Speech Signal Processing Project 1 Report

Due date: 05/04

Professor: 장길진 교수님

202222226464 전인구

아래의 링크에서 코드를 참조해 변형했습니다.

https://github.com/ln99/ssp2023/blob/main/proj1/proj1_nidr.ipynb

보고서의 연속성을 위해 github의 코드와 보고서의 구성 순서가 다릅니다.

1.HMM의 Training & Validation

튜플 자료형으로 (wave file경로, 숫자)로 Data를 정리했습니다.

Ex)

```
('segmented-train/Dandyst/0/kdigits0-0.wav', 0)
('segmented-val/org/do/0/kdigits0-0.wav', 0)
```

train_digits와 validation_digits도 튜플 자료형을 사용하도록 변형했습니다.

Ex)

```
def train_digits_modified(Data_list, tag, num_trials=10):
##########################################################################
# Data loading from Data_list

    m_trainingsetfeatures = [extmfcc(x) for (x,y) in Data_list] # Data path
    m_trainingsetlabels = [y for (x,y) in Data_list] # Data label
...
```

코드 1 train_digits_modified 함수의 일부분

숫자 별로 성능을 확인하고 싶어서 Confusion matrix를 같이 출력하도록 validation함수를 수정했습니다.
그 후, 반복해서 코드를 돌려보면서, HMM 모델을 구성하기 위한 변수들을 아래와 같이 바꿔줬습니다.

변수 명	변형 전	변형 후
n_mfcc	10	20
m_num_of_HMMStates	3	4
m_num_of_mixtures	2	3

기존의 MFCC가 충분한 특징을 포함하지 않는 것 같아서 n_mfcc를 늘렸고, 기존의 HMM 모델보다 상태와 전이를 더 많이 할 수 있게 해서 복잡한 종속성을 학습하도록 했습니다.

Training 결과

```
Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 500, 500
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =85.8


Confusion Matrix (row : True, column : Predicted)
label:0        [37.  0.  0.  0.  2.  0.  0.  0. 11.  0.]          accuracy : 74 %
label:1        [ 0. 44.  0.  0.  0.  0.  6.  0.  0.  0.]          accuracy : 88 %
label:2        [ 0.  0. 41.  0.  0.  0.  0.  1.  8.  0.]          accuracy : 82 %
label:3        [ 0.  0.  0. 39.  5.  0.  3.  0.  3.  0.]          accuracy : 78 %
label:4        [ 0.  0.  0.  2. 45.  0.  0.  0.  3.  0.]          accuracy : 90 %
label:5        [ 0.  1.  0.  0.  0. 39. 10.  0.  0.  0.]          accuracy : 78 %
label:6        [ 0.  0.  0.  0.  0.  0. 39.  0.  2.  0.]          accuracy : 78 %
label:7        [ 0.  0.  0.  0.  0.  0.  2. 46.  2.  0.]          accuracy : 92 %
label:8        [ 0.  0.  0.  0.  0.  0.  0.  0. 50.  0.]          accuracy : 100 %
label:9        [ 0.  0.  0.  0.  0.  0.  1.  0.  0. 49.]          accuracy : 98 %

================================================
```

Validation 결과

```
Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 200, 200
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =51.5


Confusion Matrix (row : True, column : Predicted)
label:0        [ 0.  0.  0.  0.  0.  0.  7.  0. 13.  0.]          accuracy : 0 %
label:1        [ 0. 19.  0.  0.  0.  0.  0.  0.  0.  1.]          accuracy : 95 %
label:2        [0. 8. 9. 0. 0. 0. 0. 1. 2. 0.]                    accuracy : 45 %
label:3        [0. 0. 1. 9. 2. 0. 2. 0. 6. 0.]                    accuracy : 45 %
label:4        [ 0.  0.  0.  2. 10.  0.  1.  2.  5.  0.]          accuracy : 50 %
label:5        [0. 6. 0. 0. 0. 5. 4. 0. 0. 5.]                    accuracy : 25 %
label:6        [ 0.  1.  1.  0.  0.  2. 10.  0.  4.  2.]          accuracy : 50 %
label:7        [ 0.  0.  0.  0.  0.  0.  0. 20.  0.  0.]          accuracy : 100 %
label:8        [ 0.  0.  2.  0.  0.  0. 12.  0.  6.  0.]          accuracy : 30 %
label:9        [ 0.  3.  0.  0.  0.  0.  0.  2.  0. 15.]          accuracy : 75 %

================================================
```

결과 1 Traing 과 validation 데이터의 예측 결과

노이즈가 있는 validation 데이터에 대해서는 SNR이 낮아질수록 낮은 성능을 보였습니다.

[nbnSNR10 데이터의 예측 결과]

```
testing segmented-val/nbnSNR10
Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 200, 200
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =49.0


Confusion Matrix (row : True, column : Predicted)
label:0        [ 0.  0.  2.  0.  0.  0. 12.  2.  4.  0.]              accuracy : 0 %
label:1        [ 0. 13.  0.  0.  0.  1.  0.  0.  0.  6.]              accuracy : 65 %
label:2        [ 0.  0. 11.  1.  0.  0.  0.  7.  0.  1.]              accuracy : 55 %
label:3        [ 0.  0.  0.  5.  2.  0. 13.  0.  0.  0.]              accuracy : 25 %
label:4        [0. 0. 0. 4. 9. 0. 3. 4. 0. 0.]              accuracy : 45 %
label:5        [0. 3. 0. 0. 0. 1. 7. 1. 0. 8.]              accuracy : 5 %
label:6        [ 0.  0.  1.  0.  0.  0. 18.  0.  0.  1.]              accuracy : 90 %
label:7        [ 0.  0.  0.  0.  0.  0.  0. 20.  0.  0.]              accuracy : 100 %
label:8        [ 0.  0.  2.  0.  0.  0. 17.  0.  1.  0.]              accuracy : 5 %
label:9        [ 0.  0.  0.  0.  0.  0.  0.  0.  0. 20.]              accuracy : 100 %

==================================================
```

[wbnSNR10 데이터의 예측 결과]

```
testing segmented-val/wbnSNR10
Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 200, 200
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =28.0


Confusion Matrix (row : True, column : Predicted)
label:0        [ 0.  0.  0.  0.  0.  0. 10.  0. 10.  0.]              accuracy : 0 %
label:1        [0. 6. 0. 0. 0. 1. 7. 3. 3. 0.]              accuracy : 30 %
label:2        [ 0.  0. 10.  0.  0.  0.  0.  4.  6.  0.]              accuracy : 50 %
label:3        [ 0.  0.  0.  7.  0.  0.  0.  0. 13.  0.]              accuracy : 35 %
label:4        [ 0.  0.  0.  3.  0.  0.  0.  0. 17.  0.]              accuracy : 0 %
label:5        [ 0.  0.  0.  0.  0.  0. 11.  4.  5.  0.]              accuracy : 0 %
label:6        [ 0.  0.  0.  0.  0.  0. 10.  1.  9.  0.]              accuracy : 50 %
label:7        [ 0.  0.  0.  0.  0.  0.  0. 13.  7.  0.]              accuracy : 65 %
label:8        [ 0.  0.  0.  0.  0.  0. 10.  0. 10.  0.]              accuracy : 50 %
label:9        [ 0.  0.  0.  0.  0.  0. 10.  8.  2.  0.]              accuracy : 0 %

==================================================
```

[nbnSNR0 데이터의 예측 결과]

```
testing segmented-val/nbnSNR0
Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 200, 200
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =35.0


Confusion Matrix (row : True, column : Predicted)
label:0        [ 0.  0.  4.  1.  0.  0. 15.  0.  0.  0.]              accuracy : 0 %
label:1        [0. 8. 0. 0. 0. 5. 1. 0. 0. 6.]              accuracy : 40 %
label:2        [ 0.  0. 16.  4.  0.  0.  0.  0.  0.  0.]              accuracy : 80 %
label:3        [ 0.  0.  0. 11.  0.  0.  8.  1.  0.  0.]              accuracy : 55 %
label:4        [ 0.  0.  1. 12.  0.  0.  6.  1.  0.  0.]              accuracy : 0 %
label:5        [ 0.  0.  0.  0.  0.  0. 14.  0.  0.  6.]              accuracy : 0 %
label:6        [ 0.  0.  0.  0.  0.  0. 19.  0.  0.  1.]              accuracy : 95 %
label:7        [ 0.  0.  2.  4.  0.  0. 10.  4.  0.  0.]              accuracy : 20 %
label:8        [ 0.  0.  0.  1.  0.  0. 19.  0.  0.  0.]              accuracy : 0 %
label:9        [ 0.  0.  0.  0.  0.  0.  8.  0.  0. 12.]              accuracy : 60 %

==================================================
```

[wbnSNR0 데이터의 예측 결과]

```
testing segmented-val/wbnSNR0
Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 200, 200
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =13.5


Confusion Matrix (row : True, column : Predicted)
label:0        [ 0.  0.  0.  0.  0.  0.  9.  0. 11.  0.]              accuracy : 0 %
label:1        [ 0.  0.  1.  0.  0.  8.  1.  0. 10.  0.]              accuracy : 0 %
label:2        [ 0.  0.  0.  0.  0.  0.  0.  0. 20.  0.]              accuracy : 0 %
label:3        [ 0.  0.  0.  2.  0.  0.  0.  0. 18.  0.]              accuracy : 10 %
label:4        [ 0.  0.  0.  1.  0.  0.  1.  0. 18.  0.]              accuracy : 0 %
label:5        [0. 0. 0. 4. 0. 4. 4. 0. 8. 0.]              accuracy : 20 %
label:6        [ 0.  0.  0.  1.  0.  0.  7.  0. 12.  0.]              accuracy : 35 %
label:7        [ 0.  0.  0.  0.  0.  0.  2.  0. 18.  0.]              accuracy : 0 %
label:8        [ 0.  0.  0.  0.  0.  0.  6.  0. 14.  0.]              accuracy : 70 %
label:9        [ 0.  0.  0.  0.  0.  5.  5.  0. 10.  0.]              accuracy : 0 %

==================================================
```

[nbnSNR-10 데이터의 예측 결과]

```
testing segmented-val/nbnSNR-10
Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 200, 200
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =10.5


Confusion Matrix (row : True, column : Predicted)
label:0        [ 0.  0.  0.  9.  0.  0. 11.  0.  0.  0.]              accuracy : 0 %
label:1        [ 0.  0.  0.  8.  0.  2. 10.  0.  0.  0.]              accuracy : 0 %
label:2        [ 0.  0.  0.  9.  0.  0. 11.  0.  0.  0.]              accuracy : 0 %
label:3        [ 0.  0.  0.  9.  0.  0. 11.  0.  0.  0.]              accuracy : 45 %
label:4        [ 0.  0.  0.  9.  0.  0. 11.  0.  0.  0.]              accuracy : 0 %
label:5        [ 0.  0.  0.  9.  0.  1. 10.  0.  0.  0.]              accuracy : 5 %
label:6        [ 0.  0.  0.  9.  0.  0. 11.  0.  0.  0.]              accuracy : 55 %
label:7        [ 0.  0.  0. 10.  0.  0. 10.  0.  0.  0.]              accuracy : 0 %
label:8        [ 0.  0.  0. 10.  0.  0. 10.  0.  0.  0.]              accuracy : 0 %
label:9        [ 0.  0.  0. 10.  0.  0. 10.  0.  0.  0.]              accuracy : 0 %

==================================================
```

[wbnSNR-10 데이터의 예측 결과]

```
testing segmented-val/wbnSNR-10
Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 200, 200
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =11.0


Confusion Matrix (row : True, column : Predicted)
label:0        [ 0.  0.  0.  2.  0.  0.  0.  0. 18.  0.]              accuracy : 0 %
label:1        [ 0.  0.  0. 11.  0.  1.  0.  0.  8.  0.]              accuracy : 0 %
label:2        [ 0.  0.  0.  0.  0.  0.  0.  0. 20.  0.]              accuracy : 0 %
label:3        [ 0.  0.  0.  0.  0.  0.  0.  0. 20.  0.]              accuracy : 0 %
label:4        [ 0.  0.  0.  0.  0.  0.  0.  0. 20.  0.]              accuracy : 0 %
label:5        [ 0.  0.  0. 10.  0.  2.  0.  0.  8.  0.]              accuracy : 10 %
label:6        [ 0.  0.  0.  9.  0.  0.  0.  0. 11.  0.]              accuracy : 0 %
label:7        [ 0.  0.  0.  0.  0.  0.  0.  0. 20.  0.]              accuracy : 0 %
label:8        [ 0.  0.  0.  0.  0.  0.  0.  0. 20.  0.]              accuracy : 100 %
label:9        [ 0.  0.  0. 11.  0.  0.  0.  0.  9.  0.]              accuracy : 0 %

==================================================
```

Wide band noise일 때 일반적으로 narrow band noise보다 성능이 더 낮게 나왔는데, 이는 입력되는 quefrency에 더 많이 영향을 주기 때문인 것 같습니다. SNR이 클수록 HMM 모델이 모든 digit에 대해 하나의 숫자로 추정하는 경향이 있음을 알 수 있습니다. 이를 통해 각 digit을 학습한 HMM이 입력의 에너지의 영향을 받음을 알 수 있습니다.

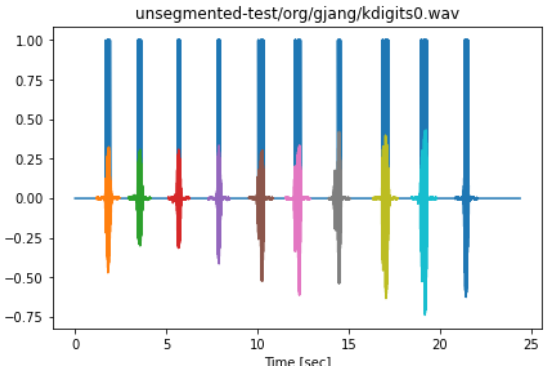
2.EPD를 이용해 unsegmented Data를 분리 후 HMM모델의 Test

```
def audio_statistics(file): # 데이터의 통계적인 특성을 확인하는 함수
    signal, Fs = librosa.load(file, sr=None, mono=False)
    Time_len=len(signal)/Fs
    mean_t=np.mean(signal)
    Var_t=np.mean((signal-mean_t)**2)
    max_t=np.max(np.abs(signal))
    return Time_len,mean_t,Var_t, max_t
```

코드 2 audio_statistic 함수

좌측의 함수를 정의해서 데이터의 통계적인 특성(평균, 분산, 각 데이터의 길이, 최대 진폭)을 확인할 수 있었습니다. Sample이 많으니까, 모든 sample의 최대 진폭을 평균내서 EPD Threshold로 정했습니다.

데이터의 길이가 1초 ~ 2초 정도되니까, 1초 간격을 내의 음성부분은 같은 숫자로 보고, 앞, 뒤로 0.5초 정도의 여유를 뒀습니다. 전체 EPD 결과의 평균보다 EPD에서 1로 예측한 시점에서 20 ms뒤의 시점까지의 평균이 작으면 무시하도록 구성했습니다.



EPD후의 분리된 음성들과 EPD의 값

좌측의 그림처럼 좌측의 EPD가 1로 예측한 부분들에서 1~10 까지의 육성이 잘 분리되었음을 알 수 있습니다. 이렇게 분리된 신호를 기존의 규칙에 맞게 kdigits0.wav -> kdigits0-x.wav
(x는 0~9)로 저장했습니다.

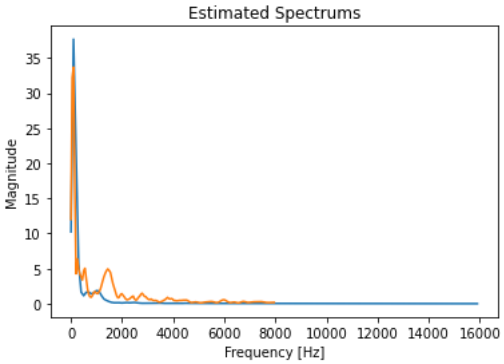
1.에서 학습된 HMM 모델을 사용해 각 digit을 잘 구분하는지 확인해 봤습니다. Org/gjjang/데이터에 대한 결과가 아래 그림과 같습니다.

```
Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 100, 100
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =29.0


Confusion Matrix (row : True, column : Predicted)
label:0         [0. 0. 0. 0. 0. 3. 7. 0. 0. 0.]                accuracy : 0 %
label:1         [ 0.  0.  0.  0.  0. 10.  0.  0.  0.  0.]               accuracy : 0 %
label:2         [0. 0. 2. 1. 0. 5. 2. 0. 0. 0.]                accuracy : 20 %
label:3         [0. 0. 0. 7. 0. 2. 1. 0. 0. 0.]                accuracy : 70 %
label:4         [0. 0. 0. 9. 0. 1. 0. 0. 0. 0.]                accuracy : 0 %
label:5         [0. 0. 0. 0. 0. 9. 1. 0. 0. 0.]                accuracy : 90 %
label:6         [ 0.  0.  0.  0.  0. 10.  0.  0.  0.  0.]               accuracy : 0 %
label:7         [0. 0. 0. 0. 0. 0. 3. 7. 0. 0.]                accuracy : 70 %
label:8         [0. 0. 0. 0. 0. 5. 5. 0. 0. 0.]                accuracy : 0 %
label:9         [0. 0. 0. 0. 0. 6. 0. 0. 0. 4.]                accuracy : 40 %

==================================================
```

결과 2 EPD된 음성신호의 예측결과

예측 성능이 validation때에 비해서는 안 나오지만, 실제로 들어봤을 때, 숫자들이 잘 분리된다는 것을 확인할 수 있습니다.

3.Noise suppression을 적용해 다양한 SNR에 대해 HMM 모델의 성능 변화

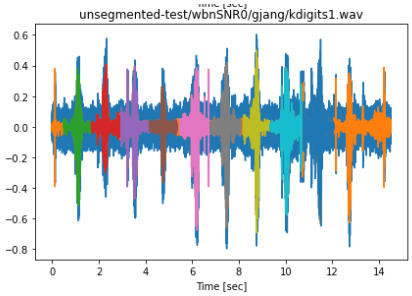


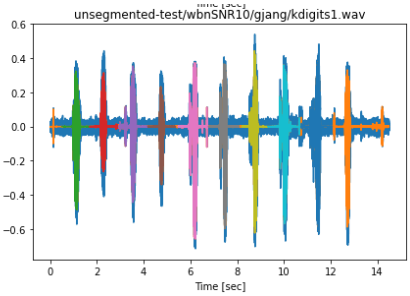
RMM으로 추정된 스펙트럼들

모든 경우 (wbn, nbn을 포함해 SNR -10dB, 0dB, 10dB)의 Validation데이터를 Multi-variable Rayleigh mixture model로 모델링하고, EM알고리즘을 적용해서 음성 신호와 노이즈를 구분했습니다. 데이터들에서 얻어진 mode 파라미터의 평균으로 노이즈의 스펙트럼 에너지를 추정했습니다.

좌측 그림에서 푸른색 선이 추정된 노이즈 스펙트럼이고, 주황색 선이 추정된 음성 신호의 스펙트럼입니다.

우측 그림이 noise suppression된 결과를 보여주고 있습니다. SNR이 0dB, 10dB 일 때의 경우에는 잘 되지만, SNR이 -10dB일 때는 잘 되지 않아서, HMM의 예측 성능이 떨어질 것으로 예상됩니다.





아래의 그림들을 통해 noise suppression이 적용되면 예측 성능이 좋아질 수 있다는 것을 확인할 수 있습니다.

```
Now, Using Test Data in nbnSNR-10
ex) ('unsegmented-test/nbnSNR-10/gjang_sep/kdigits0-0.wav', 0)

Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 100, 100
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =14.0


Confusion Matrix (row : True, column : Predicted)
label:0         [0. 0. 0. 7. 0. 1. 2. 0. 0. 0.]                accuracy : 0 %
label:1         [0. 0. 0. 1. 0. 9. 0. 0. 0. 0.]                accuracy : 0 %
label:2         [ 0.  0.  0. 10.  0.  0.  0.  0.  0.  0.]               accuracy : 0 %
label:3         [ 0.  0.  0. 10.  0.  0.  0.  0.  0.  0.]               accuracy : 100 %
label:4         [ 0.  0.  0. 10.  0.  0.  0.  0.  0.  0.]               accuracy : 0 %
label:5         [0. 0. 0. 7. 0. 3. 0. 0. 0. 0.]                accuracy : 30 %
label:6         [0. 0. 0. 5. 0. 4. 1. 0. 0. 0.]                accuracy : 10 %
label:7         [0. 0. 0. 6. 0. 3. 1. 0. 0. 0.]                accuracy : 0 %
label:8         [0. 0. 0. 6. 0. 4. 0. 0. 0. 0.]                accuracy : 0 %
label:9         [0. 0. 0. 7. 0. 3. 0. 0. 0. 0.]                accuracy : 0 %
```

```
Now, Using Test Data in nbnSNR-10
ex) ('unsegmented-test/nbnSNR-10/gjang_epded/kdigits0-0.wav', 0)

Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 100, 100
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =9.0


Confusion Matrix (row : True, column : Predicted)
label:0         [0. 0. 0. 3. 0. 0. 6. 0. 1. 0.]                accuracy : 0 %
label:1         [0. 0. 0. 1. 0. 0. 6. 0. 3. 0.]                accuracy : 0 %
label:2         [0. 0. 0. 1. 0. 0. 9. 0. 0. 0.]                accuracy : 0 %
label:3         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
label:4         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
label:5         [0. 0. 0. 0. 0. 1. 6. 0. 3. 0.]                accuracy : 10 %
label:6         [0. 0. 0. 0. 0. 0. 7. 0. 3. 0.]                accuracy : 70 %
label:7         [0. 0. 0. 2. 0. 0. 6. 0. 2. 0.]                accuracy : 0 %
label:8         [0. 0. 0. 1. 0. 0. 8. 0. 1. 0.]                accuracy : 10 %
label:9         [0. 0. 0. 1. 0. 0. 7. 0. 2. 0.]                accuracy : 0 %
```

```
Now, Using Test Data in nbnSNR0
ex) ('unsegmented-test/nbnSNR0/gjang_sep/kdigits0-0.wav', 0)

Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 100, 100
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =14.0


Confusion Matrix (row : True, column : Predicted)
label:0         [0. 0. 0. 0. 0. 2. 8. 0. 0. 0.]                accuracy : 0 %
label:1         [0. 0. 0. 0. 0. 1. 9. 0. 0. 0.]                accuracy : 0 %
label:2         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
label:3         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
label:4         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
label:5         [0. 0. 0. 0. 0. 5. 5. 0. 0. 0.]                accuracy : 50 %
label:6         [0. 0. 0. 0. 0. 1. 9. 0. 0. 0.]                accuracy : 90 %
label:7         [0. 0. 0. 0. 0. 2. 8. 0. 0. 0.]                accuracy : 0 %
label:8         [0. 0. 0. 0. 0. 1. 9. 0. 0. 0.]                accuracy : 0 %
label:9         [0. 0. 0. 0. 0. 4. 6. 0. 0. 0.]                accuracy : 0 %
```

```
Now, Using Test Data in nbnSNR0
ex) ('unsegmented-test/nbnSNR0/gjang_epded/kdigits0-0.wav', 0)

Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 100, 100
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =23.0


Confusion Matrix (row : True, column : Predicted)
label:0         [0. 0. 0. 3. 0. 0. 1. 1. 2. 3.]                accuracy : 0 %
label:1         [0. 0. 0. 2. 0. 0. 0. 7. 0. 1.]                accuracy : 0 %
label:2         [0. 2. 0. 6. 0. 0. 0. 0. 1. 1.]                accuracy : 0 %
label:3         [0. 0. 0. 6. 0. 0. 0. 0. 3. 1.]                accuracy : 60 %
label:4         [0. 0. 0. 1. 0. 0. 0. 1. 7. 1.]                accuracy : 0 %
label:5         [0. 2. 0. 0. 0. 2. 0. 2. 1. 3.]                accuracy : 20 %
label:6         [0. 0. 0. 0. 0. 2. 0. 1. 0. 7.]                accuracy : 0 %
label:7         [0. 0. 0. 0. 0. 0. 0. 7. 0. 3.]                accuracy : 70 %
label:8         [0. 2. 0. 1. 0. 0. 1. 1. 3. 2.]                accuracy : 30 %
label:9         [0. 0. 0. 1. 0. 0. 0. 4. 0. 5.]                accuracy : 50 %
```

```
Now, Using Test Data in nbnSNR10
ex) ('unsegmented-test/nbnSNR10/gjang_sep/kdigits0-0.wav', 0)

Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 100, 100
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =15.0


Confusion Matrix (row : True, column : Predicted)
label:0         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
label:1         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
label:2         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
label:3         [0. 0. 0. 4. 0. 0. 6. 0. 0. 0.]                accuracy : 40 %
label:4         [0. 0. 0. 3. 0. 0. 7. 0. 0. 0.]                accuracy : 0 %
label:5         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
label:6         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 100 %
label:7         [0. 0. 0. 0. 0. 0. 9. 1. 0. 0.]                accuracy : 10 %
label:8         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
label:9         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
```

```
Now, Using Test Data in nbnSNR10
ex) ('unsegmented-test/nbnSNR10/gjang_epded/kdigits0-0.wav', 0)

Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 100, 100
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =25.0


Confusion Matrix (row : True, column : Predicted)
label:0         [10.  0.  0.  0.  0.  0.  0.  0.  0.  0.]               accuracy : 100 %
label:1         [1. 9. 0. 0. 0. 0. 0. 0. 0. 0.]                accuracy : 90 %
label:2         [4. 6. 0. 0. 0. 0. 0. 0. 0. 0.]                accuracy : 0 %
label:3         [4. 6. 0. 0. 0. 0. 0. 0. 0. 0.]                accuracy : 0 %
label:4         [7. 3. 0. 0. 0. 0. 0. 0. 0. 0.]                accuracy : 0 %
label:5         [5. 5. 0. 0. 0. 0. 0. 0. 0. 0.]                accuracy : 0 %
label:6         [9. 1. 0. 0. 0. 0. 0. 0. 0. 0.]                accuracy : 0 %
label:7         [4. 0. 0. 0. 0. 0. 0. 6. 0. 0.]                accuracy : 60 %
label:8         [9. 0. 0. 0. 0. 0. 0. 1. 0. 0.]                accuracy : 0 %
label:9         [3. 0. 0. 0. 0. 0. 0. 7. 0. 0.]                accuracy : 0 %
```

```
Now, Using Test Data in wbnSNR-10
ex) ('unsegmented-test/wbnSNR-10/gjang_sep/kdigits0-0.wav', 0)

Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 100, 100
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =8.0


Confusion Matrix (row : True, column : Predicted)
label:0         [0. 0. 0. 5. 0. 0. 0. 0. 5. 0.]                accuracy : 0 %
label:1         [ 0.  0.  0. 10.  0.  0.  0.  0.  0.  0.]               accuracy : 0 %
label:2         [0. 0. 0. 8. 0. 0. 0. 0. 2. 0.]                accuracy : 0 %
label:3         [0. 0. 0. 3. 0. 0. 0. 0. 7. 0.]                accuracy : 30 %
label:4         [0. 0. 0. 1. 0. 0. 0. 0. 9. 0.]                accuracy : 0 %
label:5         [0. 0. 0. 8. 0. 0. 0. 0. 2. 0.]                accuracy : 0 %
label:6         [0. 0. 0. 7. 0. 0. 0. 0. 3. 0.]                accuracy : 0 %
label:7         [0. 0. 0. 8. 0. 0. 0. 0. 2. 0.]                accuracy : 0 %
label:8         [0. 0. 0. 5. 0. 0. 0. 0. 5. 0.]                accuracy : 50 %
label:9         [0. 0. 0. 8. 0. 0. 0. 0. 2. 0.]                accuracy : 0 %
```

```
Now, Using Test Data in wbnSNR-10
ex) ('unsegmented-test/wbnSNR-10/gjang_epded/kdigits0-0.wav', 0)

Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 100, 100
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =11.0


Confusion Matrix (row : True, column : Predicted)
label:0         [0. 0. 0. 1. 0. 1. 0. 0. 8. 0.]                accuracy : 0 %
label:1         [0. 0. 0. 1. 0. 0. 0. 0. 9. 0.]                accuracy : 0 %
label:2         [ 0.  0.  0.  0.  0.  0.  0.  0. 10.  0.]               accuracy : 0 %
label:3         [ 0.  0.  0.  0.  0.  0.  0.  0. 10.  0.]               accuracy : 0 %
label:4         [ 0.  0.  0.  0.  0.  0.  0.  0. 10.  0.]               accuracy : 0 %
label:5         [0. 0. 0. 1. 0. 3. 0. 0. 6. 0.]                accuracy : 30 %
label:6         [0. 0. 0. 0. 0. 0. 1. 0. 9. 0.]                accuracy : 0 %
label:7         [0. 0. 0. 2. 0. 0. 0. 0. 8. 0.]                accuracy : 0 %
label:8         [0. 0. 0. 1. 0. 1. 0. 0. 8. 0.]                accuracy : 80 %
label:9         [0. 0. 0. 2. 0. 0. 0. 0. 8. 0.]                accuracy : 0 %
```

```
Now, Using Test Data in wbnSNR0
ex) ('unsegmented-test/wbnSNR0/gjang_sep/kdigits0-0.wav', 0)

Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 100, 100
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =16.0


Confusion Matrix (row : True, column : Predicted)
label:0         [0. 0. 0. 0. 0. 8. 2. 0. 0. 0.]                accuracy : 0 %
label:1         [0. 0. 0. 0. 0. 7. 3. 0. 0. 0.]                accuracy : 0 %
label:2         [0. 0. 0. 2. 0. 4. 1. 0. 3. 0.]                accuracy : 0 %
label:3         [0. 0. 0. 4. 0. 2. 0. 0. 4. 0.]                accuracy : 40 %
label:4         [0. 0. 0. 2. 0. 1. 0. 0. 7. 0.]                accuracy : 0 %
label:5         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 100 %
label:6         [0. 0. 0. 0. 0. 8. 2. 0. 0. 0.]                accuracy : 20 %
label:7         [0. 0. 0. 2. 0. 4. 3. 0. 1. 0.]                accuracy : 0 %
label:8         [0. 0. 0. 0. 0. 8. 2. 0. 0. 0.]                accuracy : 0 %
label:9         [ 0.  0.  0.  0.  0. 10.  0.  0.  0.  0.]               accuracy : 0 %
```

```
Now, Using Test Data in wbnSNR0
ex) ('unsegmented-test/wbnSNR0/gjang_epded/kdigits0-0.wav', 0)

Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 100, 100
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =10.0


Confusion Matrix (row : True, column : Predicted)
label:0         [0. 0. 0. 0. 0. 0. 1. 0. 9. 0.]                accuracy : 0 %
label:1         [0. 0. 0. 0. 0. 0. 2. 0. 8. 0.]                accuracy : 0 %
label:2         [ 0.  0.  0.  0.  0.  0.  0.  0. 10.  0.]               accuracy : 0 %
label:3         [ 0.  0.  0.  0.  0.  0.  0.  0. 10.  0.]               accuracy : 0 %
label:4         [ 0.  0.  0.  0.  0.  0.  0.  0. 10.  0.]               accuracy : 0 %
label:5         [ 0.  0.  0.  0.  0.  0.  0.  0. 10.  0.]               accuracy : 0 %
label:6         [ 0.  0.  0.  0.  0.  0.  0.  0. 10.  0.]               accuracy : 0 %
label:7         [0. 0. 0. 0. 0. 0. 1. 0. 9. 0.]                accuracy : 0 %
label:8         [ 0.  0.  0.  0.  0.  0.  0.  0. 10.  0.]               accuracy : 100 %
label:9         [0. 0. 0. 0. 0. 0. 2. 0. 8. 0.]                accuracy : 0 %
```

```
Now, Using Test Data in wbnSNR10
ex) ('unsegmented-test/wbnSNR10/gjang_sep/kdigits0-0.wav', 0)

Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 100, 100
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =17.0


Confusion Matrix (row : True, column : Predicted)
label:0         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
label:1         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
label:2         [0. 0. 0. 1. 0. 0. 8. 0. 1. 0.]                accuracy : 0 %
label:3         [0. 0. 0. 6. 0. 0. 3. 0. 1. 0.]                accuracy : 60 %
label:4         [0. 0. 0. 5. 0. 0. 3. 0. 2. 0.]                accuracy : 0 %
label:5         [0. 0. 0. 0. 0. 0. 1. 9. 0. 0.]                accuracy : 10 %
label:6         [ 0.  0.  0.  0.  0.  0.  0. 10.  0.  0.]               accuracy : 100 %
label:7         [0. 0. 0. 0. 0. 0. 0. 9. 0. 1.]                accuracy : 0 %
label:8         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
label:9         [ 0.  0.  0.  0.  0.  0. 10.  0.  0.  0.]               accuracy : 0 %
```

```
Now, Using Test Data in wbnSNR10
ex) ('unsegmented-test/wbnSNR10/gjang_epded/kdigits0-0.wav', 0)

Words spoken: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
[validation] number of labels and features = 100, 100
Loading data completed
Prediction started

Prediction for Testing DataSet:

accuracy =35.0


Confusion Matrix (row : True, column : Predicted)
label:0         [5. 1. 0. 4. 0. 0. 0. 0. 0. 0.]                accuracy : 50 %
label:1         [0. 8. 0. 1. 0. 1. 0. 0. 0. 0.]                accuracy : 80 %
label:2         [0. 3. 0. 3. 0. 0. 0. 2. 2. 0.]                accuracy : 0 %
label:3         [0. 0. 0. 7. 0. 0. 0. 0. 3. 0.]                accuracy : 70 %
label:4         [0. 0. 0. 4. 0. 0. 0. 2. 4. 0.]                accuracy : 0 %
label:5         [0. 6. 0. 0. 0. 1. 0. 0. 3. 0.]                accuracy : 10 %
label:6         [1. 6. 0. 0. 0. 2. 0. 1. 0. 0.]                accuracy : 0 %
label:7         [ 0.  0.  0.  0.  0.  0.  0.  0. 10.  0.]               accuracy : 100 %
label:8         [2. 3. 0. 0. 0. 2. 1. 0. 1. 1.]                accuracy : 10 %
label:9         [0. 0. 0. 1. 0. 0. 0. 6. 0. 3.]                accuracy : 30 %
```

결과 3 Noise suppression 전/ 후 예측 결과

좌측의 그림에서 굵은 테두리가 쳐진 부분이 noise suppression이 적용된 결과입니다.

SNR이 10dB인 경우, noise suppression이 적용된 경우가 확연하게 더 좋은 예측 성능을 보입니다.

SNR이 0dB인 경우, narrow band noise에서는 예측 성능이 좋아지지만, wide band에서는 성능이 조금 떨어지는 경향을 보였습니다.

SNR이 -10dB인 경우, wide band noise에서는 성능이 좋아지지만, narrow band에서는 조금 성능이 떨어지는 경향을 보였습니다.

따라서, MFCC의 차원을 작게 가져가면, Narrow band 노이즈에 대해서는 숫자 인식이 잘 되지만, wide band의 경우에는 MFCC의 차원이 더 필요하다고 생각됩니다.