HW2: EEG Analysis

Wei-Yun Hsu
Institute of Multimedia Engineering
National Yang Ming Chiao Tung University

1. Multiple Choice

Problem 1

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

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 together 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

Answer: (C)

Explanations:

If we want to increase the signal-to-noise ratio in the averaged ERP waveform from 1:2 to 2:1, we need to increase the signal amplitude relative to the noise by a factor of 4. This means we need to reduce the noise amplitude to 2.5 μV while keeping the signal amplitude at 5 μV.

SNR_single = signal amplitude / noise amplitude = 5/10 = 0.5

SNR_avg = sqrt(N) * SNR_single, where SNR_avg = 2 and SNR_single = 0.5

So we get N = 16

Problem 2

The following are techniques that are commonly applied to EEG data. Which ones are unsupervised ?

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

Answer: (A)(D)(E)

Explanations:

PCA and ICA are used for feature extraction and dimensionality reduction, while K-means clustering is used for grouping similar patterns together in the data. Therefore, they don't require prior knowledge of the class labels or targets and belong to unsupervised techniques in EEG analysis.

On the other hand, LDA is used for classification task, CSP is used for feature extraction and classification task, they are supervised methods.

2. Programming Problem

2.1 EEG Dataset

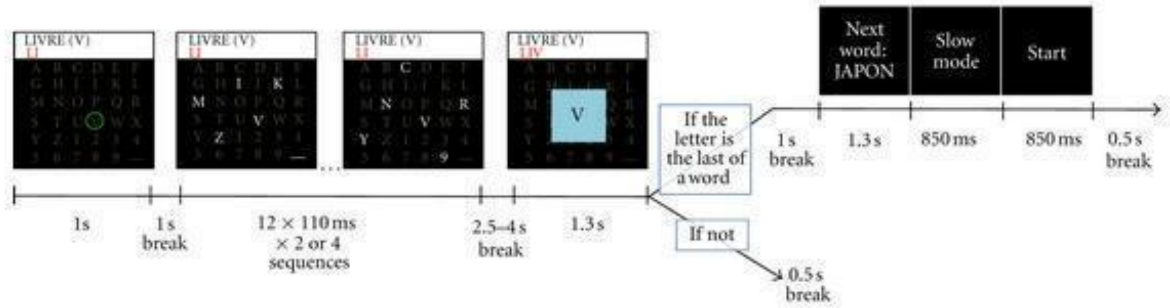
Please download the dataset at the link below

Dataset

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

Dataset description:

The dataset is acquired from 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.

2.2 EEG Dataset Preprocessing

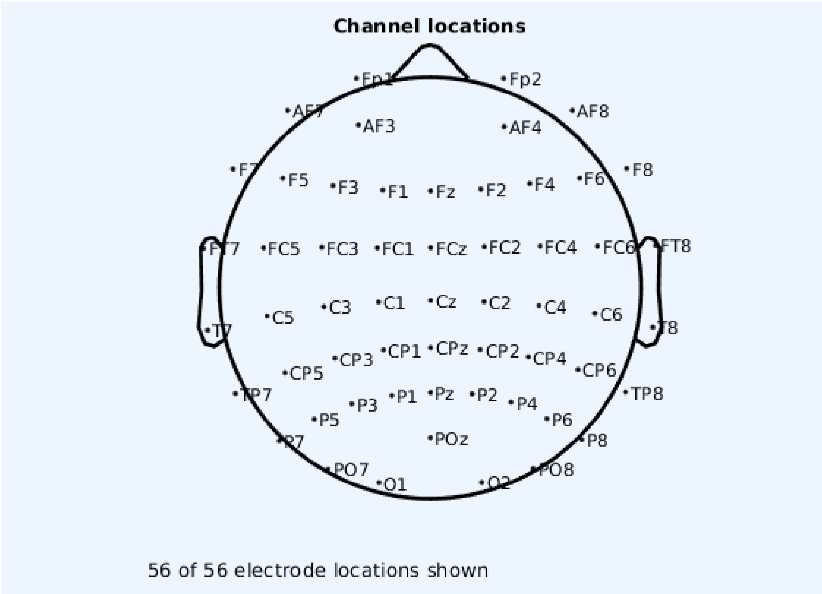
Problem 1

Please follow the following steps:

1. Plot 2D channel location map
2. Run ICA and record 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 artifacts.
5. Plot first 10-second channel data before and after deleting noise/artifact component(s).

Answer (**refer to problem1.m**)

1. Plot 2D channel location map

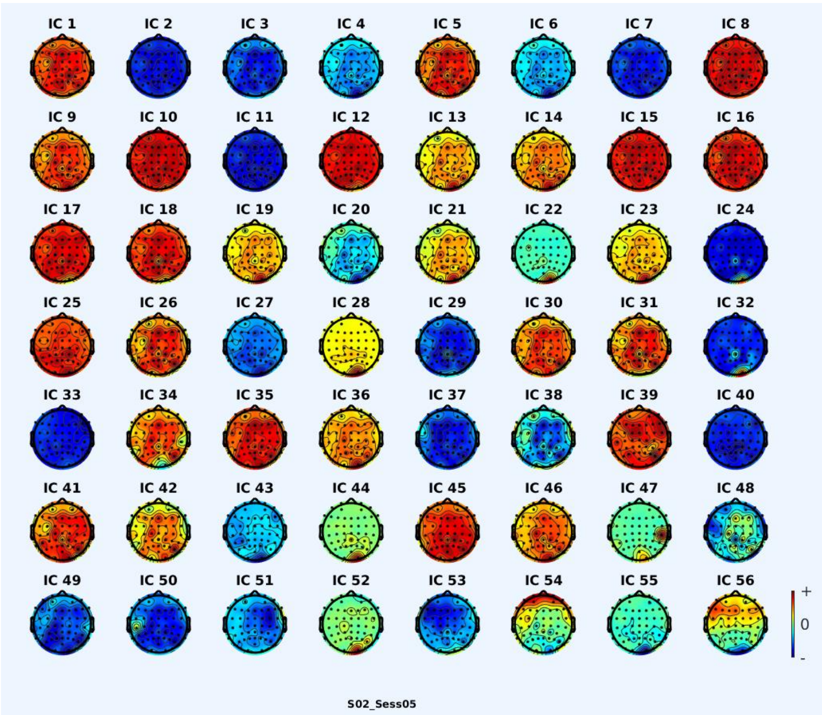


2. Run ICA and record computational time of ICA by code.

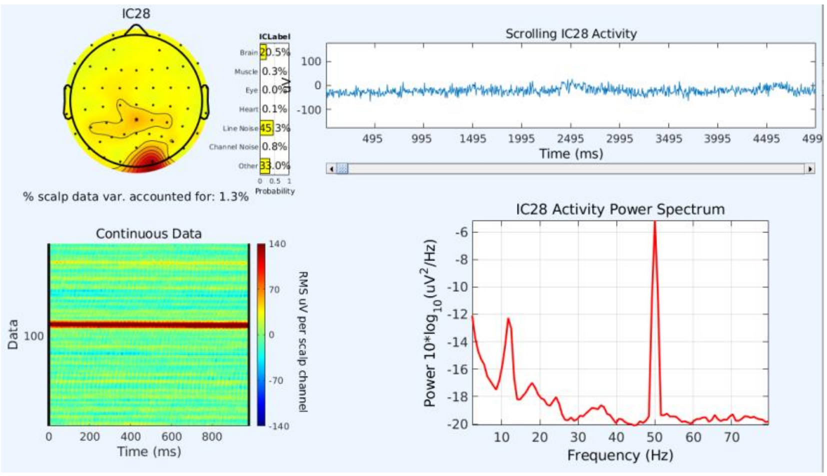
```
% Start the timer
tic;
EEG = pop_runica(EEG, 'icatype', 'runica', 'extended',1,'interrupt','on');
% Stop the timer and print the elapsed time in seconds
elapsed_time = toc;
disp(['ICA running time: ', num2str(elapsed_time), ' seconds']);
```

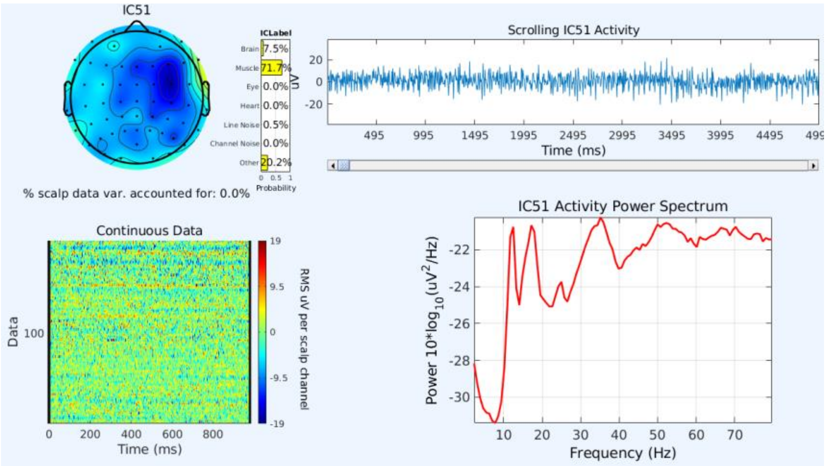
```
ICA running time: 272.8838 seconds
```

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 artifacts.

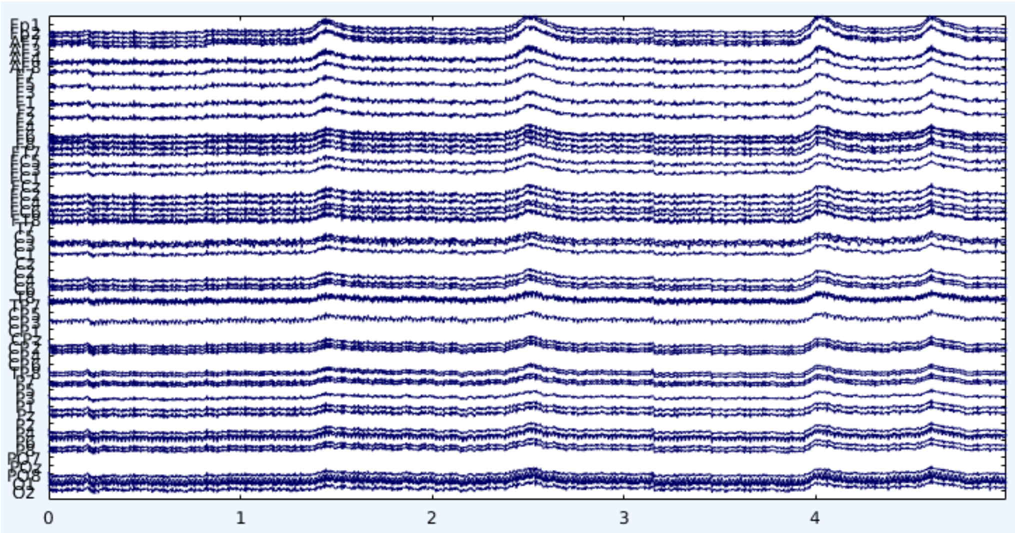




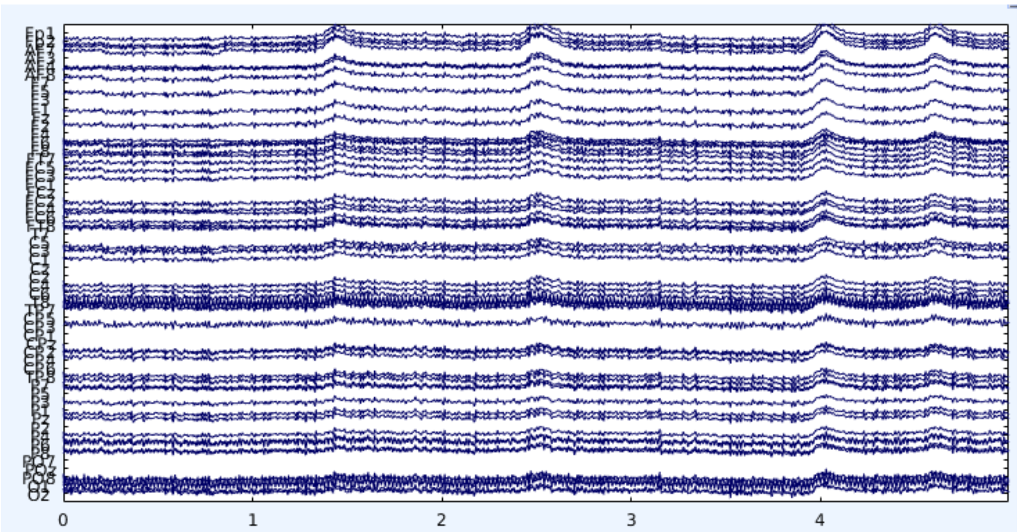
After decomposing data by ICA and classifying components using ICLabel, I observed this data exists five noise/artifacts (28, 37, 39, 47, 51), one of them is line noise, another are muscle artifacts. Here, I only selected two components to figure out and explain. The upper figure is a line noise example, we can see that it has a peak at 50 Hz, so this component contain line noise problem. For the bottom figure, it is a muscle artifact example, it is obvious that the spectrum has very unstable power (low frequency) and high-amplitude situations, we can say that this component maybe exist the muscle artifact problem.

5. Plot first 10-second channel data before and after deleting noise/artifact component(s).

Before



After



Observations:

In my observation, there doesn't seem to be much difference between the two figures above.

Problem 2

Please follow the following steps:

1. Plot 2D channel location map

2. **Bandpass filtering [1, 48]*Hz*.**

3. Run ICA and record 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 artifacts.

6. Plot 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.**

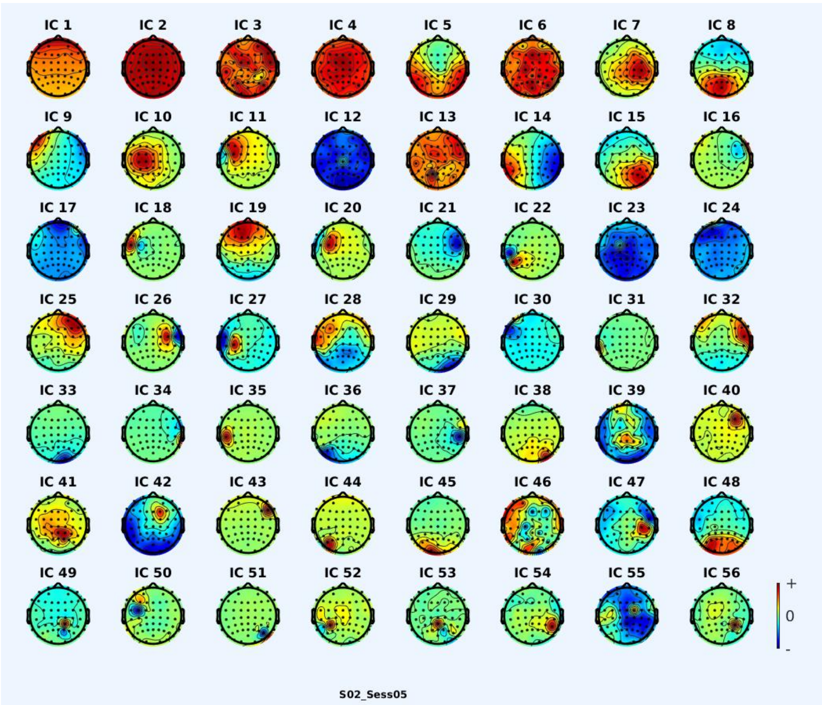
Answer **(refer to problem2.m)**

1. Same as the Problem 1
2. **Bandpass filtering [1, 48]Hz**
3. Run ICA and record computational time of ICA by code.

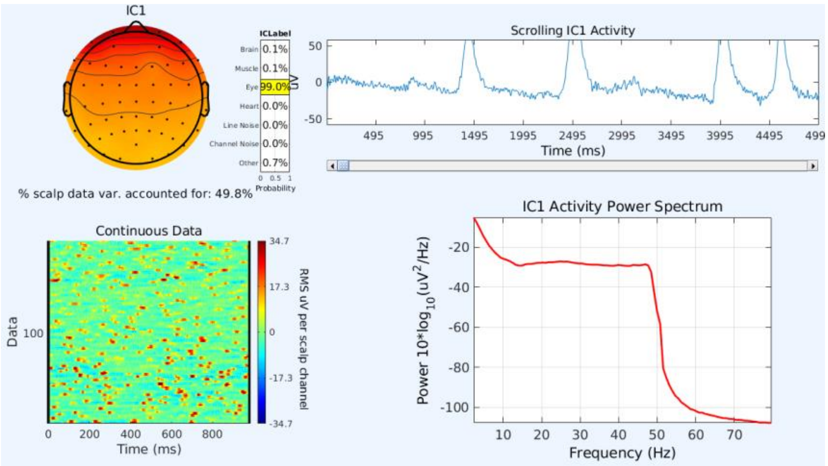
```
% Start the timer
tic;
EEG = pop_runica(EEG, 'icatype', 'runica', 'extended',1,'interrupt','on');
% Stop the timer and print the elapsed time in seconds
elapsed_time = toc;
disp(['ICA running time: ', num2str(elapsed_time), ' seconds']);
```

```
ICA running time: 279.1334 seconds
```

4. Plot component maps in 2D.

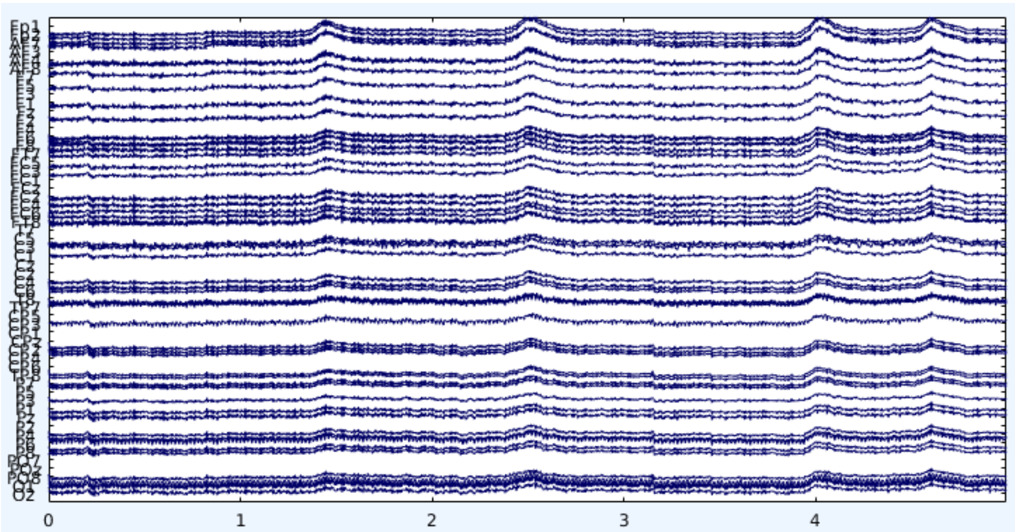


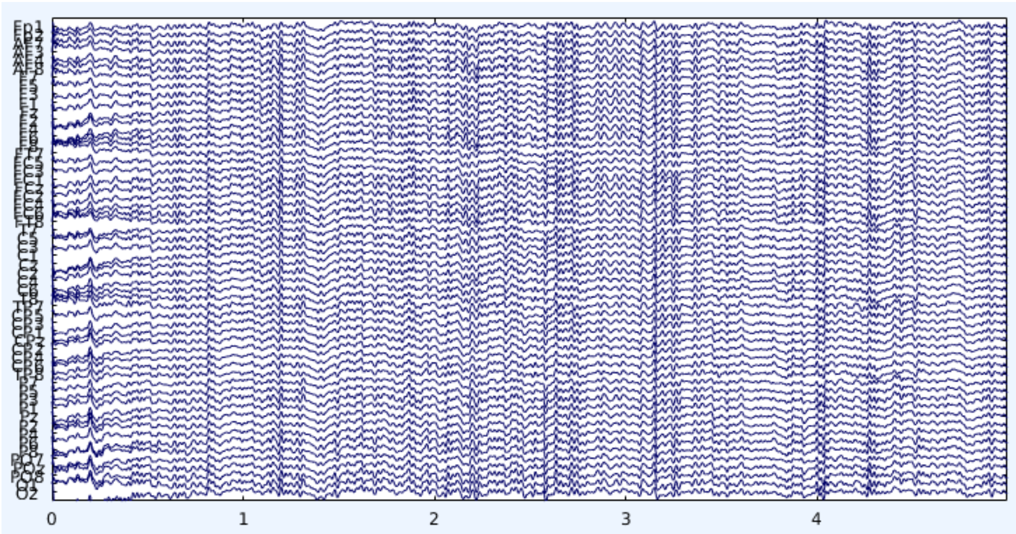
5. Indicate noise component(s) if they exist and explain the reason why you identify this component as noise or artifacts.



For the above figure, I think it is an eye artifact component, because eye artifacts are typically characterized by transient high-amplitude deflections that occur at specific frequencies (1-10Hz).

6. Plot first 10-second channel data before and after deleting noise/artifact component(s).

Before

After



Observations:

By comparison, I observed that the channel data was reduced the amplitude. Noise and artifacts can contribute significant amounts of signal energy to EEG data, leading to increased the amplitude and variability. After removing these components, the overall amplitude of the signal may be reduced, resulting in a cleaner and more stable recording.

7. Discuss the effect of bandpassing (highpassing) the signal before running ICA.

Compare with Problem1, we bandpass the signal before running ICA, I found that the more components were identified as having noise or artifacts. I think the reason is bandpass filtering can improve separation of signal and noise/artifacts, it make easier for ICA to separate the remaining sources. From the question 4 in Problem 1 and Problem 2, bandpass filtering would change the component topographies, specifically if the cutoff frequencies are close to the frequency range of the source. And the other hand, it still has some disadvantages, one of them is potential loss of information, if the important signal components are located near the frequency cutoff.

2.3 Independent Component Analysis and Artifact Removal (refer to problem3.m)

Problem 3

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

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

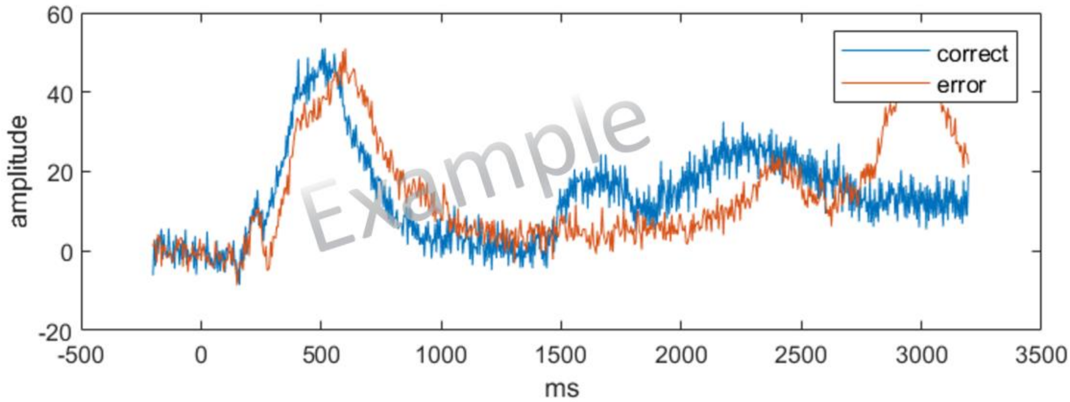
Since the error-related potential originates from the **anterior cingulate cortex(ACC)**, we focus on the **FCz** channel.

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)

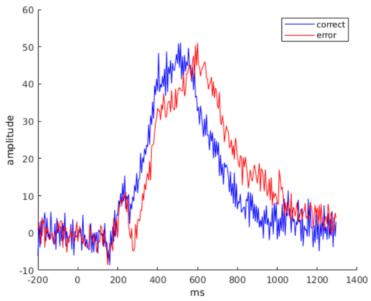
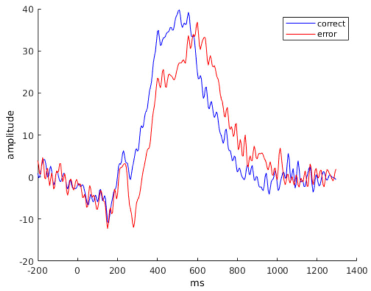
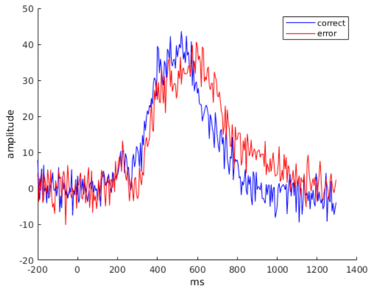
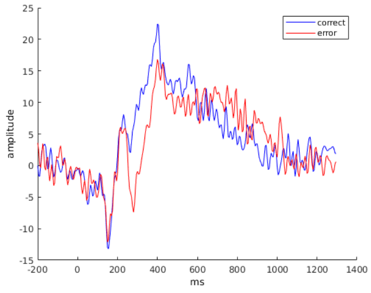
3. Remove the epoch baseline mean.

4. Plot the ERP at FCz time-locked to the two different events(i.e the correct and error feedbacks)
   (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 shape (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		21.5509
Bandpass only		17.0426
IC removal only		10.7136
Bandpass+IC removal		9.3673