CSIC30100: Brain-Computer Interface **(Due: 11/07/2025)**

HW2: EEG Analysis

Instructor: Chun-Shu Wei
TA: Vasilii (瓦希利)

Student ID: 412540005

Submission Policy

Read all the instructions below carefully before you start working on the assignment, and before you make a submission. Each student must hand in your report (pdf) for this assignment.

- **PLAGIARISM IS STRICTLY PROHIBITED. (0 points for Plagiarism)**
- For mathematical problem(s), please show your work step by step and clarify the statement of the theorem you use (if any). Answering without mathematical derivations will get 0 points.
- Submission deadline: **2025.11.07 23:59:59 AM**.
- Late submission penalty formula:

$$\text{original score} \times (0.7)^{\#(\text{days late})}$$

Submission Format

- Each student submits **1** zip file including the following:
  - **1 report** (.pdf file)
  - **3 files with the code** (.m file or .ipynb) for each programming problem with dataset 1
- The report must contain **observations, results, and explanations**. Please name your zip file as hw2_studentID_Name.zip
- Illegal format penalty: **`-5 points`** for violating each rule of file format.

Pre-requisite

For the programming problems, it is suggested to use Matlab and EEGLab

Matlab

- [The official MATLAB page]
- [MATLAB Installation tutorial]
- EEGLab official installation page (v2020.0+ is recommended)

(Using the Python MNE package is allowed, but tutorials will not be provided.)

---

1. Multiple Choice

Problem 1

Assume the signal-to-noise ratio is defined as $SNR \equiv \frac{\textit{the amplitude of signal in voltage}}{\textit{the amplitude of noise in voltage}}$

Imagine that we are looking for a 5 µV ERP effect, and the noise is 10 µV in the single-trial EEG, giving us a 5:10 (or 1:2) signal-to-noise ratio on single trials. How many trials would we need to average to get a 2:1 signal-to-noise ratio in the averaged ERP waveform?

(Hint: event-related potential)

(A) 4
(B) 8
(C) 16
(D) 32
(E) 64

Problem 2

The following are techniques that are commonly applied to EEG data. Which ones are unsupervised? (**there may be more than one correct answer**)

(A) PCA
(B) LDA
(C) CSP
(D) ICA
(E) K-means clustering

Problem 3

What's the number of components ICA provides when applying to EEG data? (**there may be more than one correct answer**)

(A) Twice the number of EEG channels
(B) A fixed number, typically 20, regardless of EEG channels
(C) The same as the number of EEG channels
(D) Half the number of EEG channels
(E) As many as the number of independent brain sources

2. Programming Problems

2.1. EEG Dataset

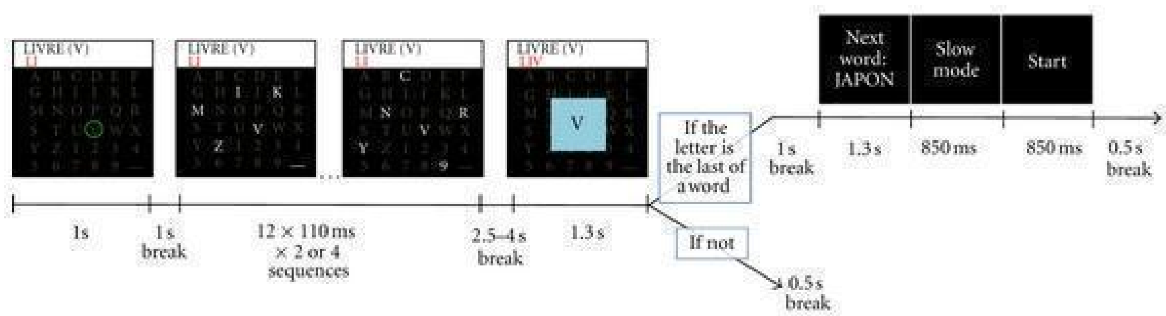
Dataset 1 Description:

Please download the dataset at the link below

[Dataset]

After downloading the file, open it using EEGLab. EEGLab is a MATLAB toolbox for analyzing EEG data. The official tutorial is provided on the EEGLab wiki website; we also provide a simple tutorial video at the link. To open the dataset, click **File** and choose the sub-menu item **Load existing dataset** in the EEGLAB graphic interface.

The dataset is acquired from the BCI challenge and has been transformed into a .set file. This specific dataset used in this assignment contains the EEG recording of subject 02, the fifth session. The experimental paradigm of one trial is shown below.



In each trial, the subject is asked to focus on a particular letter, so that **P300** occurred when that letter flashed. The spelling system would determine the letter through the subject's brain wave. Then 2.5-4s after the flash period, the letter selected by the system would show up on the screen, which is called the feedback event. The dataset above contains two event types - one is **"FeedBack_correct"** and the other is **"FeedBack_wrong"**. The **"FeedBack_correct"** event means that the selected letter matches the subject's intention. Otherwise, the **"FeedBack_wrong"** event corresponds to the wrong letter selected by the system. Error-related potential occurs after the onset of the "FeedBack_wrong" event.

Dataset 2 Description:

In Lab 1, you have collected a 4-channel EEG recording for at least 5 minutes. To complete this assignment, you have to:

- Convert exported .csv file from EEG cap into EEGLAB (or Python MNE) compatible format.
- Find electrode position from the EEGLAB (or Python MNE) packages for topographical plots.

You can find our tutorial for the above procedure using EEGLAB at [link] and the location file at [link]

2.2. EEG Dataset Preprocessing

Problem 1-1

Please follow the following steps for **Dataset 1**:

1. Plot 2D channel location map
2. Run ICA and record the computational time of ICA by code.
3. Plot component maps in 2D.
4. Indicate noise component(s) if they exist and explain the reason why you identify this component as noise or artifact.
5. Plot the first 10-second channel data before and after deleting noise/artifact component(s).

Problem 1-2

Please follow the following steps for **Dataset 2**:

1. Plot 2D channel location map
2. Plot spectra and map in 2D.
3. Plot the first 10-second channel data, and discuss anything you observed.

Problem 2

Please follow the following steps for **Dataset 1**:

1. Plot 2D channel location map
2. **Bandpass filtering [1, 48]*Hz*.**
3. Run ICA and record the computational time of ICA by code.
4. Plot component maps in 2D.
5. Indicate noise component(s) if they exist and explain the reason why you identify this component as noise or artifact.
6. Plot the first 10-second channel data before and after deleting noise/artifact component(s).
7. **Discuss the effect of bandpassing(highpassing) the signal before running ICA.**

2.3. Independent Component Analysis and Artifact Removal

Problem 3

Please follow the following steps for **Dataset 1**:

According to Hu et al, the SNR of an ERP waveform can be defined as below:

$$SNR = \frac{peak\ amplitude\ of\ error{-}related\ potential\ (0\ to\ 1000ms)}{standard\ deviation\ of\ the\ ERP\ waveform\ in\ the\ pre{-}stimulus\ interval(-200\ to\ 0\ ms)}$$

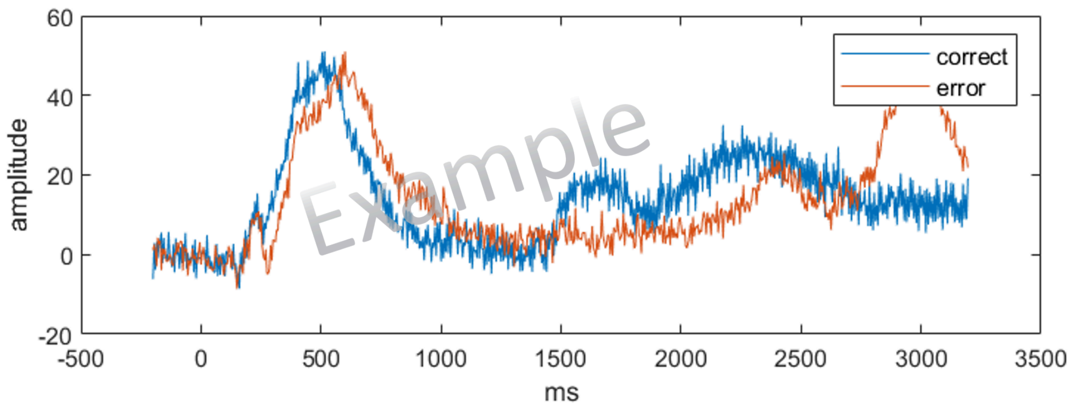
Since the error-related potential originates from the **anterior cingulate cortex(ACC)**, we focus on the **FCz** channel.

1. Apply all four following preprocessing flows before calculating ERP at FCz:
   A. Without any operation
   B. Bandpass the signal (1~48 Hz)
   C. Run ICA and remove bad components(Hint: using ICLabels)
   D. Bandpass the signal (1~48 Hz) first and run ICA to remove bad components

2. After the preprocessing, epoch the continuous EEG with a time interval [-0.2 1.3] sec, where t=0 is the feedback onset. (Hint: EEGLAB epoch)

3. Remove the epoch baseline mean.

4. Plot the ERP at FCz time-locked to the two different events(i.e the correct and error feedback) (Hint: In the MATLAB workspace, you can see an EEG structure that contains all the information of the current EEGLAB dataset. EEG.data is an array of shapes (num_channel, num_sample, num_trial))
   Example:



5. Fill out the table below

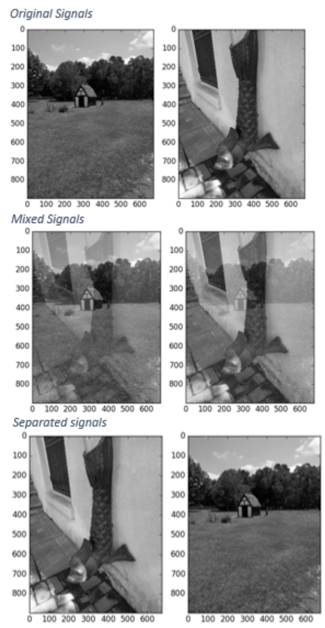
Preprocessing Methods	**ERP plot for 2 types of feedback**	**SNR(error feedback only)**
Without any operation		
Bandpass only		
IC removal only		
Bandpass+IC removal		

3. Bonus Problem

3.1. Independent Component Analysis

3.1.1. Motivation: Blind Source Separation

Blind Source Separation is a problem in which we try to separate a set of source signals from a set of mixed signals without the aid of or with little aid of information about the source signals or mixing process. In short, the objective of this problem is to recover the original components from mixtures of signals. Below is a visual example of blind source separation.



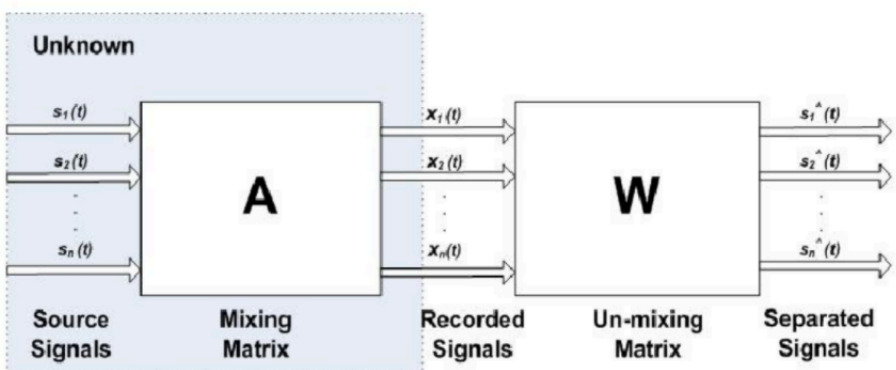


Figure: Blind Source Separation (BSS) [Naik and Kumar,2011]

The goal of BSS is to estimate A and S so that S' provides unknown source signals as possible.

$$X = AS + E \leftarrow X = A'S'$$

3.1.2. Cocktail Party Problem

One classic example of Blind Source Separation (BSS) is the Cocktail Party Problem, where people are talking simultaneously in a room and the listener/observer is trying to listen to one of the conversations. While humans can easily solve this problem, it is a hard problem in Digital Signal Processing

Let X be a recorded signal and S is a source signal according to the below formalization. We assume that $\{s_i \mid i = 1, 2, ..., n\}$ is statistically independent.

$$X = \hat{A}\hat{S} \iff \begin{bmatrix} - & x_1^T & - \\ - & x_2^T & - \\ \vdots & \vdots & \vdots \\ - & x_m^T & - \end{bmatrix} = \hat{A}_{m\times n} \begin{bmatrix} - & \hat{s}_1^T & - \\ - & \hat{s}_2^T & - \\ \vdots & \vdots & \vdots \\ - & \hat{s}_n^T & - \end{bmatrix}$$

Independent Component Analysis is to estimate the independent component S from X.

Hypothesis of ICA

- $\{s_j \in \mathbb{R}^{d\times 1} | j \in \mathbb{Z}_n\}$ statistically independent, that is, $P(s_1, ..., s_n) = \prod_{j=1}^{n} P(s_j)$
- $\{s_j \in \mathbb{R}^{d\times 1} | j \in \mathbb{Z}_n\}$ follows the Non-Gaussian distribution.
- A is regular

Therefore, we could rewrite the model as $\hat{S} = \hat{B}X$ where $\hat{B} = \hat{A}^{-1}$. It's only necessary to estimate B (compute $\hat{B}$) so that $\{s_j \in \mathbb{R}^{d\times 1} | j \in \mathbb{Z}_n\}$ is independent.

Definition 1.2 *White signal*
White signals are defined as any $z \in \mathbb{R}^{d\times 1}$ which satisfying

- *Zero mean:* $E[z] = 0 = m_z$
- *Unit covariance:* $C_z = E[(z - m_z)(z - m_z)^T] = E[zz^T] = I_d$

From now on we assume that $m = n$ to simplify the model. Whitening is useful for PCA and simplifies ICA problems. If we denote whitening signal as

$$Z_{d\times m} = V_{d\times d} X^T_{d\times m} \iff \begin{bmatrix} | & | & \cdots & | \\ z_1 & z_2 & \cdots & z_m \\ | & | & \cdots & | \end{bmatrix} = V_{d\times d} \begin{bmatrix} | & | & \cdots & | \\ x_1 & x_2 & \cdots & x_m \\ | & | & \cdots & | \end{bmatrix}$$

Where $V \in R_{d\times d}$ is a whitening matrix of $X_{m\times d}$, then model becomes

$$\hat{S}^T_{d\times m} = U_{d\times d} Z_{d\times m} = U_{d\times d} V_{d\times d} X^T_{d\times m} = \hat{B}_{d\times d} X^T_{d\times m} \iff \begin{bmatrix} | & | & \cdots & | \\ \hat{s}_1 & \hat{s}_2 & \cdots & \hat{s}_m \\ | & | & \cdots & | \end{bmatrix} = \hat{B}_{d\times d} \begin{bmatrix} | & | & \cdots & | \\ x_1 & x_2 & \cdots & x_m \\ | & | & \cdots & | \end{bmatrix}$$

Where $U \in R_{d\times d}$ is an orthogonal transformation matrix.

Hence it's necessary to estimate U!

The gaussianity of X (sums of non-gaussian random variables) must be larger than S (original) according to the Central Limit Theorem. Let $\{x_j \in R_{d\times 1} | j \in Z_m\}$ be the observed signals, we want to maximize the non-gaussianity of source signals $s_j = B \times j$.

Kurtosis is a measure of non-gaussianity

Definition 1.3 *Kurtosis*
for a random variable $y \in \mathbb{R}^{d\times 1}$,

$$kurt(y) = E[y^4] - 3(E[y^2])^2$$

That is, for white signal $z \in \mathbb{R}^{d\times 1}$,

$$kurt(z) = E[z^4] - 3(E[z^2])^2 = E[z^4] - 3$$

Which means we could solve ICA problem by

$$\hat{b} = \max_b \|kurt(b^T x)\| \tag{1.11}$$

Bonus Problem: Solving the CA problem by kurtosis (z is a white signal)

$$argmax_w \|kurt(w^T z)\| \text{ with } w^T w = 1$$

References

L. Hu, A. Mouraux, Y. Hu, G.D. Iannetti, A novel approach for enhancing the signal-to-noise ratio and detecting automatically event-related potentials (ERPs) in single trials, NeuroImage, Volume 50, Issue 1,2010, Pages 99-111. (link)

Perrin, M., Maby, E., Daligault, S., Bertrand, O., & Mattout, J. Objective and subjective evaluation of online error correction during P300-based spelling. Advances in Human-Computer Interaction, 2012, 4. (link)

Winkler I, Debener S, Müller KR, Tangermann M. On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. Annu Int Conf IEEE Eng Med Biol Soc. 2015;2015:4101-5. doi: 10.1109/EMBC.2015.7319296. PMID: 26737196 (link)