IOC 5275: Brain Computer Interface **(Due: 05/04/2021)**

Lab3: Event Related Potential Analysis

Instructor: Chun-Shu Wei *Group Number:* 1
TA: Min-Jiun Tsai *0810834, 0713223, 309591027, 0716307, 0860838, 0617028*

Submission Policy

Read all the instructions below carefully before you start working on the assignment, and before you make a submission. For this assignment, please hand in the following your report (pdf) and code (.ipynb or .m file).

- **PLAGIARISM IS STRICTLY PROHIBITED. (0 point for Plagiarism)**
- For mathematical problem(s), please show your work step by step and clarify statement of theorem you use (if any). Answering without mathematical derivations will get 0 point.
- Submission deadline: **2021.05.04 09:00:00 AM**.
- **Late submission penalty formula:**

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

File Format

- Each group submits 1 report (.pdf and .tex file) and 1 code (.ipynb or .m).
- **Report** must contains observations, results and explanations. Please name your .pdf and .tex file as **5275_Lab3_GroupNum.pdf** and **5275_Lab3_GroupNum.tex**,respectively.
- Paper submission is not allowed. **Please use our LaTeX template to complete your report**.
- **Code** file must contains comments to explain your code. Please name your code file as **5275_Lab3_GroupNum.ipynb/.m**
- Implementation will be graded by completeness, algorithm correctness, model description, and discussion.
- **Illegal format penalty:** −5 points for violating each rule of file format.

Prerequest

To finish programming problem, you could choose Matlab or Python base on your programming preference.

Matlab 2020a+

- NYCU installation page
- NCTU installation tutorial
- EEGLab official installation page (v2020.0+ is recommended)

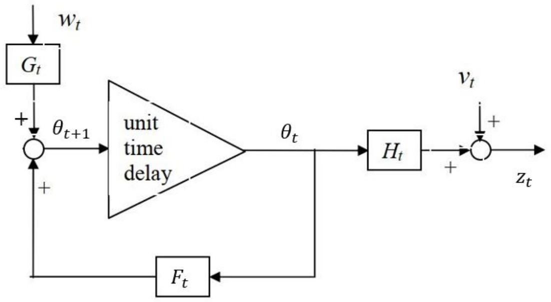
Python 3.7+

- MNE official installation page (0.20.7+ is recommended)

1 Mathematical problem

1.1 Denoising the ERP signals–Kalman Filtering

1.1.1 Framework of Kalman filtering



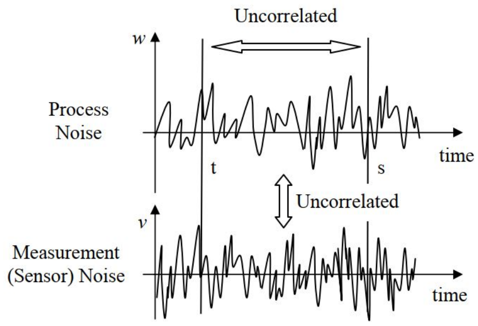


Figure: (Left) State space representation of linear time varying system (ERP) with process noise and measurement noise. (Right) Noise characteristics.

Model Formalization

Let $\theta_t \in \mathbb{R}^{k\times 1}$ denotes the input signal (state) and $z_t \in \mathbb{R}^{M\times 1}$ denotes the output signal at time t. According to *S. D. Georgiadis et al.*, both θ_t and z_t are defined as vector-valued processes.We could construct a state-space model for the linear dynamic systems (ERP) as below:

$$\theta_{t+1} = F_t\theta_t + G_t w_t,\ z_t = H_t\theta_t + v_t \tag{1.1}$$

with initial condition θ_0, Here we define $w_t \in \mathbb{R}^{k\times 1}$ as the process noise with zero mean, $v_t \in \mathbb{R}^{M\times 1}$ as the measurement noise with zero mean, $F_t \in \mathbb{R}^{k\times k}$ as the transition matrix, and $H_t \in \mathbb{R}^{M\times k}$, $G_t \in \mathbb{R}^{k\times k}$.

Hypothesis of ERP system

- F_t, G_t, and H_t are known sequences of matrices
- (θ_0, w_t, v_t) is a sequence of mutually uncorrelated random vectors with finite variance
- $E[w_t] = \mathbf{0}_{k\times 1},\ E[v_t] = \mathbf{0}_{M\times 1}\ \forall t$
- The covariances C_w, C_v, and C_{wv} are known sequences of matrices.

Given the following conditions:

$$C_v(t,s) = E[v_t \cdot v_s^T] = \begin{cases} \mathbf{0}_{M\times M},\ where\ t \neq s \\ R_{t,M\times M},\ where\ t = s \end{cases} \tag{1.2}$$

$$C_w(t,s) = E[w_t \cdot w_s^T] = \begin{cases} \mathbf{0}_{k\times k},\ where\ t \neq s \\ Q_{t,k\times k},\ where\ t = s \end{cases} \tag{1.3}$$

$$C_{wv}(t,s) = E[w_t \cdot v_s^T] = \mathbf{0}_{k\times M},\ \forall t,\ \forall s \tag{1.4}$$

To obtain an optimal value of θ_t based on measurements z_t, we minimize the mean square error:

$$J_t = E[(\hat{\theta}_t - \theta_t)^T(\hat{\theta}_t - \theta_t)] \tag{1.5}$$

with the constraints (1.1) to (1.4). We call $\hat{\theta}_{t|t-1}$ as expected state transition based on model, $\hat{z}_t$ as expected output.

$$\begin{aligned} \theta_t &= F_{t-1}\theta_{t-1} + G_{t-1}w_{t-1} \\ \hat{\theta}_{t|t-1} &= E[F_{t-1}\hat{\theta}_{t-1} + G_{t-1}w_{t-1}] = F_{t-1}\hat{\theta}_{t-1} + G_{t-1}E[w_{t-1}] \\ \hat{z}_t &= E[H_t\hat{\theta}_{t|t-1} + v_t] = H_t\hat{\theta}_{t|t-1} + E[v_t] = H_t\hat{\theta}_{t|t-1} \end{aligned} \tag{1.6}$$

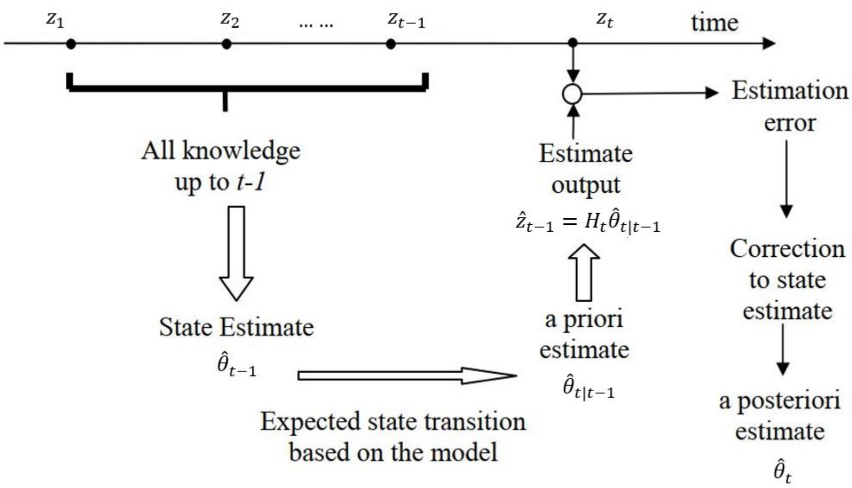


Figure: Outline of the Kalman filter algorithm

1.1.2 Correction of the state estimate

Assimilating the new measurement z_t, we can update the state estimate in proportion to the output estimation error.

$$\hat{\theta}_t = \hat{\theta}_{t|t-1} + K_t(z_t - \hat{z}_t) \tag{1.7}$$

Let $\epsilon_t = \hat{\theta}_{t|t-1} - \theta_t$ be a priori estimation error, then

$$\begin{aligned} e_t \equiv \hat{\theta}_t - \theta_t &= \hat{\theta}_{t|t-1} + K_t\big(z_t - H_t\hat{\theta}_{t|t-1}\big) - \theta_t \\ &= \hat{\theta}_{t|t-1} + K_t\big(H_t\theta_t + v_t - H_t\hat{\theta}_{t|t-1}\big) - \theta_t = \big(\hat{\theta}_{t|t-1} - \theta_t\big) - K_tH_t\big(\hat{\theta}_{t|t-1} - \theta_t\big) + K_tv_t \\ &= \big(I_k - K_tH_t\big)\epsilon_t + K_tv_t \end{aligned} \tag{1.8}$$

Equation (1.7) provides a structure of linear filter in recursive form. Denoting $K_t \in \mathbb{R}^{k\times M}$ as a gain matrix to be optimized so that the mean squared error (expectation of $e_t^Te_t$) of state estimation may be minimized.

Problem 1 (5 points)

Please show that

$$e_t^Te_t = \epsilon_t^T\epsilon_t + \epsilon_t^TH_t^TK_t^TK_tH_t\epsilon_t - 2\epsilon_t^TK_tH_t\epsilon_t + 2\epsilon_t^TK_t^Tv_t - 2v_t^TK_t^TK_tH_t\epsilon_t + v_t^TK_t^TK_tv_t \tag{1.9}$$

Solution:
Because
$e_t = \epsilon_t v_t^t - K_tH_t\epsilon_t + K_tv_t$

So
$e_t^T = \epsilon_t^T - \epsilon_t^TH_t^TK_t^T + v_t^TK_t^T$

Then, we can get:

$e_t^Te_t = \epsilon_t^T\epsilon_t^T - \epsilon_t^TK_tH_t\epsilon_t + \epsilon_t^TK_tv_t - \epsilon_t^TH_t^TK_t^T\epsilon_t + \epsilon_t^TH_t^TK_t^TK_tH_t\epsilon_t - \epsilon_t^TH_t^TK_t^TK_tv_t + v_t^TK_t^T\epsilon_t - v_t^TK_t^TK_tH_t\epsilon_t + v_t^TK_t^TK_tv_t$

$= e_t^Te_t = \epsilon_t^T\epsilon_t + \epsilon_t^TH_t^TK_t^TK_tH_t\epsilon_t - 2\epsilon_t^TK_tH_t\epsilon_t + 2\epsilon_t^TK_t^Tv_t - 2v_t^TK_t^TK_tH_t\epsilon_t + v_t^TK_t^TK_tv_t$

Let us differentiate the scalar function $e_t^Te_t$ with respect to matrix K_t by using the following matrix differentiation rules.

Matrix differentiation rule 1
Let $a \in \mathbb{R}^{k\times 1}$, $b \in \mathbb{R}^{M\times 1}$, and $K \in \mathbb{R}^{k\times M}$ is same as above K_t (We omit the subscript t for brevity).

$$f \equiv [a_1, ..., a_k] \begin{bmatrix} K_{11} & \dots & K_{1M} \\ \vdots & \ddots & \vdots \\ K_{k1} & \dots & K_{kM} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_M \end{bmatrix} = a^T K b \Rightarrow \frac{df}{dK} = \left[\frac{\partial f}{\partial K_{ij}}\right] = [a_i b_j] = ab^T,\ \forall i \in \mathbb{Z}_k, j \in \mathbb{Z}_M \tag{1.10}$$

Matrix differentiation rule 2
Let $b, c \in \mathbb{R}^{M\times 1}$, and $K \in \mathbb{R}^{k\times M}$, and $g = c^T K^T K b$, then

$$\frac{dg}{dK_t} = \left[\frac{\partial g}{\partial K_{im}} \sum_{i=1}^{M}\sum_{j=1}^{k}\sum_{l=1}^{k} K_{il} c_l K_{ij} b_j\right] = \left[\sum_{j=1}^{k} c_m K_{ij} b_j + \sum_{j=1}^{k} K_{il} c_l b_m\right] = K b c^T + K c b^T \tag{1.11}$$

Problem 2 (15 points)

Use the matrix differentiation rule 1 and 2 to show that

$$\frac{de_t^T e_t}{dK_t} = 2K_t H_t \epsilon_t \epsilon_t^T H_t^T - 2[K_t H_t \epsilon_t v_t^T + K_t v_t \epsilon_t^T H_t^T] + 2K_t v_t v_t^T + 2[\epsilon_t v_t^T - \epsilon_t \epsilon_t^T H_t^T] \tag{1.12}$$

Solution:

From problem1, we get
$e_t^T e_t = \epsilon_t^T \epsilon_t + \epsilon_t^T H_t^T K_t^T K_t H_t \epsilon_t - 2\epsilon_t^T K_t H_t \epsilon_t + 2\epsilon_t^T K_t^T v_t - 2v_t^T K_t^T K_t H_t \epsilon_t + v_t^T K_t^T K_t v_t$

Now, we want to get the differentiation of $e_t^T e_t$ by K_t. We can get the answer by dividing $e_t^T e_t$ into several parts and using the matrix differentiation rule 1 and 2 as follow.
0. $\epsilon_t^T \epsilon_t$
The differentiation of this part is zero

1. $\epsilon_t^T H_t^T K_t^T K_t H_t \epsilon_t$
Use rule 2:
Now, b= $H_t \epsilon_t$ and c=$H_t \epsilon_t$, so the differentiation of this part is: $K_t H_t \epsilon_t \epsilon_t^T H_t^T + K_t H_t \epsilon_t \epsilon_t^T H_t^T = 2K_t H_t \epsilon_t \epsilon_t^T H_t^T$

2. $-2\epsilon_t^T K_t H_t \epsilon_t$ Use rule 1:
Now, b=$H_t \epsilon_t$ and a=ϵ_t , so the differentiation of this part is: $-2\epsilon_t \epsilon_t^T H_t^T$

3. $2\epsilon_t^T K_t^T v_t$
Use rule 1:
Now, b=v_t and a=ϵ_t , so the differentiation of this part is: $2\epsilon_t v_t^T$

4. $-2v_t^T K_t^T K_t H_t \epsilon_t$ Use rule 2:
Now, b= $H_t \epsilon_t$ and c=v_t, so the differentiation of this part is: -2$K_t H_t \epsilon_t v_t^T + K_t v_t \epsilon_t^T H_t^T$

5. $v_t^T K_t^T K_t v_t$ Use rule 2:
Now, b= v_t and c=v_t, so the differentiation of this part is: $K_t v_t v_t^T + K_t v_t v_t^T = 2K_t v_t v_t^T$

So, the differentiation of the $e_t^T e_t$ by K_t is the sum of result from 1 to 5:
$de_t^T e_t / dK_t = 2K_t H_t \epsilon_t \epsilon_t^T H_t^T - 2[K_t H_t \epsilon_t v_t^T + K_t v_t \epsilon_t^T H_t^T] + 2K_t v_t v_t^T + 2[\epsilon_t v_t^T - \epsilon_t \epsilon_t^T H_t^T]$

The necessary condition for the mean squared error of state estimate with respect to the gain matrix K_t is:

$$\frac{dJ_t}{dK} = \mathbf{0}_{k\times M}$$

Taking expectation of $e_t^T e_t$, differentiating it w.r.t. K_t and setting it to zero yield:

$$E[K_t H_t \epsilon_t \epsilon_t^T H_t^T - K_t H_t \epsilon_t v_t^T - K_t v_t \epsilon_t^T H_t^T + K_t v_t v_t^T + \epsilon_t v_t^T - \epsilon_t \epsilon_t^T H_t^T] = \mathbf{0}_{k\times M} \tag{1.13}$$

Which means that K_t and H_t can be factored out,

$$K_t H_t E[\epsilon_t \epsilon_t^T] H_t^T - K_t H_t E[\epsilon_t v_t^T] - K_t E[v_t \epsilon_t^T] H_t^T + K_t E[v_t v_t^T] + E[\epsilon_t v_t^T] - E[\epsilon_t \epsilon_t^T] H_t^T = \mathbf{0}_{k\times M} \tag{1.14}$$

Since we know that

$$\hat{\theta}_{t|t-1} = E[F_{t-1}\hat{\theta}_{t-1} + G_{t-1} w_{t-1}] = F_{t-1}\hat{\theta}_{t-1} \tag{1.15}$$

and

$$\epsilon_t = \hat{\theta}_{t|t-1} - \theta_t,$$

we can examine $E[\epsilon_t v_t^T] = E[(\hat{\theta}_{t|t-1} - \theta_t) v_t^T] = E[(\hat{\theta}_{t|t-1} v_t^T] - E[\theta_t v_t^T] = \mathbf{0}_{k\times M}$

Problem 3 (10 points)

Please show that

$$E[\epsilon_t v_t^T] = \mathbf{0}_{k\times M} \ \& \ E[v_t \epsilon_t^T] = \mathbf{0}_{M\times k} \tag{1.16}$$

Solution:

(1) show that $E[\epsilon_t v_t^T] = \mathbf{0}_{k\times M}$

$$E[\epsilon_t v_t^T] = E[(\hat{\theta}_{t|t-1} - \theta_t)v_t^T] = E[(\hat{\theta}_{t|t-1} v_t^T] - E[\theta_t v_t^T]$$

since (θ_0, w_t, v_t) is mutually uncorrelated random vectors

$$\Rightarrow \ E[\epsilon_t v_t^T] = \Big[E[(\hat{\theta}_{t|t-1})]\Big]_{(k\times 1)} \Big[E[v_t^T]\Big]_{(1\times M)} - \Big[E[\theta_t]\Big]_{(k\times 1)} \Big[E[v_t^T]\Big]_{(1\times M)}$$

$$E[\hat{\theta}_{t|t-1}] = E[F_{t-1}\hat{\theta}_{t-1} + G_{t-1}E[w_{t-1}]] = E[F_{t-1}\hat{\theta}_{t-1}] + E[G_{t-1}E[w_{t-1}]] = E[F_{t-1}\hat{\theta}_{t-1}]$$

$$\begin{aligned} E[F_{t-1}\hat{\theta}_{t-1}] &= E[F_{t-1}(\hat{\theta}_{t-1|t-2} + K_{t-1}(z_{t-1} - \hat{z_{t-1}}))] \\ &= E[F_{t-1}(\hat{\theta}_{t-1|t-2} + K_{t-1}(H_{t-1}\theta_{t-1} + v_{t-1} - H_t\hat{\theta}_{t-1|t-2}))] \end{aligned}$$

since $\hat{\theta}_{t-1|t-2}$, θ_{t-1}, v_{t-1} are uncorrelated to v_t

$$\Rightarrow \ E[F_{t-1}\hat{\theta}_{t-1}]E[v_t^T] = \mathbf{0}_{K\times M}$$

$$\Rightarrow \ E[\hat{\theta}_{t|t-1}]E[v_t^T] = \mathbf{0}_{K\times M}$$

Also:

$$E[\theta_t] = E[F_{t-1}\theta_{t-1}] + E[G_t w_{t-1}] = E[F_{t-1}\theta_{t-1}] + G_t E[w_{t-1}] = E[F_{t-1}\theta_{t-1}]$$

$$\Rightarrow \ E[F_{t-1}\theta_{t-1}] = E[F_{t-1}(F_{t-2}\theta_{t-2} + G_{t-2}w_{t-2})]$$

since θ_{t-2}, w_{t-2} are uncorrelated to v_t

$$\Rightarrow \ E[F_{t-1}\theta_{t-1}]E[v_t^T] = \mathbf{0}_{K\times M}$$

$$\Rightarrow \ E[\theta_t]E[v_t^T] = \mathbf{0}_{K\times M}$$

$$\Rightarrow \ E[\epsilon_t v_t^T] = \mathbf{0}_{K\times M}$$

(2) show that $E[v_t\epsilon_t^T] = \mathbf{0}_{M\times K}$

$$E[v_t\epsilon_t^T] = \Big[E[v_t]\Big]_{M\times 1}\Big[E[(\hat{\theta}_{t|t-1} - \theta_t)^T]\Big]_{1\times K} = E[v_t]E[(\hat{\theta}_{t|t-1})^T] - E[v_t]E[(\theta_t)^T]$$

from (1) we can know that:

$$\begin{aligned} E[(\hat{\theta}_{t|t-1})^T] &= E[(F_{t-1}\hat{\theta}_{t-1})^T] \\ &= E[\Big(F_{t-1}(\hat{\theta}_{t-1|t-2} + K_{t-1}(H_{t-1}\theta_{t-1} + v_{t-1} - H_t\hat{\theta}_{t-1|t-2}))\Big)^T] \end{aligned}$$

since $\hat{\theta}_{t-1|t-2}$, θ_{t-1}, v_{t-1} are uncorrelated to v_t

$$\Rightarrow \; E[v_t]E[(F_{t-1}\hat{\theta}_{t-1})^T] = \mathbf{0}_{M\times K}$$

$$\Rightarrow \; E[v_t]E[(\hat{\theta}_{t|t-1})^T] = \mathbf{0}_{M\times K}$$

Also from (1):

$$E[(\theta_t)^T] = E[(F_{t-1}\theta_{t-1})^T] = E[(F_{t-1}(F_{t-2}\theta_{t-2} + G_{t-2}w_{t-2}))^T]$$

since θ_{t-2}, w_{t-2} are uncorrelated to v_t

$$\Rightarrow \; E[v_t]E[(F_{t-1}\theta_{t-1})^T] = \mathbf{0}_{M\times K}$$

$$\Rightarrow \; E[v_t]E[(\theta_t)^T] = \mathbf{0}_{M\times K}$$

$$\Rightarrow \; E[v_t\epsilon_t^T] = \mathbf{0}_{M\times K}$$

Let us define the error covariance of a priori state estimation

$$C_{\hat{\theta}_{t|t-1}} \equiv E[\epsilon_t\epsilon_t^T] = E[(\hat{\theta}_{t|t-1} - \theta_t)(\hat{\theta}_{t|t-1} - \theta_t)^T] \tag{1.17}$$

Problem 4 (5 points)

Please show that

$$K_t = C_{\hat{\theta}_{t|t-1}} H_t^T \left(H_t C_{\hat{\theta}_{t|t-1}} H_t^T + C_v(t,t) \right)^{-1} \tag{1.18}$$

Such K_t is called the **Kalman Gain**.

Solution:

By eq.(1.14), we know that:

$$K_t H_t E[\epsilon_t \epsilon_t^T] H_t^T - K_t H_t E[\epsilon_t v_t^T] - K_t E[v_t \epsilon_t^T] H_t^T + K_t E[v_t v_t^T] + E[\epsilon_t v_t^T] - E[\epsilon_t \epsilon_t^T] H_t^T = \mathbf{0}_{k\times M}$$

since $E[\epsilon_t v_t^T] = \mathbf{0}_{k\times M}$ & $E[v_t \epsilon_t^T] = \mathbf{0}_{M\times k}$ *by problem3*

$$\Rightarrow \; K_t H_t E[\epsilon_t \epsilon_t^T] H_t^T - \mathbf{0}_{k\times M} - \mathbf{0}_{k\times M} + K_t E[v_t v_t^T] + \mathbf{0}_{k\times M} - E[\epsilon_t \epsilon_t^T] H_t^T = \mathbf{0}_{k\times M}$$

$$\Rightarrow \; K_t H_t E[\epsilon_t \epsilon_t^T] H_t^T + K_t E[v_t v_t^T] = \Big[E[\epsilon_t \epsilon_t^T] H_t^T \Big]_{(k\times M)}$$

$$\Rightarrow \; K_t \Big(H_t E[\epsilon_t \epsilon_t^T] H_t^T + E[v_t v_t^T] \Big) = \Big[E[\epsilon_t \epsilon_t^T][H_t^T \Big]_{(k\times M)}$$

since $C_{\hat{\theta}_{t|t-1}} \equiv E[\epsilon_t \epsilon_t^T]$, and by definition of eq.(1.2): $C_w(t,s) = E[w_t \cdot w_s^T]$

$$\Rightarrow \; K_t \Big(H_t C_{\hat{\theta}_{t|t-1}} H_t^T + C_w(t,t) \Big) = C_{\hat{\theta}_{t|t-1}} H_t^T$$

$$\Rightarrow \; K_t = C_{\hat{\theta}_{t|t-1}} H_t^T \left(H_t C_{\hat{\theta}_{t|t-1}} H_t^T + C_w(t,t) \right)^{-1}$$

1.1.3 Updating the Error Covariance

The above Kalman gain contains the a priori error covariance $C_{\hat{\theta}_{t|t-1}}$. This must be updated recursively based on each new measurement and the state transition model. Define the a posteriori state estimation error covariance

$$C_{\hat{\theta}_t} = E[e_t e_t^T] = E[(\hat{\theta}_t - \theta_t)(\hat{\theta}_t - \theta_t)^T] \tag{1.19}$$

Problem 5 (5 points)

Please show that

$$C_{\hat{\theta}_t} = (I_k - K_t H_t) C_{\hat{\theta}_{t|t-1}} \tag{1.20}$$

Solution:

$C_{\hat{\theta}_t} = Cov(e_t)$

$\quad = Cov[(I_k - K_t H_t)(\hat{\theta_{t|t-1}} - \theta_t) + K_t V_t]$

$\because$ V_t is not correlated with other items,

$\therefore C_{\hat{\theta}_t} = Cov[(I_k - K_t H_t)(\hat{\theta_{t|t-1}} - \theta_t)] + Cov[K_t V_t]$

$\quad = (I_k - K_t H_t) Cov[\hat{\theta}_{t|t-1} - \theta_t](I_k - K_t H_t)^T \ + \ K_t Cov(V_t) K_t^T$

$\because Cov(\hat{\theta}_{t|t-1} - \theta_t) = C_{\hat{\theta}_{t|t-1}}$

$\therefore C_{\hat{\theta}_t} \ = \ (I_k - K_t H_t) C_{\hat{\theta}_{t|t-1}} (I_k - K_t H_t)^T \ + \ K_t Cov(V_t) K_t^T$

$\because I^T = I$ and $(AB)^T = B^T A^T$

$\therefore C_{\hat{\theta}_t} \ = \ (I_k - K_t H_t) C_{\hat{\theta}_{t|t-1}} (I_k - H_t^T K_t^T) \ + \ K_t Cov(V_t) K_t^T$

From equation (1.14) and (1.16),

$K_t E[v_t v_t^T] = C_{\hat{\theta}_{t|t-1}} H_t^T - K_t H_t C_{\hat{\theta}_{t|t-1}} H_t^T$

$K_t E[v_t v_t^T] K_t^T = C_{\hat{\theta}_{t|t-1}} H_t^T K_t^T - K_t H_t C_{\hat{\theta}_{t|t-1}} H_t^T K_t^T$

$\therefore C_{\hat{\theta}_t} \ = \ (C_{\hat{\theta}_{t|t-1}} - K_t H_t C_{\hat{\theta}_{t|t-1}})(I_k - K_t H_t) \ + \ C_{\hat{\theta}_{t|t-1}} H_t^T K_t^T \ - \ K_t H_t C_{\hat{\theta}_{t|t-1}} H_t^T K_t^T$

$= \ C_{\hat{\theta}_{t|t-1}} - K_t H_t C_{\hat{\theta}_{t|t-1}} \ - \ C_{\hat{\theta}_{t|t-1}} H_t^T K_t^T \ + \ K_t H_t C_{\hat{\theta}_{t|t-1}} H_t^T K_t^T \ + \ C_{\hat{\theta}_{t|t-1}} H_t^T K_t^T \ - \ K_t H_t C_{\hat{\theta}_{t|t-1}} H_t^T K_t^T$

$= \ C_{\hat{\theta}_{t|t-1}} - K_t H_t C_{\hat{\theta}_{t|t-1}} + (-C_{\hat{\theta}_{t|t-1}} H_t^T K_t^T + C_{\hat{\theta}_{t|t-1}} H_t^T K_t^T) + (K_t H_t C_{\hat{\theta}_{t|t-1}} H_t^T K_t^T - K_t H_t C_{\hat{\theta}_{t|t-1}} H_t^T K_t^T)$

$\Rightarrow C_{\hat{\theta}_t} \ = \ C_{\hat{\theta}_{t|t-1}} - K_t H_t C_{\hat{\theta}_{t|t-1}}$

$\Rightarrow C_{\hat{\theta}_t} \ = \ (I_k - K_t H_t) C_{\hat{\theta}_{t|t-1}}$

Derive (1.21), one can compute $C_{\hat{\theta}_{t+1|t}}$ by using the state transition equation (1.1). Consider

$$\epsilon_{t+1} = \hat{\theta}_{t+1|t} - \theta_{t+1|t} = F_t \hat{\theta}_t - (F_t \theta_t + G_t w_t) = F_t e_t - G_t w_t. \tag{1.21}$$

From (1.17)

$$\begin{aligned} C_{\hat{\theta}_{t+1|t}} &= E[\epsilon_{t+1} \epsilon_{t+1}^T] \\ &= E[(F_t \theta_t + G_t w_t)(F_t \theta_t + G_t w_t)^T] \\ &= F_t E[e_t e_t^T] F_t^T - G_t E[w_t e_t^T] F_t^T - F_t E[e_t w_t^T] G_t^T + G_t E[w_t w_t^T] G_t^T \end{aligned} \tag{1.22}$$

Problem 6 (5 points)

Please show that

$$C_{\hat{\theta}_{t|t-1}} = F_{t-1} C_{\hat{\theta}_{t-1}} F_{t-1}^T + G_{t-1} C_w(t-1, t-1) G_{t-1}^T \tag{1.23}$$

Solution:

Based on equation (1.15) and (1.16),

$\epsilon_t = \hat{\theta}_{t|t-1} - \theta_{t|t-1}$

$= F_{t-1}\hat{\theta}_{t-1} - [F_{t-1}\theta_{t-1}]$, and by equation (1.8)

$= F_{t-1}e_{t-1} - G_{t-1}w_{t-1}$

From equation (1.22),

$C_{\hat{\theta}_{t|t-1}} \ = \ E[\epsilon_t \epsilon_t^T]$

$= \ E[(F_{t-1}e_{t-1} - G_{t-1}w_{t-1})(F_{t-1}e_{t-1} - G_{t-1}w_{t-1})^T]$

$= \ F_{t-1}E[e_{t-1}e_{t-1}^T]F_{t-1}^T \ - \ G_{t-1}E[w_{t-1}e_{t-1}^T]F_{t-1}^T \ - \ F_{t-1}E[e_{t-1}w_{t-1}^T]G_{t-1}^T \ + \ G_{t-1}E[w_{t-1}w_{t-1}^T]G_{t-1}^T$

$\because$ e_{t-1} and w_{t-1} have zero cross-correlation because they are uncorrelated. In addition, noise w_{t-1} accumulates between $t-1$ and $t-2$ whereas the error e_{t-1} is error up until $t-1$

$\therefore \ E[w_{t-1}e_{t-1}^T] \ = \ 0, \ E[e_{t-1}w_{t-1}^T] \ = \ 0$

$\Rightarrow C_{\hat{\theta}_{t|t-1}} \ = \ F_{t-1}E[e_{t-1}e_{t-1}^T]F_{t-1}^T \ + \ G_{t-1}E[w_{t-1}w_{t-1}^T]G_{t-1}^T$

$\Rightarrow C_{\hat{\theta}_{t|t-1}} \ = \ F_{t-1}C_{\hat{\theta}_{t-1}}F_{t-1}^T \ + \ G_{t-1}C_w(t-1, t-1)G_{t-1}^T.$

1.1.4 The Recursive Calculation Procedure for the Discrete Kalman Filter

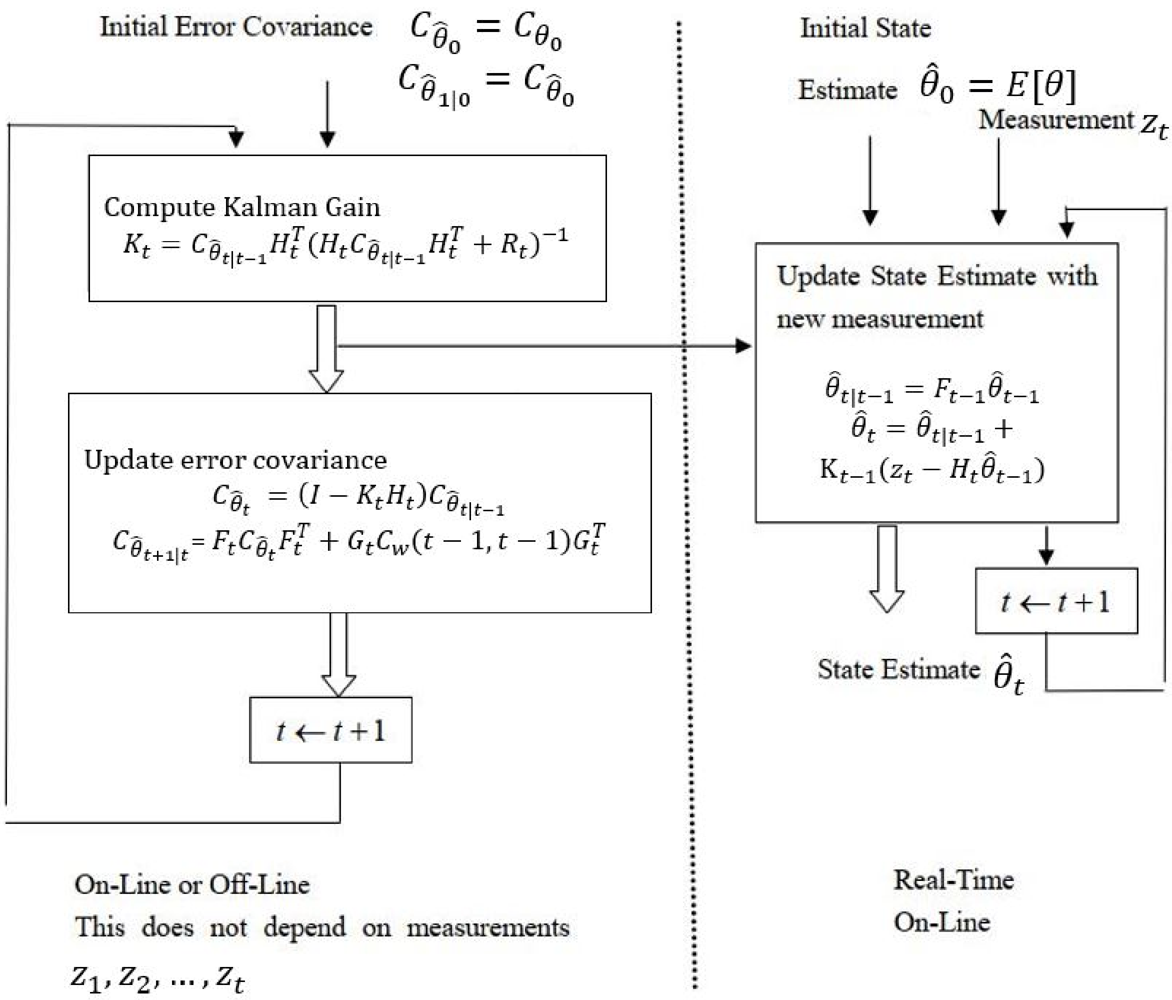


Figure: Solve discrete Kalman Filter recursively

2 Multiple choices

Please give a brief explanation for option(s) you choose. Answering without any description will get 0 point.

Problem 7 (5 points)

Which of the following statements are true regarding baseline correction? Assume that we are using **stimulus-locked epochs**.

(A) To perform baseline correction, the mean voltage is calculated during the prestimulus portion of the epoch, and this value is then subtracted from every point in the prestimulus period.

(B) To perform baseline correction, the mean voltage is calculated during the prestimulus portion of the epoch, and this value is then subtracted from every point in the waveform.

(C) We take the mean of the prestimulus period (rather than just taking the voltage at time zero) so that we can average out random noise during the prestimulus period and obtain a better estimate of the voltage offset.

(D) We take the mean of the prestimulus period (rather than just taking the voltage at time zero) so that we can obtain a better estimate of the noise level.

Answers are : B and C.
Before the stimulation occurs, the various internal and external source may cause temporal drifts. These drifts cause voltage offset. Therefore, Baseline correction is taking the mean of the recording baseline (-200ms to 0ms) to estimate the voltage offset, and then substrate the voltage offset from every point in the waveform.

Problem 8: Single choice (5 points)

Imagine that a researcher conducts a study comparing a patient group with a control group in an oddball paradigm. The researcher conducts a separate patient/control x rare/frequent ANOVA for each time point from 0-800 ms at each electrode site. Each ANOVA yielded 3 p values (main effect of patient/control, main effect of rare/frequent, and interaction). The sampling rate was 250 Hz, so there were 200 time points between 0 and 800 ms. There were 32 electrode sites. If there are no true differences between groups, how many significant p values would you expect the researcher to obtain as a result of noise in the data? [For simplicity, assume that every time point and electrode site is independent of every other time point and electrode site. Assume that $\alpha = 0.05$.]
(A) 1280 (B) 507 (C) 320 (D) 960
Answer: D.
For conducting ANOVA for each time point, and each ANOVA gives three p values (patient/control, rare/frequent, interaction), there will be $timepoints \times electrodes \times 3$ numbers of p values,which is $(250 \times 0.8) \times 32 \times 3 = 19200$. For 0.05 significant level, there are expected $19200 \times 0.05 = 960$ significant p-values from the data.

Problem 9 (5 points)

Imagine that a researcher conducts a go/no-go experiment in which subjects are supposed to press a button every time they see the word GO (written in green, 80% of trials) and to make no response when they see the word STOP (written in red, 20% of trials). And imagine that they find a larger N1 wave (at 170 ms) for the STOP stimulus than for the GO stimulus. Could this effect be plausibly explained by a physical stimulus confound?

(A) No. It is unlikely that the small physical differences between GO and STOP could explain a difference at 170 ms.

(B) Yes, because subjects may have been looking at a different places on the screen when the words GO and STOP appeared, which would change the position of the stimulus on the retina.

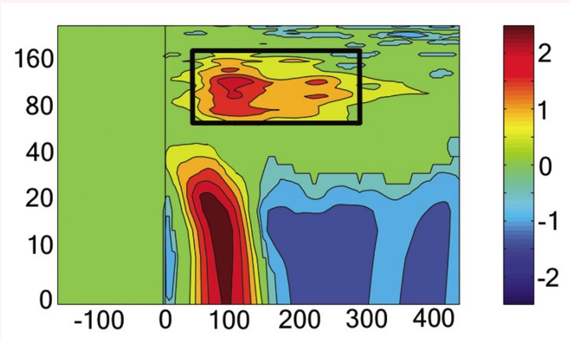
(C) Yes, because the word STOP contains more letters than the word GO and therefore might elicit a larger N1.

(D) Yes, because it is possible that red stimuli elicit a larger N1 than green stimuli.

Ans: B, C and D
From reference papers, color red seems to guide attention[7], while N1 would be affected when adapting to different visual target[8], and is related to subjects's spatial attention[9], we suppose that B,C and D are correct.

Problem 10 (5 points)

The time-frequency plot below shows data from the oddballs in a mismatch negativity paradigm. Which of the following statements are true about this plot?



(A) The X axis is time.

(B) The Y axis is a measure of magnitude (typically amplitude or power).

(C) If we wanted to reproduce the original data from a set of wavelets, we would need some high-amplitude wavelets centered at 100 ms with frequencies ranging from 0 to 20 Hz (in addition to wavelets at other frequencies).

(D) The activity shown in the box from 80-280 milliseconds is probably a genuine oscillation.

Answers: A and C
The x axis is time, y axis is frequency, contour lines are amplitude, hence A is correct, B is incorrect. From observing the plot, there are high amplitude activations of 0-20 frequency waves between 50-150 Hz, so C is correct. For option D, 80-120 Hz is too high to be EEG, the activation is probably from noise.

3 Coding problem

Problem 11: Auditory Oddball paradigm (2+2+2+2+2+2+2+2=16 points)

Please use the data: Day1 _ERP.set to answer these questions.
Data Information
This data contains 2 sessions, been down-sampled to $250Hz$, and been band-pass filtered.

Trigger	Event
10	Response
2	High pitch
3	Low pitch
4	End of session

(a) Please guess the range of band-pass filtered and show how you find this range.
(b) Please guess the portion of High Pitch: Low pitch and show how you find it.
(c) Plot the Fz, Cz, and Pz's average ERP for Response respectively.
(d) Plot the Fz, Cz, and Pz's average ERP for High pitch and Low pitch respectively.

For subproblem (e) and (f), please plot topolots for P300.
Suppose that for each channel, P300 occurs during $[300, 400]msec$ when $t = 0$ indicates onset time of High pitch and Low pitch. That is, $P300^{ch} \in \mathbb{R}^{51\times1}\ \forall ch \in \mathbb{Z}_{30}$.
(e) Plot topoplot for High pitch and Low pitch respectively. (Use $mean(P300^{ch})\ \forall ch \in \mathbb{Z}_{30}$)
$\forall ch \in \mathbb{Z}_{30}$, define Min-Max normalization as below

$$\frac{mean(P300^{ch}) - \min\left\{mean(P300^{ch})\middle|ch \in \mathbb{Z}_{30}\right\}}{\max\left\{mean(P300^{ch})\middle|ch \in \mathbb{Z}_{30}\right\} - \min\left\{mean(P300^{ch})\middle|ch \in \mathbb{Z}_{30}\right\}} \tag{3.1}$$

(f) Plot topoplot for High pitch and Low pitch respectively with Min-Max normalization.
If we define signal-to-noise ratio (SNR) for each channel as below:

$$SNR^{ch} = \frac{P300^{ch}}{std(Baseline^{ch})} \tag{3.2}$$

where baseline interval $[-200, 0]msec$, $Baseline^{ch}$ is mean by trial, and $t = 0$ indicates onset time of High pitch and Low pitch.
(g) For each channel, plot SNR for High picth and Low pitch. (Bar plot)
(h) Plot cumulative (by trial) SNR for Fz, Cz, and Pz channel and give a description of your observation.

a) From Figure 1 we can assume that filter that was applied was 32Hz low-pass. But if was applied band-pass filter from Figure 2 and 3 we can infer that probably filter bandpass range was 0.8-32 Hz.

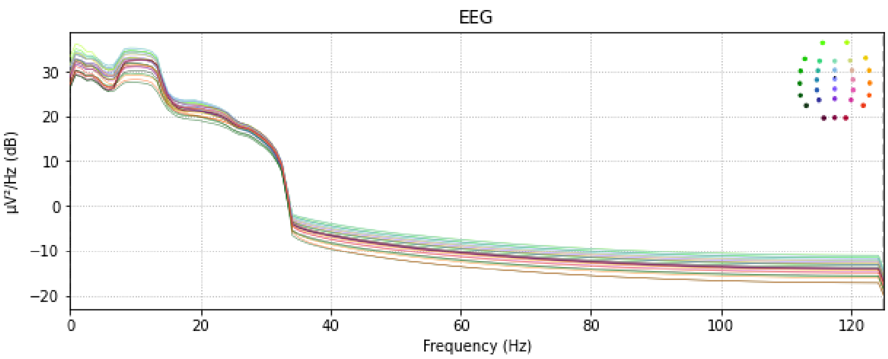


Figure 1: PSD plot of EEG signal

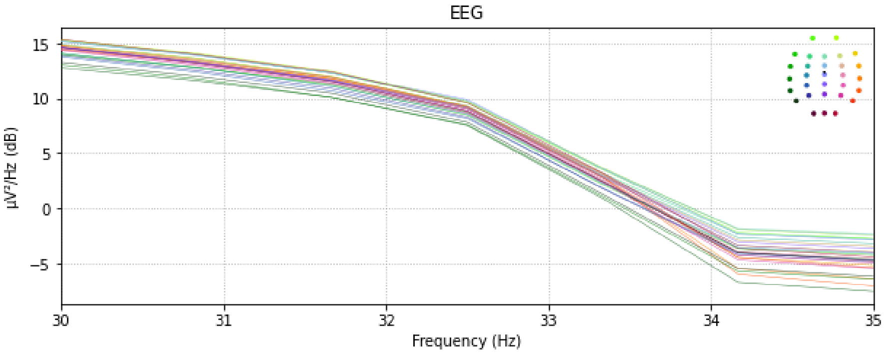


Figure 2: PSD plot of EEG signal, zoomed part [30 - 35

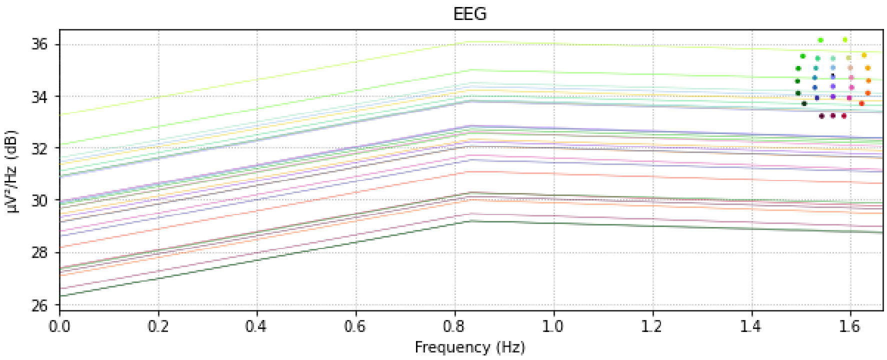


Figure 3: PSD plot of EEG signal, zoomed part [0 - 1.6]

b) For visualizing events distribution we used *viz.plot_events()* function that will generates Figure 4, and you just can count events. We also count events automatically based on internal mapping, you can find implementation in script file. What we found was 100 events for High pitch, and 400 events for Low pitch, that you can clearly see on the graph.

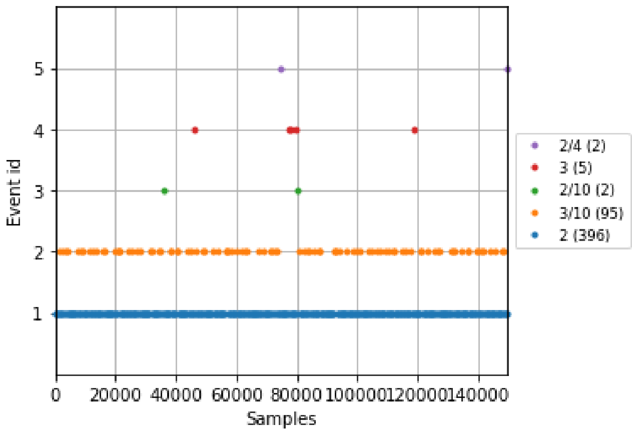


Figure 4: Distribution of events

c) Figure 5 is the summary for Fz, Cz and Pz's average ERP for Response (10). More figure you can find in the script for this part of the assignment.

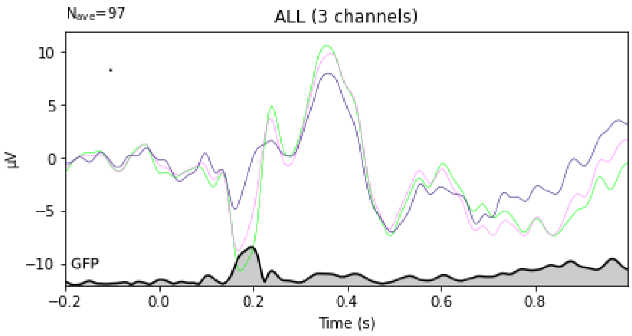


Figure 5: Average ERP for Response

d) Figures 6 and 7 is the summary for Fz, Cz and Pz's average ERP for Low and High pitch (3, 2). More figure you can find in the script for this part of the assignment.

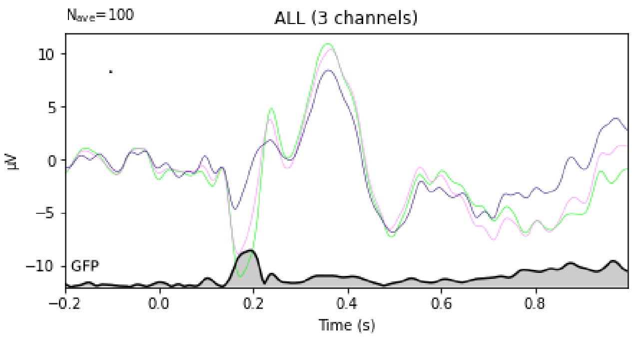


Figure 6: Average ERP for Low pitch

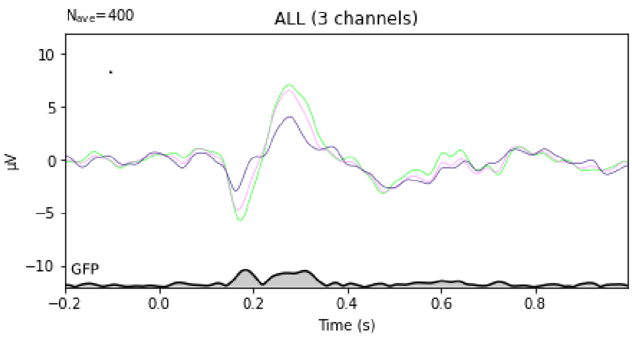


Figure 7: Average ERP for High pitch

e,f) In a Figure 8 you can see topoplots for Low pitch before and after Min-Max normalization. Same stands for the Figure 9 that shows topoplots before and after for a High pitch.

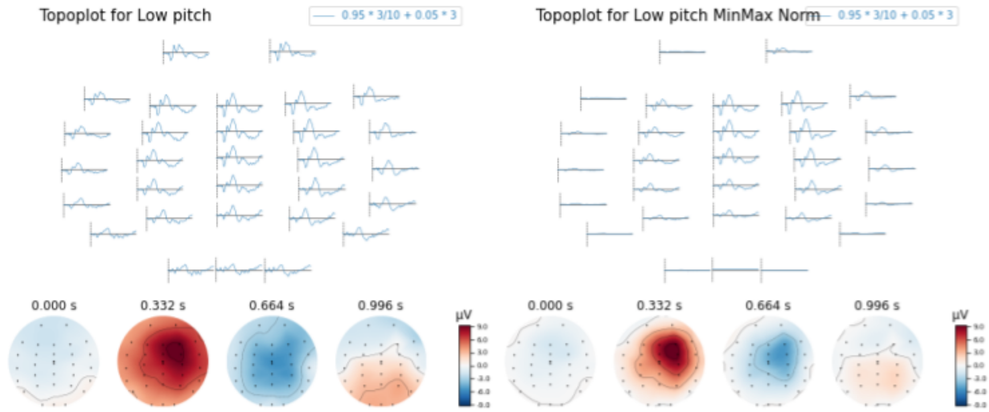


Figure 8: Topoplots for Low pitch before and after Min-Max normalization

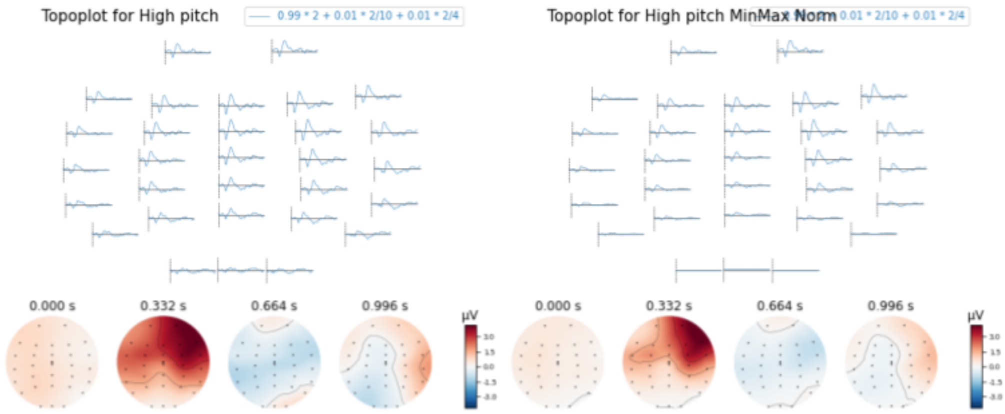


Figure 9: Topoplots for High pitch before and after normalization

g) Signal to Noise Ratio is higher when the signal is high and standard deviation of noise is low, and it's low when the signal is low and standard deviation of noise is high. From Figure 10 we can clearly see that for a low pitch we get high SNR values for F4, FT8, C4, CP4, P4 that corresponds to the right homosphere that correlates with our observations on the Figure 8. Highest SNR for High pitch, on Figure 2, lay between F4 and C3 that roughly corresponds to the frontal cortex, that also matches our observations form Figure 9.

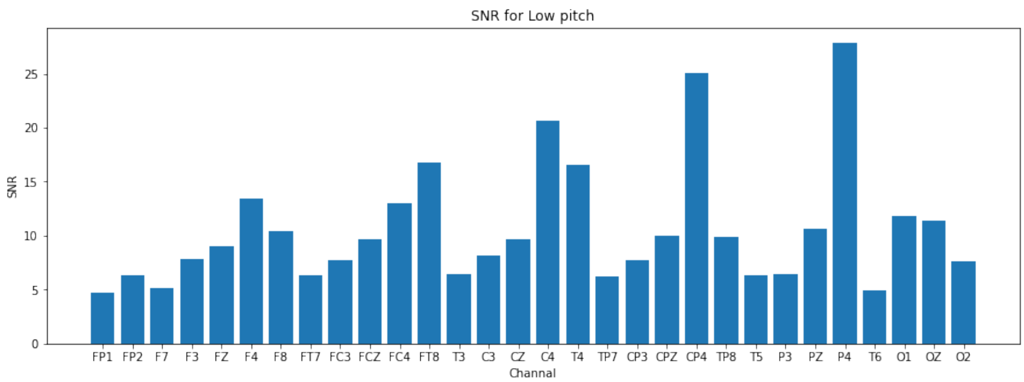


Figure 10: SNR for Low pitch

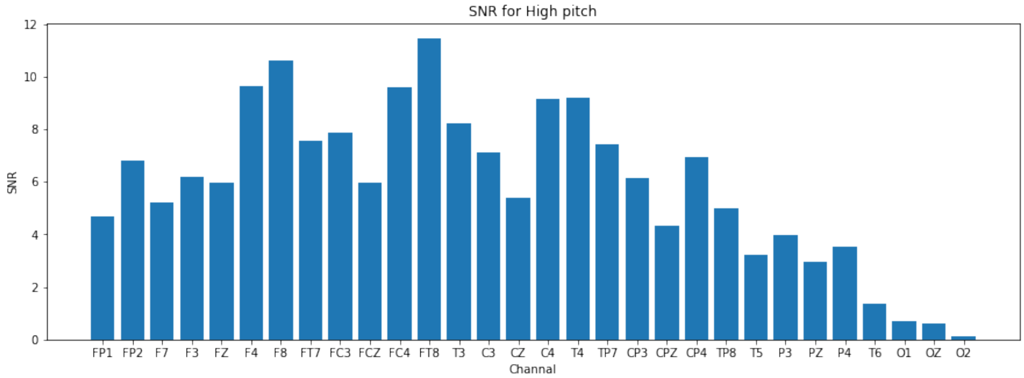


Figure 11: SNR for High pitch

h) Figure 12 shows cumulative (by trial) SNR for Fz, Cz, and Pz channels, you can clearly see grows over time with FZ that have a highest value, PZ with lowest value and CZ in the middle. From that we can conclude that the most amount of the valuable signal comes from FZ, that fit our previous observations.

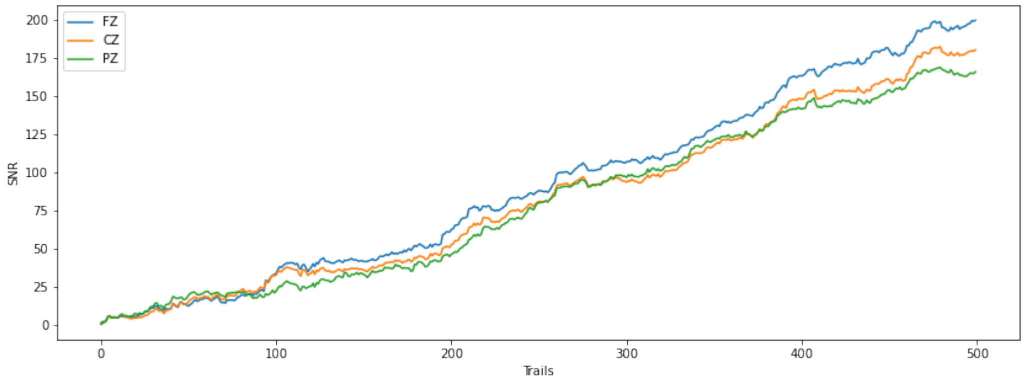


Figure 12

Problem 12: 5 target SSVEP paradigm (4+8+7=19 points)

Please use the data: Day 2_SSVEP.set to answer these questions.
Data Information

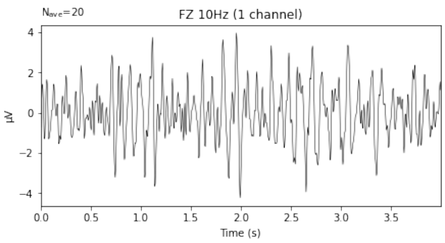
Trigger	Event
11	10 Hz
21	11 Hz
31	12 Hz
41	13 Hz
51	Nan

(a) For Fz and Oz, please plot average ERP for each type of stimuli.
Apply short time Fourier transform (spectrogram in matlab) with the following parameters to answer subproblem (b).

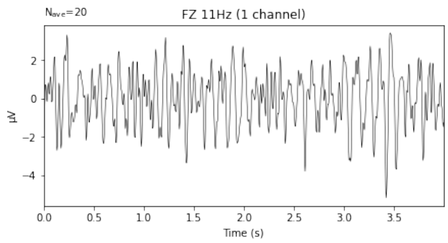
```
% B: SSVEP for certain channel, sfreq:sampling rate
% P is a power spectrum density matrix with size (N_freq, N_time)
[S,F,T,P]=spectrogram(B,sfreq,sfreq/2,sfreq,sfreq);
```

(b) Plot power v.s. frequency for each stimuli at F_z and O_z channel and give description of your observation.
We extract $\{10, 11, 12, 13\}Hz$ **from PSD you get from subproblem (b), and called it as response frequency.**
(c) Plot topolot of response Hz v.s. stimuli Hz with same Min-Max normalization technique. That is, you will plot $4 \times 5 = 20$ topoplots this time.

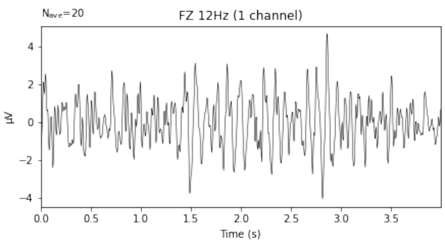
a)



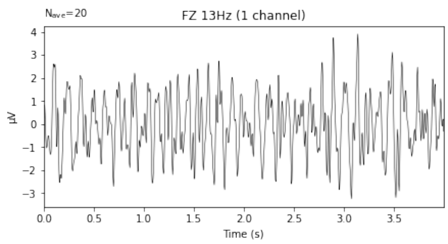
(a) average ERP for FZ channel at 10Hz stimuli



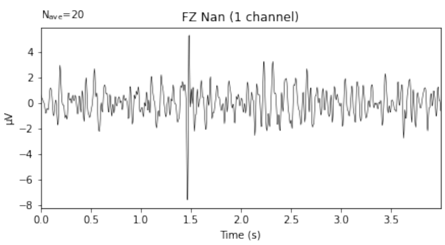
(b) average ERP for FZ channel at 11Hz stimuli



(c) average ERP for FZ channel at 12Hz stimuli

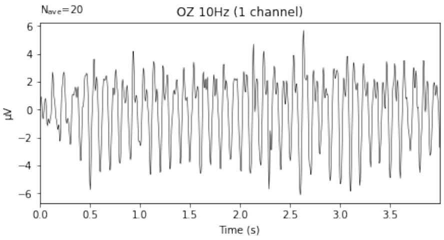


(d) average ERP for FZ channel at 13Hz stimuli

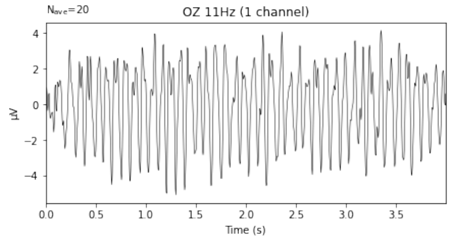


(e) average ERP for FZ channel at nan stimuli

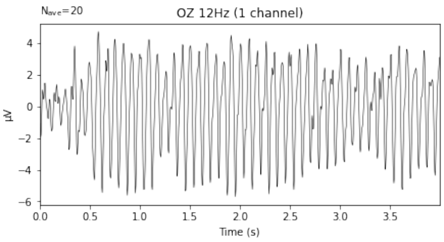
Figure 13: Average ERP for FZ channel at different frequencies



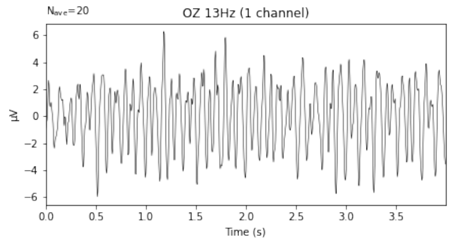
(a) average ERP for OZ channel at 10Hz stimuli



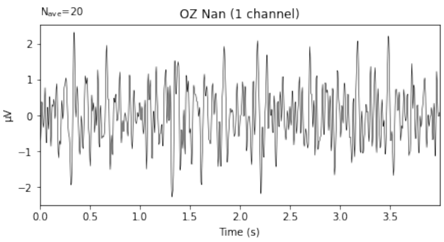
(b) average ERP for OZ channel at 11Hz stimuli



(c) average ERP for OZ channel at 12Hz stimuli



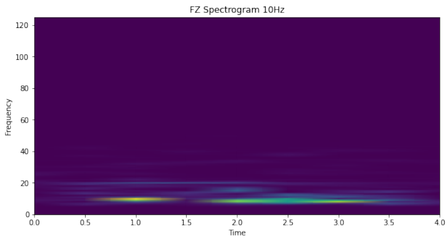
(d) average ERP for OZ channel at 13Hz stimuli



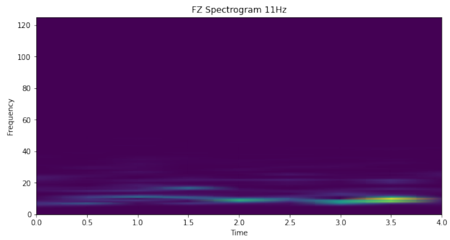
(e) average ERP for OZ channel at nan stimuli

Figure 14: Average ERP for Oz channel at different frequencies

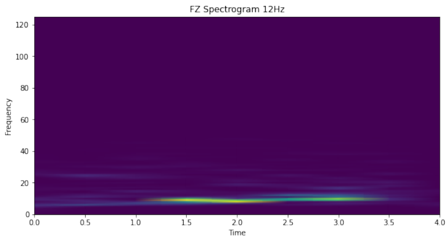
b)



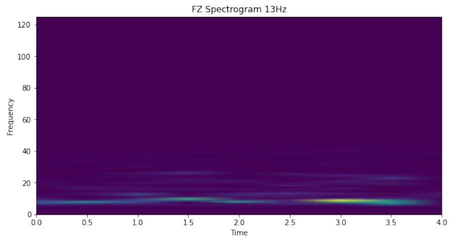
(a) spectrogram for 10Hz stimuli at Fz channel



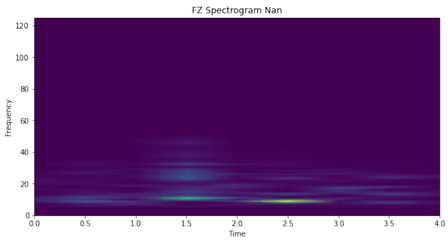
(b) spectrogram for 11Hz stimuli at Fz channel



(c) spectrogram for 12Hz stimuli at Fz channel



(d) Plot power v.s. frequency for 13Hz stimuli at Fz channel



(e) spectrogram for nan stimuli at Fz channel

Figure 15: Spectrogram for different frequencies at Fz channel

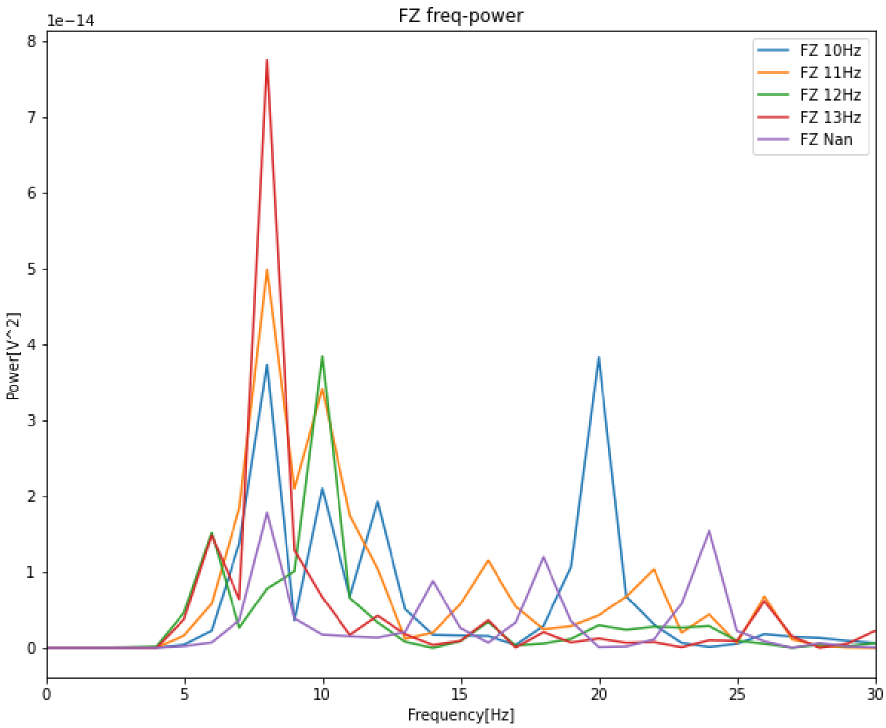
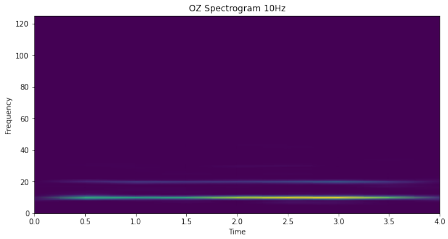
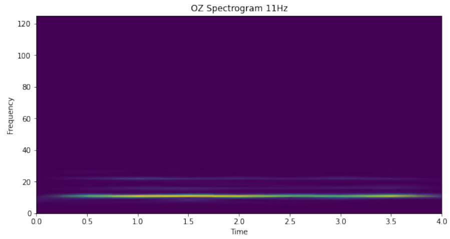


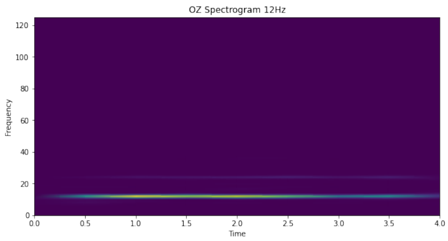
Figure 16: power v.s. frequency for every stimuli at Fz channel



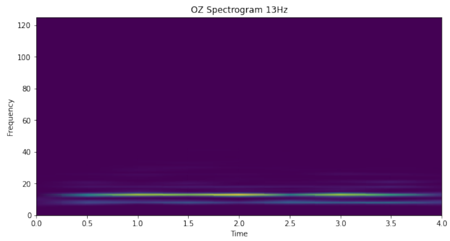
(a) spectrogram for 10Hz stimuli at Oz channel



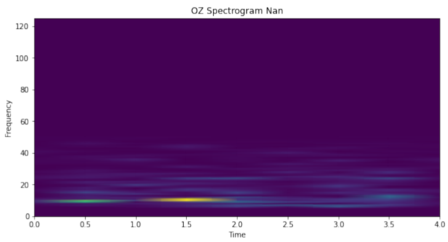
(b) spectrogram for 11Hz stimuli at Oz channel



(c) spectrogram for 12Hz stimuli at Oz channel



(d) spectrogram for 13Hz stimuli at Oz channel



(e) spectrogram for nan stimuli at Oz channel

Figure 17: Spectrogram for different frequencies at Oz channel

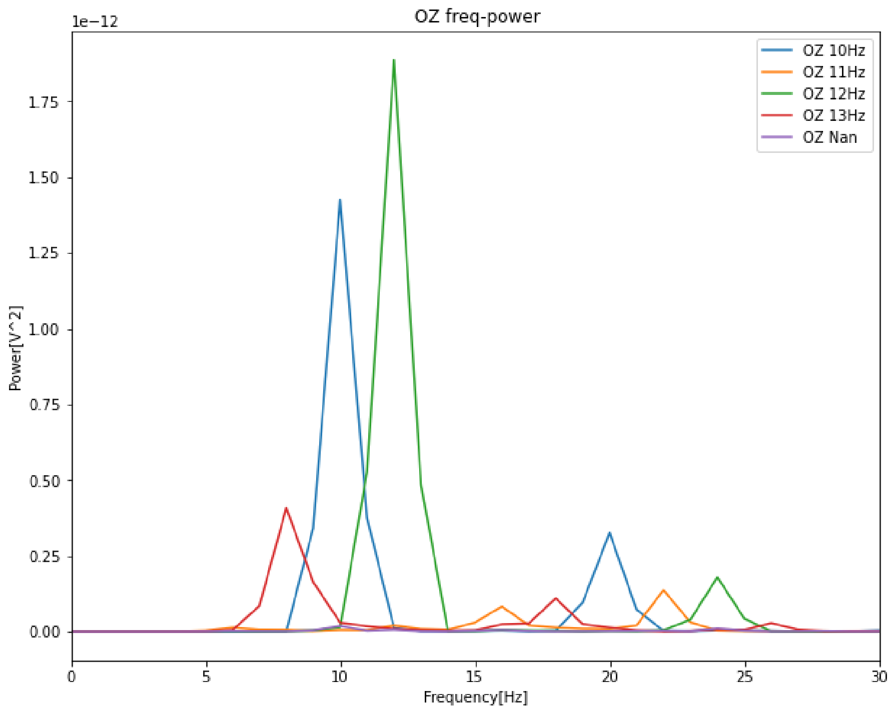


Figure 18: power v.s. frequency for every stimuli at OZ channel

Based on the plot of power v.s frequency of Fz channel, the lines, which correspond to different frequencies of stimulus, the value are quite random. The harmonic phenomenon is not obvious.

In contrast, the plot of power v.s frequency of Oz channel, the harmonic phenomenon is obvious. The 10 Hz stimulus have strong responses at 10Hz and 20 Hz. The 11Hz stimulus has a strong response at 22Hz, and the 12Hz stimulus have strong responses at 12Hz and 24Hz. However, the 13 Hz stimulus have strong responses at 9 Hz, 18 Hz, 27 Hz. It's suspicious that this stimulus is actually 9 Hz, but not 13Hz.

In conclusion, the harmonic phenomenon is more obvious at Oz. This indicates that the stimulus triggered more response at Occipital than Frontal region.

(c)

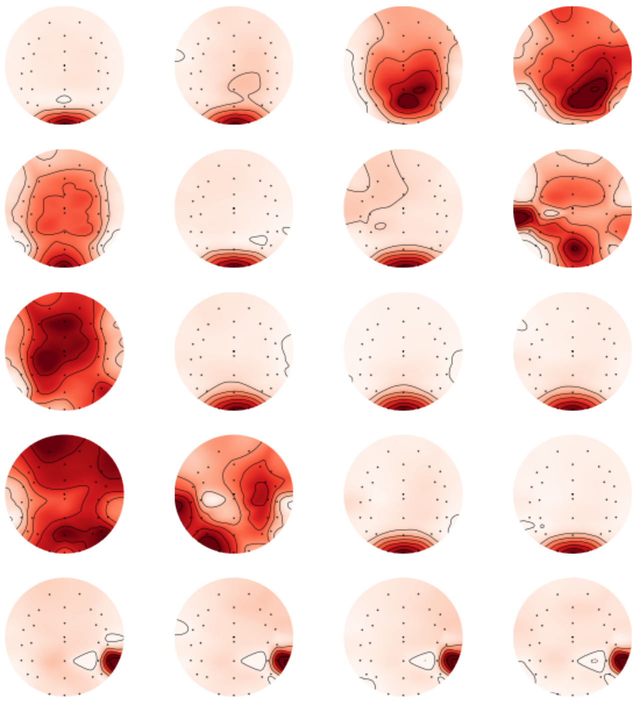


Figure 19: Topoplot of response Hz v.s. stimuli Hz (x: Response Hz, y: Stimuli Hz)

There's a clear response in diagonal, which means the stimuli Hz produces corresponding response Hz. However, we observe distinct psd at non-diagonal region, it can be predicted from frequency power plot. There are some noises.

References

[1] Stefanos D. Georgiadis, Perttu O. Ranta-aho, Mika P. Tarvainen, Pasi A. Karjalainen, *Single-Trial Dynamical Estimation of Event-Related Potentials: A Kalman Filter-Based Approach*, IEEE Transactions on Biomedical Engineering, 52(8), 2005.

[2] Harry Asada, Lecture notes for *Identification, Estimation, and Learning*, Massachusetts Institute of Technology, Department of Mechanical Engineering, 2006.

[3] S. Sanei, J.A. Chambers, *EEG Signal Processing*, Wiley, 2007.

[4] Yuan-Pin Lin, Lecture notes for 3^{rd} *EEG summer workshop*, National Sun Yat-sen University, Institute of Medical Science and Technology, 2020.

[5] Mike X Cohen. *Analyzing neural time series data : theory and practice.* Cambridge, Massachusetts :The MIT Press, 2014.

[6] Donald L. Schomer and Fernando H. Lopes da Silva, *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*,Lippincott William & Wilkins, 2011. **ISBN** 9780781789424.

[7] Michał Kuniecki, Joanna Pilarczyk, and Szymon Wichary, *The color red attracts attention in an emotional context. An ERP study*, 2015

[8] Jens-Max Hopf, Edward Vogel, Geoffrey Woodman, Hans-Jochen Heinze, Steven J Luck *Localizing visual discrimination processes in time and space*, 2002

[9] George R Mangun, *Neural mechanisms of visual selective attention*, 1995

4 Feedback for Lab 3

This part is not for grading but for understanding learning situation of each student. Please give us your feedback and comments.

4.1 Work Division

Student ID	Name	Be response for...
0810834	Yi-Ching Chiu	problem 1 and 2.
0713223	Jian-Xue Huang	problem 3 and 4.
309591027	Chang-Ling Tsai	problem 5,6 and 7.
0716307	Chia-Ying Hsieh	problem 8, 9, and 10.
0860838	Ivan Kakhaev	problem 11.
0617028	Ya-Lin Huang	problem 12.

4.2 Suggestions and Comments

4.2.1 For instructor

4.2.2 For teaching assistant(s)

Ok first of all task was not clear for normalization in task 11 and 12

4.2.2.a For Min-Jiun

4.2.2.b For Eric

Office Hour Information

We'll have limited time to teach EEGLab and MNE on our course; therefore, if you have any question about lab 2, feel free to make an appointment or come to ask me during my office hour.

Day	Time	Office
Tue.	12:20 p.m.-13:10 p.m.	EC120
Thur.	06:30 p.m.-09:30 p.m.	SC207

Note

Actually, my office hour on Thursdays is main for calculus consultation. If there are undergraduate students come to ask calculus problems, I need to teach them first and then to solve your problem during the rest of the office hour on Thursday nights.