

# SLEEP EDF

---

AI Accelerator Computing Lab

수면 데이터 분석

**Younghoon Na**

**Email : [nayounghoon0223@gmail.com](mailto:nayounghoon0223@gmail.com)**

**Github : [younghoonNa@github.com](https://github.com/younghoonNa)**

# Polysomnogram

---

PSG는 수면연구의 일종으로 **Polysomnogram** 혹은 **PSG**라고 불림

- EEG(뇌의 움직임)
- EOG(안구의 움직임)
- EMG(근육 혹은 골격은 움직임)
- ECG (심장 움직임)

이 이외에도 여러가지 신호를 조사함.

# Detrended Fluctuation Analysis

DFA : Detrended fluctuation analysis

- bio-signal analysis
- Robustness and capability recognize nonlinear time series
- Revaling the extent of long-range correlations in time series

이 논문의 특징

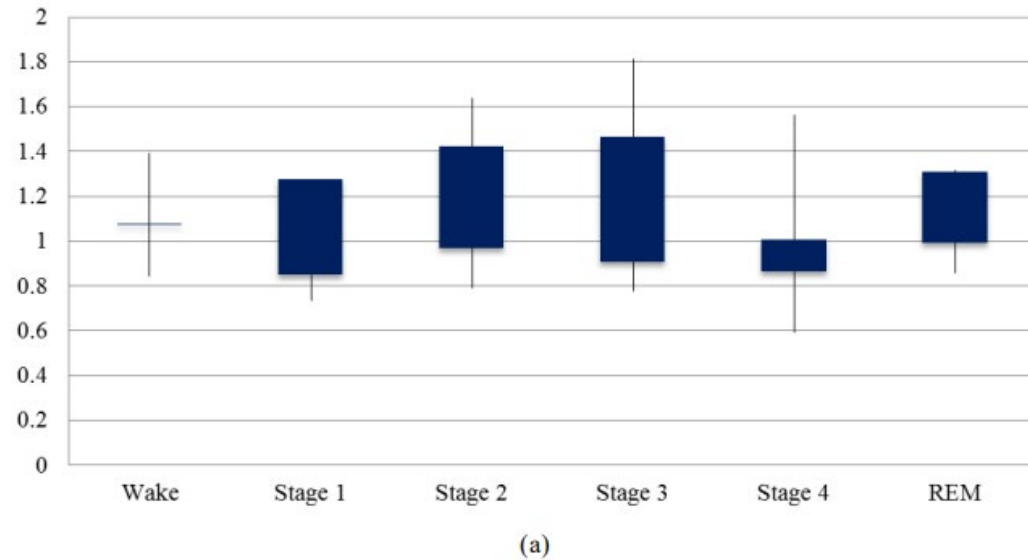
1. EEG Fpz-Cz, EEG Pz – Oz 만 사용하여 DFA 분석을 진행
2. EEG의 변동 정도를 나타내는  $\alpha$ 를 통해 DFA score를 계산
3.  $\alpha = 0.5$  는 백색소음.  $\alpha < 0.5$  일 경우 반지속적 상관관계, 반대는 지속적인 상관관계  
0.5 ~ 1 사이의 값은 지속적인 멱함수( $1/f$ ) 상관관계  
1 보다 클 경우 상관관계 존재 다만 멱함수 관계는 아님

$n$  = 샘플을  $n$ 개로 나눔

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_n(k)]^2}$$

$$F(n) \propto n^{\alpha}$$

# Detrended Fluctuation Analysis



EEG Fpz-Cz

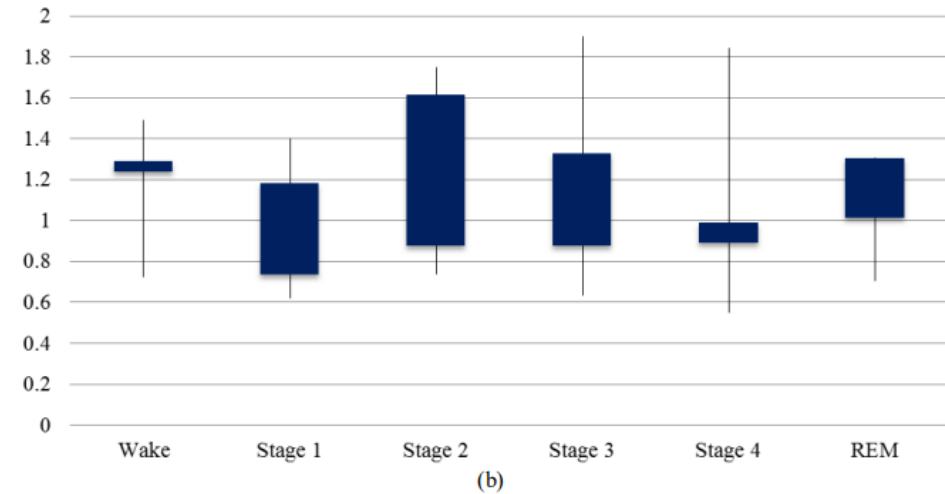


Figure 3. The mean and standard deviation of the scaling exponents in (a) Fpz-Cz channel and in (b) Pz-Oz

EEG Pz-Oz

1. 단계 및 신호마다 상이한  $\alpha$ 값 을 가지고 있음
2. 2단계에서 안정적인 신호를 4단계에서 불안정한 신호를 보여줌.

# Sleep Stage Classification

---

수면단계는 크게 Wake -> NREM -> REM 깊어진다.

NREM : non Rapid Eye Movement 의 줄임말

REM : Rapid Eye Movement

Wake(Wakefulness) 단계는  $100\mu V$  의 낮은 전압과 혼합된 EEG 신호가 발생된다.

NREM은 크게 1단계부터 4단계로 나눌 수 있다.

NREM 1단계 – drowsiness (졸음)

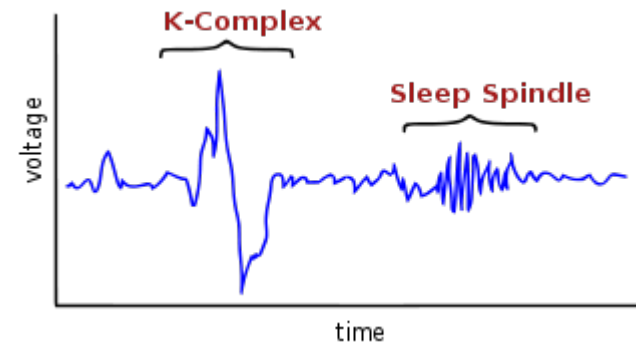
# Sleep Stage Classification -NREM

NREM 1단계 – drowsiness (졸음)

- 낮은 전압 ( $200\mu V$ )
- 혼합된 2-7Hz의 높은 진폭을 가진 EEG 신호
- 눈의 움직임은 매우 느림

NREM 2단계 – light sleep stage (얕은 수면 단계)

- 눈의 움직임 멈춤
- 뇌파 또한 매우 느려짐
- Sleep spindle 과 K complexes 나타남



# Sleep Stage Classification -NREM

---

NREM 3단계 – deep sleep (깊은 수면 단계)

- 매우 느린 뇌파 (delta wave라 불림)
- 높은 진폭의 1-3Hz EEG 신호 발생

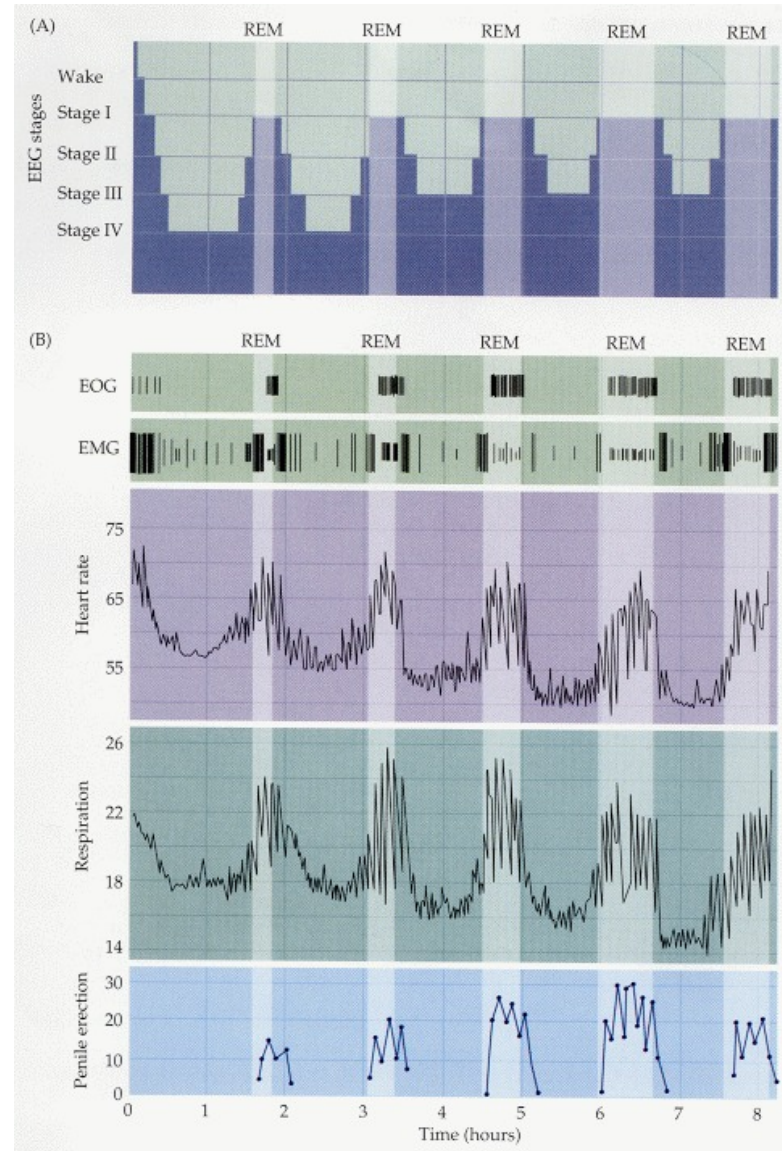
NREM 4단계 – deep sleep & slow wave sleep

- delta wave가 뇌에서 나타남.
- 높은 진폭 2Hz 이하의 EEG 신호 발생
- 3, 4 단계에서는 잠에 깨기가 매우 힘들

# Sleep Stage Classification - REM

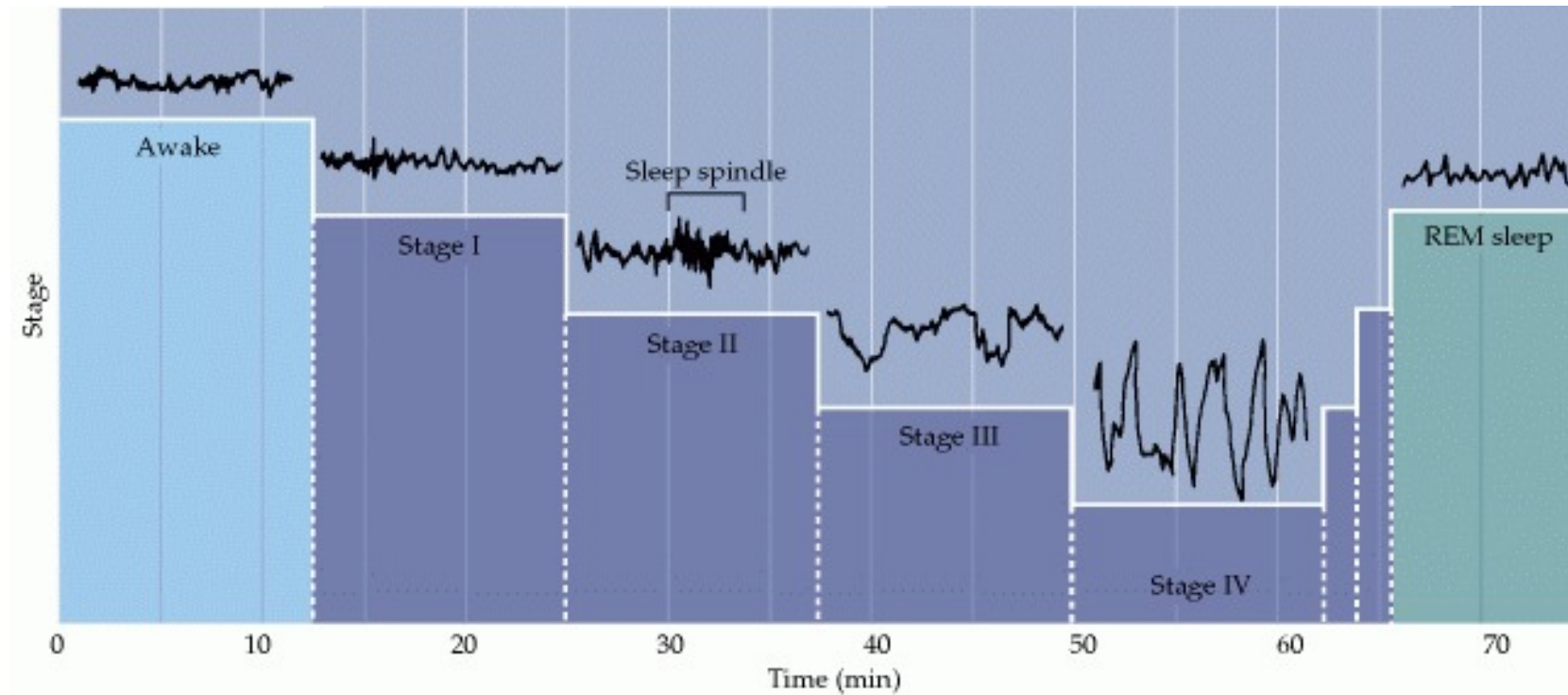
## REM 수면단계

- 5단계 라고도 불림
- Rapid Eye Movement
- 가끔 근육이 움직이는 현상 발생
- Theta wave 발생

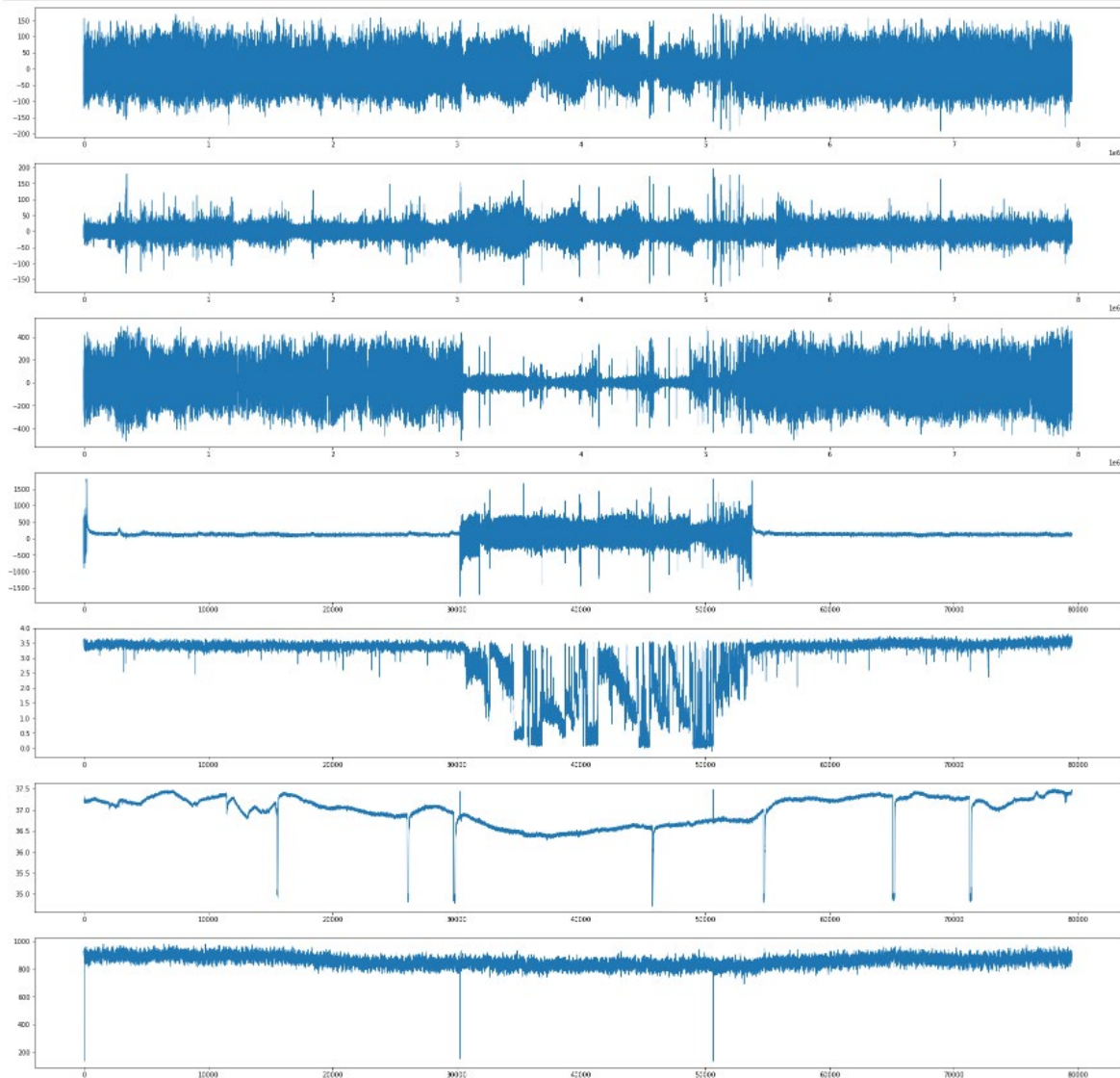




# Sleep EDF



# Sleep EDF – Data Preprocessing



```
for idx in range(7):  
    print(sleep_data1_label[idx]['label'])
```

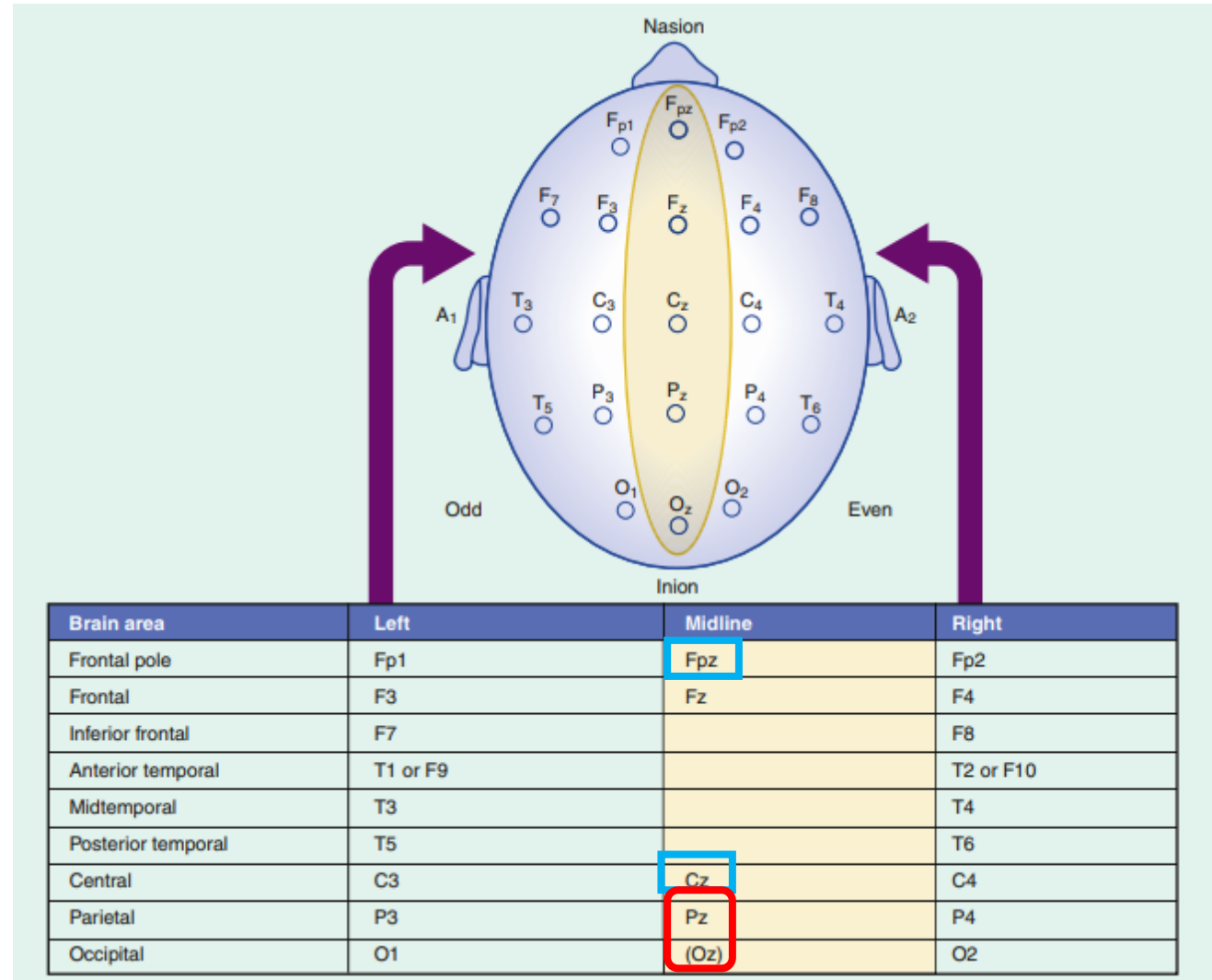
EEG Fpz-Cz  
EEG Pz-Oz  
EOG horizontal  
Resp oro-nasal  
EMG submental  
Temp rectal  
Event marker

# Sleep EDF – Signal

```
for idx in range(7):
    print(sleep_data1_label[idx]['label'])
```

EEG Fpz-Cz  
 EEG Pz-Oz  
 EOG horizontal  
 Resp oro-nasal  
 EMG submental  
 Temp rectal  
 Event marker

- EOG : 안구의 움직임



# Sleep EDF – Data Preprocessing

## 100Hz 데이터

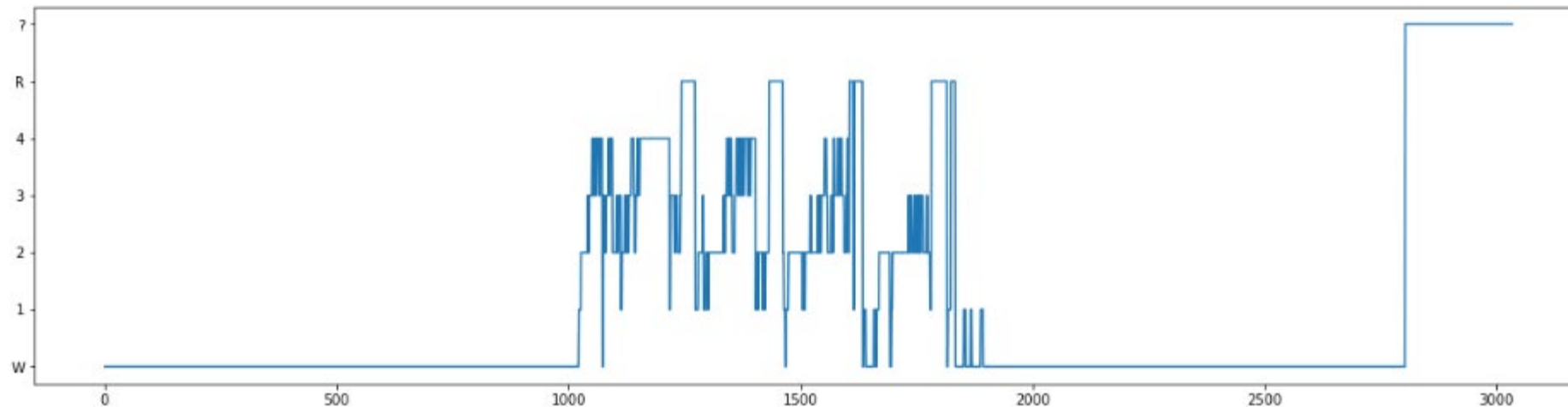
```
for idx in range(7):  
    print(sleep_data1_label[idx]['label'])
```

EEG Fpz-Cz  
EEG Pz-Oz  
EOG horizontal  
Resp oro-nasal  
EMG submental  
Temp rectal  
Event marker

Sample rate : 3000

Sample rate : 30

```
] : plt.figure(figsize = (20,5))  
    plt.plot(sleep_stage)  
    plt.show()
```



# Sleep EDF – Data Preprocessing

## ? 라벨 처리하기

```
sample = 0
sleep_data_signal = highlevel.read_edf(trains[sample])[0]
sleep_data_label = highlevel.read_edf(trains[sample])[1]
sleep_data1_test = highlevel.read_edf(labels[sample])

# print(sleep_data1)

for signal, info in zip(sleep_data_signal, sleep_data_label):
    print(signal.shape, info['label'], info['sample_rate'], signal.shape[0]//info['sample_rate'])

(7950000,) EEG Fpz-Cz 3000.0 2650.0
(7950000,) EEG Pz-Oz 3000.0 2650.0
(7950000,) EOG horizontal 3000.0 2650.0
(79500,) Resp oro-nasal 30.0 2650.0
(79500,) EMG submental 30.0 2650.0
(79500,) Temp rectal 30.0 2650.0
(79500,) Event marker 30.0 2650.0
```

```
0 [79500.0, 6900.0, 'Sleep stage ?'] ??? check 7950000
1 [84900.0, 1500.0, 'Sleep stage ?'] ??? check 8490000
2 [84060.0, 2340.0, 'Sleep stage ?'] ??? check 8406000
```

# Sleep EDF – Data Preprocessing

## ? 라벨 처리하기

```
tmp = 0
# Max_num = 152+1

for idx in range(Max_num): # 총 153개의 데이터
    d1 = highlevel.read_edf(trains[idx])[0] # 순서대로 읽기 (PSG.edf)
    d2 = highlevel.read_edf(labels[idx])    # 라벨 데이터 (Hypnogram.edf)

    for element in d2[-1]['annotations']: # 라벨 데이터 순회 [start, step, sleep stage] 로 구분
        if element[0] == 0:
            tmp = element[1]
            continue

        if element[0] != tmp:
            print("Exception")
            break

        if element[2][-1] == "?":
            print(idx, element, "???? check", d1[0].shape[0])

        tmp = element[0] + element[1] # tmp값은 이전 값의 start+step값

    if int(element[0] * 100) == d1[0].shape[0]: # 100Hz 데이터 이기 때문에 100을 곱해줄
        # print("complete", int(element[0] * 100), d1[0].shape[0]) # 너무 많아서 출력 x
        pass
    else:
        print(idx, " data", "dimension Error", int(element[0] * 100), d1[0].shape[0]) # 100Hz 데이터 이기 때문에

print("===== Process End =====")
```

```
0 [79500.0, 6900.0, 'Sleep stage ?'] ???? check 7950000
1 [84900.0, 1500.0, 'Sleep stage ?'] ???? check 8490000
2 [84060.0, 2340.0, 'Sleep stage ?'] ???? check 8406000
3 [85440.0, 960.0, 'Sleep stage ?'] ???? check 8550000
3 data dimension Error 8544000 8550000
4 [84120.0, 2280.0, 'Sleep stage ?'] ???? check 8412000
5 [82680.0, 3720.0, 'Sleep stage ?'] ???? check 8268000
6 [84600.0, 1800.0, 'Sleep stage ?'] ???? check 8460000
7 [81960.0, 4440.0, 'Sleep stage ?'] ???? check 8196000
8 [77100.0, 9300.0, 'Sleep stage ?'] ???? check 7710000
9 [83760.0, 2640.0, 'Sleep stage ?'] ???? check 8376000
10 [81660.0, 4740.0, 'Sleep stage ?'] ???? check 8166000
11 [84180.0, 2220.0, 'Sleep stage ?'] ???? check 8418000
12 [83100.0, 3300.0, 'Sleep stage ?'] ???? check 8310000
13 [84900.0, 1500.0, 'Sleep stage ?'] ???? check 8490000
14 [84300.0, 2100.0, 'Sleep stage ?'] ???? check 8430000
15 [83100.0, 3300.0, 'Sleep stage ?'] ???? check 8310000
16 [83880.0, 2520.0, 'Sleep stage ?'] ???? check 8388000
17 [79020.0, 7380.0, 'Sleep stage ?'] ???? check 7902000
18 [81960.0, 4440.0, 'Sleep stage ?'] ???? check 8196000
```

1. 대체로 ? 라벨은 마지막에 등장
2. ? 라벨을 제거 한 크기는 대체로 PSG 신호 데이터의 Shape과 동일
3. ? 값은 진짜 모르는 값이거나 쓰레기 값으로 넣었거나 끝을 알리기 위해서 쓰지 않았을까?



# Sleep EDF – Data Preprocessing

1. ? 라벨 값이 중간에도 등장하는 데이터
2. 2. ? 라벨을 제거한 라벨 데이터의 크기가 PSG 데이터의 shape과 다른 데이터

```
3 [85440.0, 960.0, 'Sleep stage ?'] ??? check 8550000
3 data dimension Error 8544000 8550000
```

```
19 [61650.0, 24060.0, 'Sleep stage ?'] ??? check 8574000
19 [85740.0, 660.0, 'Sleep stage ?'] ??? check 8574000
```

```
24 [80550.0, 30.0, 'Sleep stage ?'] ??? check 8358000
24 [80580.0, 30.0, 'Sleep stage ?'] ??? check 8358000
24 [80610.0, 180.0, 'Sleep stage ?'] ??? check 8358000
24 [80790.0, 5610.0, 'Sleep stage ?'] ??? check 8358000
24 data dimension Error 8079000 8358000
```

```
71 [83100.0, 3300.0, 'Sleep stage ?'] ??? check 6798000
71 data dimension Error 8310000 6798000
```

```
79 [82950.0, 3450.0, 'Sleep stage ?'] ??? check 8574000
79 data dimension Error 8295000 8574000
```

```
82 [81690.0, 4710.0, 'Sleep stage ?'] ??? check 8430000
82 data dimension Error 8169000 8430000
```

```
91 [83700.0, 2700.0, 'Sleep stage ?'] ??? check 8370000
92 data dimension Error 8598000 8640000
93 data dimension Error 8280000 8640000
```

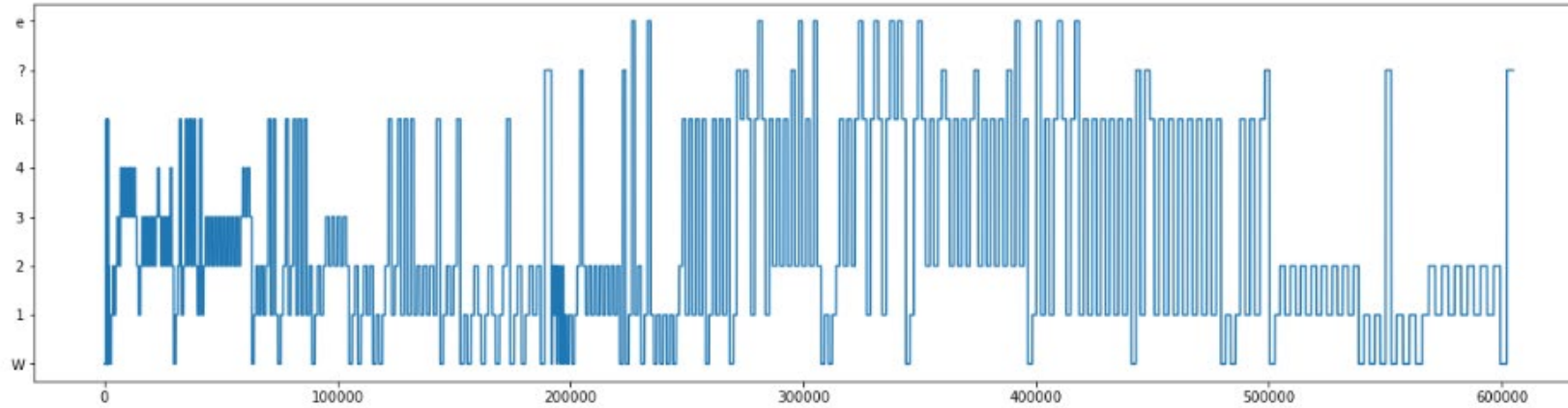
```
109 [62610.0, 420.0, 'Sleep stage ?'] ??? check 8550000
109 [85500.0, 900.0, 'Sleep stage ?'] ??? check 8550000
110 data dimension Error 6000000 8640000
```

```
117 [47760.0, 60.0, 'Sleep stage ?'] ??? check 7950000
117 [48150.0, 90.0, 'Sleep stage ?'] ??? check 7950000
117 [53700.0, 30.0, 'Sleep stage ?'] ??? check 7950000
117 [79500.0, 6900.0, 'Sleep stage ?'] ??? check 7950000
```

```
126 [49830.0, 30.0, 'Sleep stage ?'] ??? check 8520000
126 [85200.0, 1200.0, 'Sleep stage ?'] ??? check 8520000
127 [79920.0, 2280.0, 'Sleep stage ?'] ??? check 8220000
127 [82200.0, 4200.0, 'Sleep stage ?'] ??? check 8220000
```

```
143 [41880.0, 480.0, 'Sleep stage ?'] ??? check 7890000
143 [42420.0, 120.0, 'Sleep stage ?'] ??? check 7890000
143 [78900.0, 7500.0, 'Sleep stage ?'] ??? check 7890000
144 [36210.0, 30.0, 'Sleep stage ?'] ??? check 8430000
144 [37950.0, 90.0, 'Sleep stage ?'] ??? check 8430000
144 [45690.0, 30.0, 'Sleep stage ?'] ??? check 8430000
144 [45780.0, 30.0, 'Sleep stage ?'] ??? check 8430000
144 [47610.0, 30.0, 'Sleep stage ?'] ??? check 8430000
144 [51750.0, 30.0, 'Sleep stage ?'] ??? check 8430000
144 [52890.0, 30.0, 'Sleep stage ?'] ??? check 8430000
144 [53400.0, 30.0, 'Sleep stage ?'] ??? check 8430000
144 [57180.0, 2910.0, 'Sleep stage ?'] ??? check 8430000
144 [60210.0, 450.0, 'Sleep stage ?'] ??? check 8430000
144 [64320.0, 300.0, 'Sleep stage ?'] ??? check 8430000
144 [70050.0, 30.0, 'Sleep stage ?'] ??? check 8430000
144 [84300.0, 2100.0, 'Sleep stage ?'] ??? check 8430000
```

# Sleep EDF – Data Preprocessing



Movement time과 ? 는 앞에 나온 라벨로 맵핑

```
[46500.0, 60.0, 'Sleep stage R'],
[46560.0, 30.0, 'Movement time'],
[46590.0, 240.0, 'Sleep stage R'],
```

```
[47640.0, 180.0, 'Sleep stage 2'],
[47820.0, 30.0, 'Movement time'],
[47850.0, 30.0, 'Sleep stage 2'],
```

1. Movement time은 앞 뒤의 라벨이 같았기 때문
2. ?가 뒤에 나온 경우는 검사가 끝나고 주렁주렁 달린 장치를 제거하는 과정에서 찍힌 데이터지 않을까 생각.
3. ?가 중간에 나온 경우에는 두 개 이상의 신호가 생긴 애매한 상황에서의 라벨이라고 생각.

-> 따라서 그냥 앞의 라벨을 그대로 붙여 넣는 식으로 보간 진행해야겠다 생각.