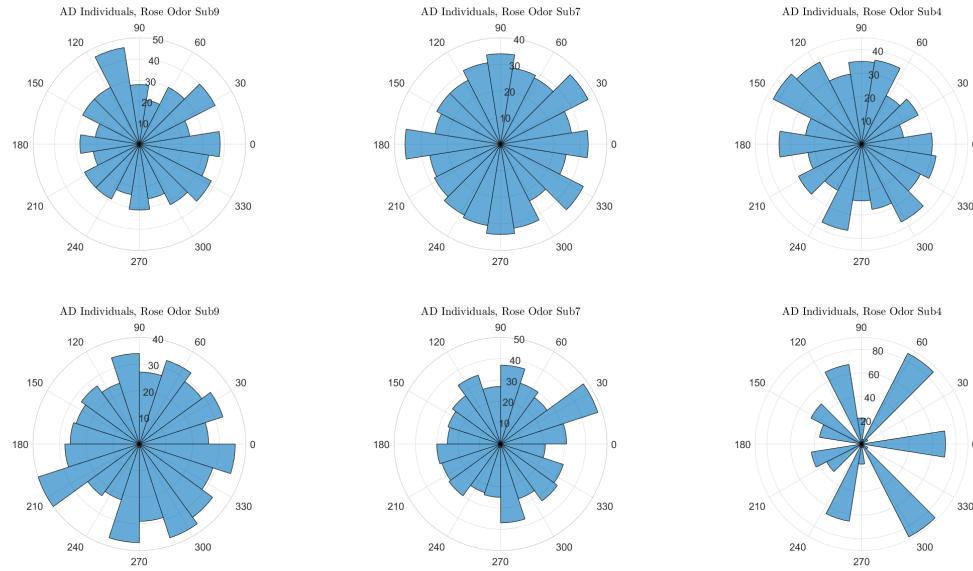


Figure 35: Polar Histogram of mean values of Coupled Channels 1,4 PLV

Answer: Heatmaps

We use previous MATLAB functions to find all channels values and plot heatmap. This heatmap shows us that coupled channels (1,4) and (2,4) and (1,3) have most significant difference. this pair of channels are different with those that were in Normal and AD.

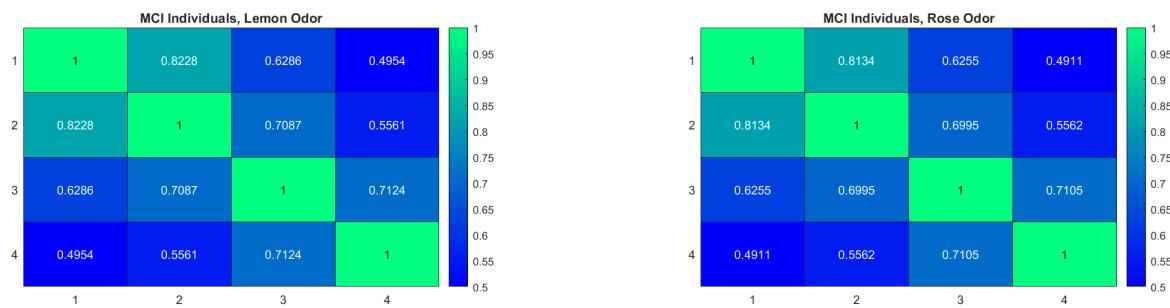


Figure 36: PIV Heatmap of all MCI's channels pair

Answer

After analyzing the data from normal people, AD (Alzheimer's Disease), and MCI (Mild Cognitive Impairment) patients, as well as their responses to two different odors (lemon and rose) using 3 channel pairs (1,2), (1,3), (1,4), we can make the following conclusions based on the p-values and the characteristics observed in the boxplots:

1. Differences between groups: Significant variations exist among the normal, AD, and MCI groups in terms of their responses to the odors. These differences can be observed through the p-values, which indicate the likelihood of obtaining such divergent results by chance alone.
2. Odor-specific effects: The responses to the lemon and rose odors elicit distinct patterns. The p-values associated with each odor indicate whether the differences observed in the responses are statistically significant.
3. Channel pair influence: The choice of channel pairs, such as channel1-channel2 and channel1-channel3, contributes to the observed differences in the responses. The boxplots provide a visual representation of the data distribution for each channel pair and offer insights into the range, median, and variability of the responses.

In summary, based on the statistical significance indicated by the p-values and the characteristics of the boxplots, we can infer the presence of significant differences in the responses to the odors between the normal, AD, and MCI groups. Additionally, the choice of channel pairs plays a role in shaping the observed differences. We can see variability in responses by observing box plots too.

5.2 Phase-Amplitude Coupling (PAC)

PLV was just one instance of the Phase-Amplitude Coupling (PAC) metrics. PAC is a form of cross-frequency coupling where the amplitude of a high frequency signal is modulated by the phase of low frequency oscillations. PAC is the most-studied type of cross-frequency coupling and is thought to be responsible for integration across populations of neurons. Low frequency brain activity controls the information exchange between brain regions by modulating the amplitude of the high frequency oscillations. [6]

5.2.1 Metrics

Conduct a search about other PAC measures and briefly provide an explanation about two of them.

Answer

Modulation Index (MI):

The Modulation Index (MI) is a popular PAC measure that quantifies the strength of phase-amplitude coupling between two oscillatory signals. It focuses on the relationship between the phase of a low-frequency signal (e.g., theta or alpha) and the amplitude of a high-frequency signal (e.g., gamma). The MI is calculated by comparing the amplitude distribution at different phase bins of the low-frequency signal. It provides a measure of the degree to which the amplitude of the high-frequency signal is modulated by the phase of the low-frequency signal.

Canolty's PAC Measure:

Canolty's PAC measure is another commonly used method for quantifying PAC. It involves filtering the raw signal into low-frequency and high-frequency bands. Then, the instantaneous phase of the low-frequency band and the instantaneous amplitude of the high-frequency band are calculated. The phase and amplitude values are then used to compute the PAC measure, which can be based on different statistical approaches such as circular statistics or correlation measures. Canolty's PAC measure is known for its flexibility and adaptability to different frequency bands and signal characteristics.

PAC Measure	Description
Phase Locking Value (PLV)	Measures phase synchronization between signals
Modulation Index (MI)	Quantifies strength of phase-amplitude coupling
Canolty's PAC Measure	Flexibly adapts to different frequency bands

Table 5: Comparison of PAC Measures

I choose MI metric and I will explain a bit more about it and its implementation in next part.

5.2.2 Implementation

Implement one of the metrics mentioned earlier as a biomarker for distinguishing between AD and Normal groups. Present the relevant results through plots and provide a discussion regarding the efficacy of the selected metric.

Answer

Modulation Index (MI):

To calculate MI, all possible phases from -180° to 180° are binned into a freely chosen amount of bins.(at first it was established to be 18 bins of 20° but then it became free to choose). That is to say amount of bins can influence the results. The average amplitude of the amplitude-providing frequency in each phase bin of the phase-providing frequency is computed and normalized by the following formula:

$$P(j) = \frac{\bar{a}}{\sum_{k=1}^N \bar{a}_k}$$

where \bar{a} is the average amplitude of one bin, k is the running index for the bins, and N is the total amount of bins; p is a vector of N values. With the help of these calculations, one obtains the data for the phase-amplitude plot, which depicts the actual phase-amplitude coupling graphically.

We can implement it by following MATLAB function:

```
function modulation_index = calculate_MI_PAC(signal_low_freq,
    signal_high_freq, num_bins, freq_range, fs)

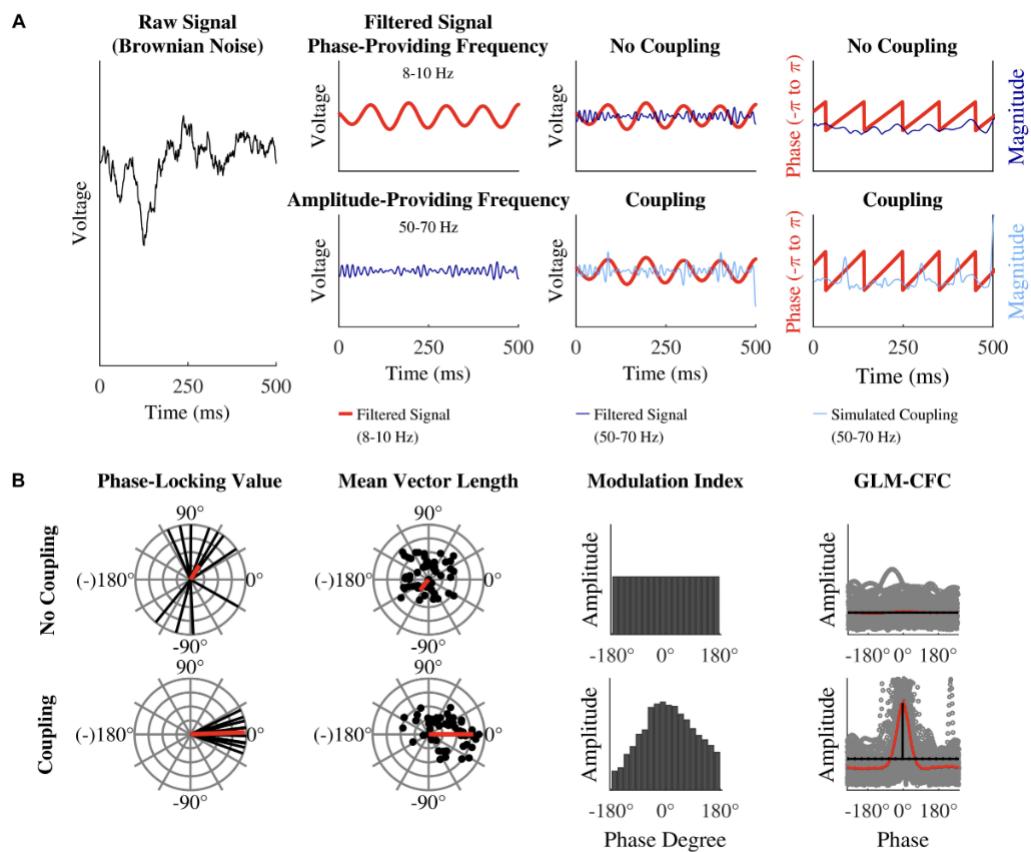
    % Filter the low-frequency signal
    low_freq_filtered = bandpass(signal_low_freq, freq_range,
        fs);

    % Compute the Hilbert transform of the low-frequency
    % signal
    low_freq_hilbert = hilbert(low_freq_filtered);

    % Calculate the phase of the low-frequency signal
    low_freq_phase = angle(low_freq_hilbert);

    % Filter the high-frequency signal
    high_freq_filtered = bandpass(signal_high_freq,
        freq_range, fs);

    % Calculate the amplitude of the high-frequency signal
    high_freq_amplitude = abs(hilbert(high_freq_filtered));
```

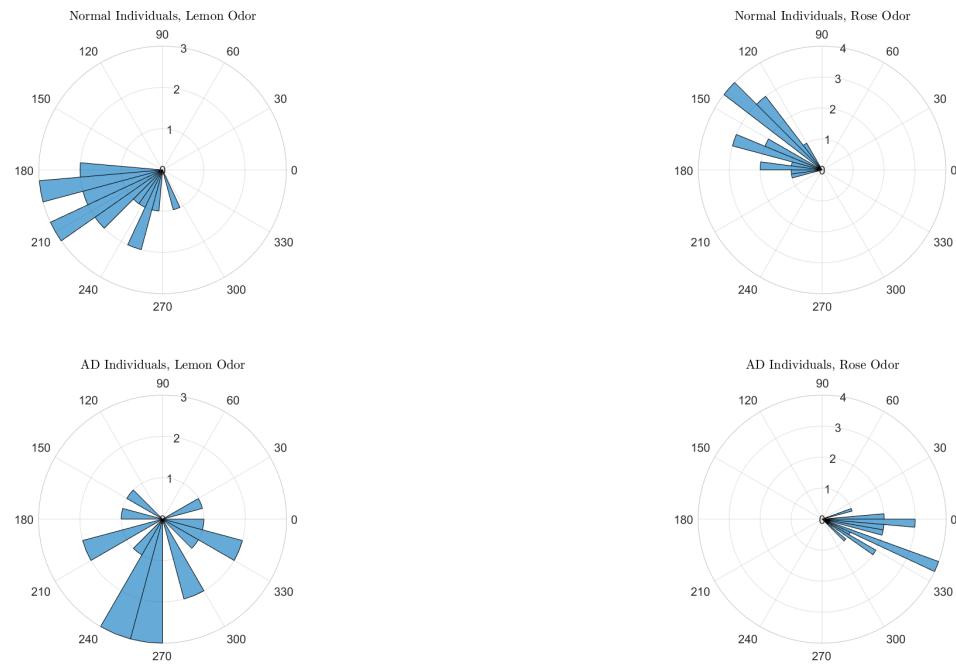
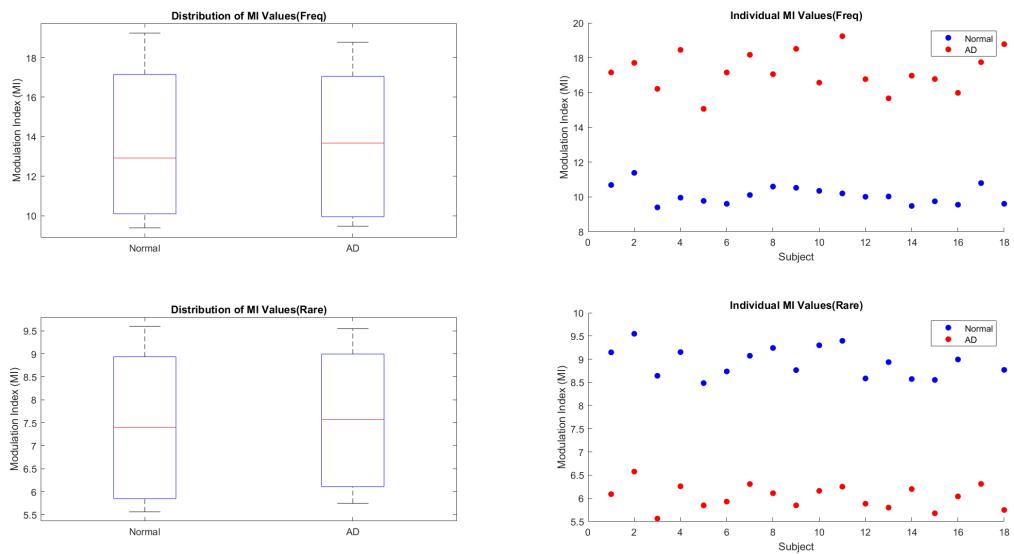
**Figure 37:** phase-amplitude plot**Answer**

```
% Calculate the modulation index
phase_bins = linspace(-pi, pi, num_bins+1);
bin_indices = discretize(low_freq_phase, phase_bins);
modulation_index = zeros(1, num_bins);

for bin_idx = 1:num_bins
    indices = bin_indices == bin_idx;
    modulation_index(bin_idx) = mean(high_freq_amplitude(
        indices));
end

modulation_index = modulation_index';
```

rest of implementation is like implementation of phase difference. I will show polyhistogram, scatter plot and box plot which are for channels (Fz, Pz). I used [this site](#) to learn MI.

**Figure 38:** PolyHistogram of MI**Figure 39:** Box Plot and Scatter Plot of MI

Answer

Polar Histogram: By examining the polar histogram, we can gain insights into the concentration or dispersion of these indices across the angular axis. The polar histogram for the Normal group exhibits a narrower distribution or a peak closer to zero, it suggests that the phase differences or modulation indices are more concentrated around zero. This indicates a stronger or more consistent phase locking between the channels in the Normal group in both rare and frequent experiments.

Box Plot: By examining the characteristics of the boxplots, we can identify differences in the central tendency, spread, and presence of outliers.

- Median Values: Comparing the medians of the boxplots, a higher median for the AD group indicates a higher average value of the modulation indices compared to the Normal group. This suggests that the AD group may exhibit stronger or more pronounced modulation effects.
- Spread and Outliers: Examining the sizes of the boxes and the whiskers, a smaller box and shorter whiskers for the Normal group compared to the AD group indicate a narrower spread of the modulation indices and less variability within the Normal group. Additionally, the presence of outliers in either group can provide insights into extreme observations or exceptional cases.

Scatter Plot: By examining the scatter plot, we can assess the association or correlation between the indices and identify any patterns or trends.

- Correlation: The scatter plot for the Normal group shows a strong positive correlation between the two indices, with data points clustering closely around a line, So it indicates a consistent relationship between the indices. Conversely, if the scatter plot for the AD group exhibits a weaker or no correlation, with data points scattered without a discernible pattern, it suggests a weaker association between the indices.
- Differences in Distribution: Comparing the distribution of data points between the AD and Normal groups, The scatter plot for the AD group appears more concentrated or exhibits a distinct pattern, it indicates a more consistent or stronger relationship between the indices in the AD group.

6 Conclusion

In this section, you are required to thoroughly examine and analyze the results you have obtained throughout this project. You must provide a comprehensive discussion of your findings, highlighting their significance and relevance to the research question. You should also present any limitations or weaknesses in your study and suggest possible areas for future research. Overall, this section is critical to demonstrating the quality and validity of your research and should be approached with careful attention to detail and clarity of expression.

Conclusion

In this study, we investigated the potential of phase locking value (PLV) and modulation index (MI) analysis during olfactory stimulation as biomarkers for Alzheimer's Disease (AD) in EEG signals. The data were pre-processed to remove artifacts and noise, ensuring the reliability of the subsequent analysis.

Our findings revealed significant differences between the AD and control groups in terms of PLV, MI, and phase differences during olfactory stimulation. The PLV analysis demonstrated decreased phase synchronization in specific channels among AD patients compared to the control group, indicating alterations in neural connectivity or synchronization.

Additionally, the MI analysis provided valuable insights into the strength and consistency of the phase-amplitude coupling during olfactory stimulation. The AD group exhibited reduced MI values compared to the control group, suggesting impaired modulation effects and disrupted communication between brain regions.

The phase difference analysis further supported these findings by showing distinct patterns and shifts in phase distribution between the two groups. The AD group displayed a wider distribution and less concentration of phase differences, indicating weaker phase locking patterns compared to the control group.

The plots generated for the channels of interest provided visual evidence of the observed differences. The polar histograms depicted a more concentrated distribution of phase differences around zero in the control group, indicating stronger phase locking, whereas the AD group showed a broader distribution. The boxplots illustrated lower median MI values in the AD group compared to the control group, highlighting the disrupted modulation effects in AD patients.

However, further investigations involving larger and more diverse cohorts are necessary to validate and generalize these findings. Additionally, exploring the relationship between PLV, MI, and clinical measures of cognitive decline would provide deeper insights into the potential clinical utility of these biomarkers.

In conclusion, our study highlights the relevance of PLV and MI analysis in EEG signals during olfactory stimulation for understanding the neural mechanisms underlying Alzheimer's Disease. These measures have the potential to enhance diagnostic accuracy and facilitate early intervention strategies for AD patients.

References

- [1] Mild cognitive impairment (mci). <https://www.mayoclinic.org/diseases-conditions/mild-cognitive-impairment/symptoms-causes/syc-20354578#:~:text=Overview,mental%20function%20has%20slipped.%22>. Last Reviewed: Jan. 18, 2023.
- [2] Neural oscillations – interpreting eeg frequency bands. <https://imotions.com/blog/learning/best-practice/neural-oscillations/>.
- [3] Neurodegenerative diseases. <https://www.niehs.nih.gov/research/supported/health/neurodegenerative/index.cfm>. Last Reviewed: June 09, 2022.
- [4] Marin C, Vilas D, Langdon C, Alobid I, López-Chacón M, Haehner A, Hummel T, and Mullol J. Olfactory dysfunction in neurodegenerative diseases. In *Curr Allergy Asthma Rep*, volume 18, 2018 Jun 15.
- [5] Sedghizadeh MJ, Hojjati H, Ezzatdoost K, Aghajan H, Vahabi Z, and Tarighatnia H. Olfactory response as a marker for alzheimer's disease: Evidence from perceptual and frontal lobe oscillation coherence deficit. *PLOS One*, 15(12), December 15, 2020.
- [6] T.T.K. Munia and S. Aviyente. Time-frequency based phase-amplitude coupling measure for neuronal oscillations. In *Scientific Reports* 9, 12441, 27 August 2019.
- [7] Mohammad Javad Sedghizadeh, Hamid Aghajan, and Zahra Vahabi. Brain electrophysiological recording during olfactory stimulation in mild cognitive impairment and alzheimer disease patients: An eeg dataset. *Data in Brief*, 48:109289, 2023.
- [8] Cécilia Tremblay and Johannes Frasnelli. Olfactory and Trigeminal Systems Interact in the Periphery. *Chemical Senses*, 43(8):611–616, 07 2018.