EE624 : SPEECH TECHNOLOGY

PROJECT REPORT

NAME : S. SRI SAI KOUSHIK **BRANCH** : EEE

ROLL NO : 190108050 **DATE** : 13-05-2023

1.

For 1st question, data folder contains all the recordings of 25 utterances in their specific folders.

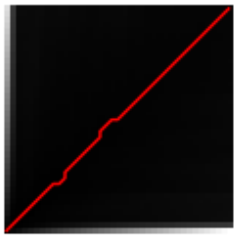
Code working:

The **project1.ipynb** first calculates the mfcc features of utterances and sent it to dtw function from dtw library, it gives the dtw score and alignment path. With the help of matplotlib library we plotted the DTW curves. The output from the is given below.

DTW curves and plots are given from next page:

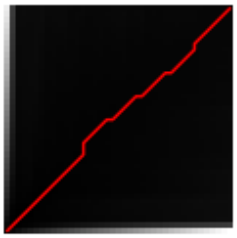
for [a] :

5860.476947784424

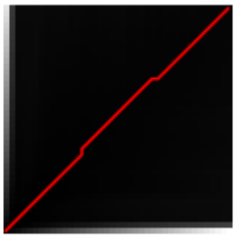


5481.437183380127

6518.602493286133

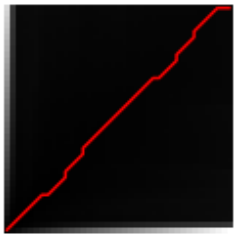


5678.330627441406



6052.8681564331055

7157.3231201171875



6715.903480529785

6720.908123016357

6426.443958282471

6656.870704650879

7350.622161865234

5555.6189041137695

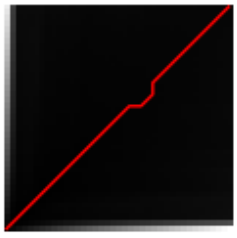
5849.977569580078

5686.926490783691

6305.209335327148

6272.684585571289

6646.127361297607

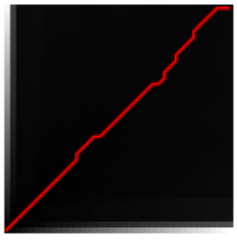


7307.729824066162

6243.60954284668

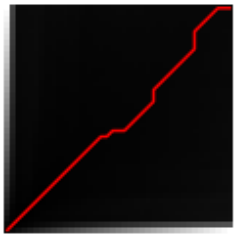
7492.180450439453

6476.92077255249



6156.865898132324

6727.251567840576



6284.222785949707

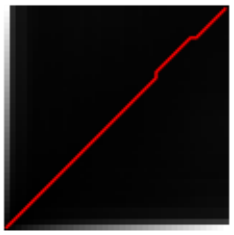
for [e]:

4950.397323608398

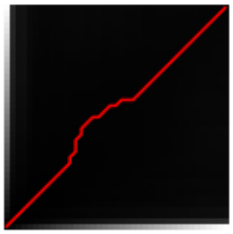
4962.649040222168

6062.584030151367

5571.547698974609



5633.637538909912



6123.176471710205

4967.033004760742

5883.8112144470215

5101.546276092529

4767.162658691406

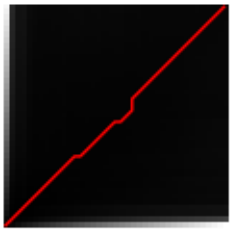
5261.9112548828125

4960.085121154785

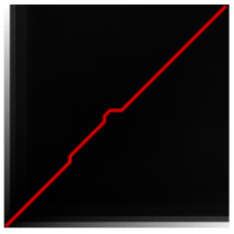
5118.559989929199

5220.152770996094

7226.272239685059



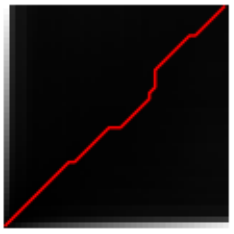
6769.628448486328



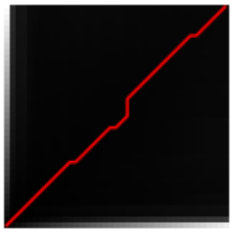
7541.721076965332

7393.409980773926

6791.341815948486

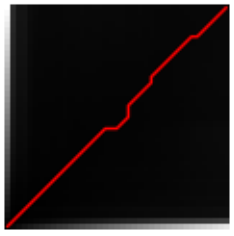


7432.761985778809

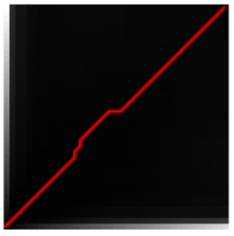


7848.406063079834

7862.687644958496



7364.1635093688965



7963.1549072265625

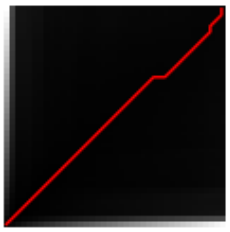
for [i]:

4763.098434448242

4708.850193023682

5879.63996887207

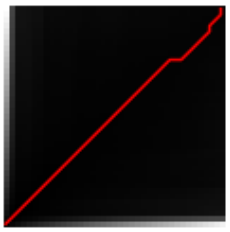
5334.4758224487305



5739.894153594971

6420.506332397461

4796.019096374512



6360.241535186768

5772.963188171387

5068.459461212158

5463.027847290039

5224.025554656982

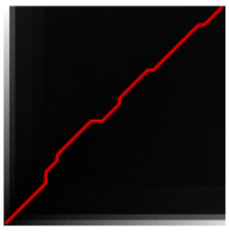
5515.148906707764

5623.83629989624

7547.671722412109

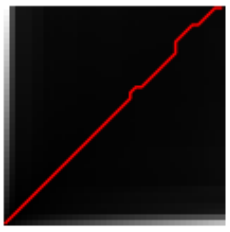
8085.818981170654

7757.916549682617



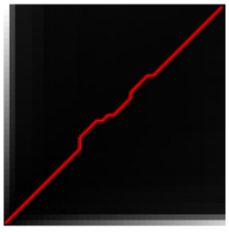
7069.765739440918

8228.993923187256



7033.024810791016

7411.604042053223



7765.93087387085

6826.466522216797

8639.177528381348

for [o]:

5116.930938720703

5139.677921295166

5294.030967712402

5675.454360961914

5930.075889587402

5323.666610717773

5497.400428771973

6880.8889236450195

5823.290725708008

4502.360019683838

5832.6262550354

4951.688873291016

5774.764503479004

5388.863952636719

6212.326545715332

5916.839954376221

6954.0908126831055

6736.008350372314

6519.221202850342

6760.871326446533

6989.930213928223

7043.062740325928

6765.253311157227

6942.930530548096

for [u]:

4374.713912963867

5699.643829345703

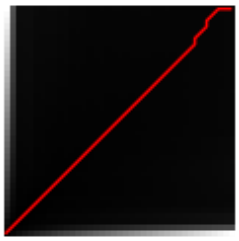
8161.4702224731445

6490.068965911865

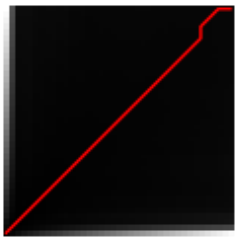
6833.607933044434

7408.06213760376

6705.895923614502



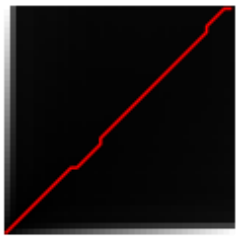
6998.453067779541



7641.94051361084

6290.432441711426

6903.027000427246

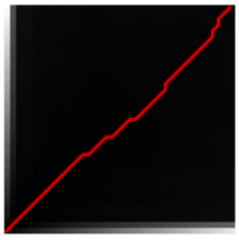


7317.2627029418945

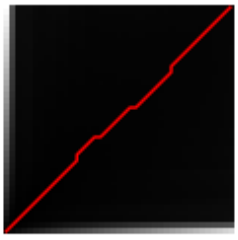
6681.615345001221

6800.070556640625

6544.487236022949

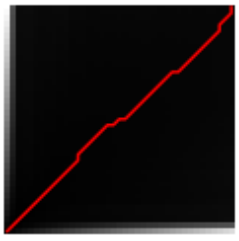


6402.278923034668



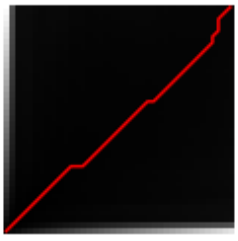
7144.449974060059

7913.660266876221



7984.7710037231445

6439.452983856201



7066.861427307129

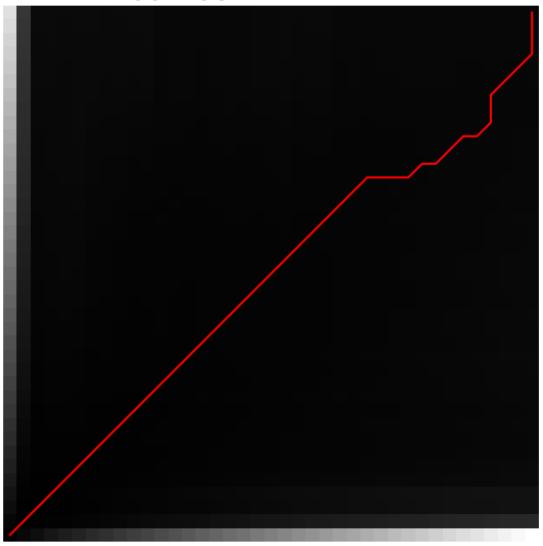
6238.503402709961

6440.87829208374

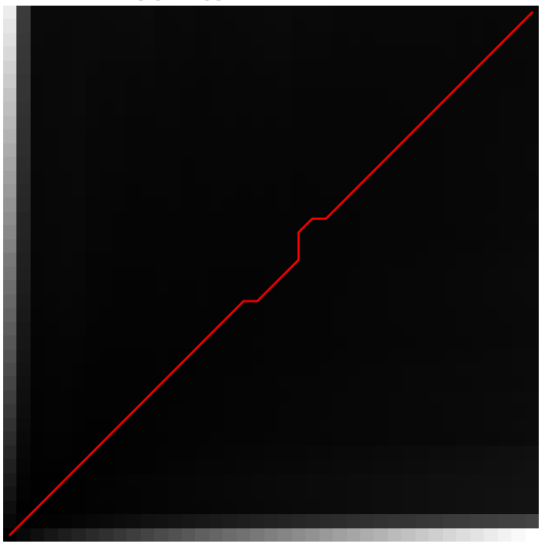
5895.094146728516

for [a] and different items:

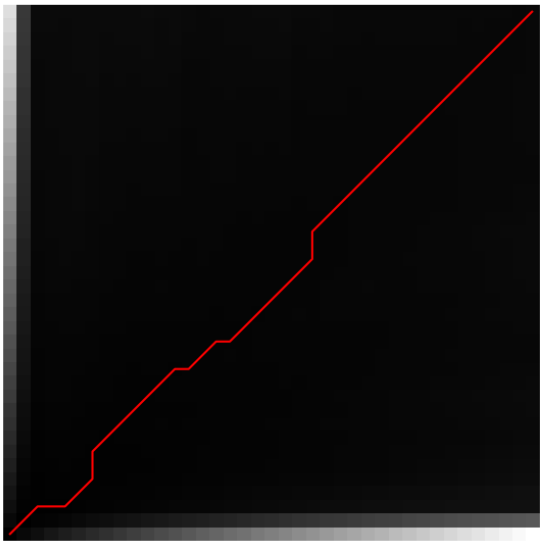
[a] and [e]:7524.409202575684



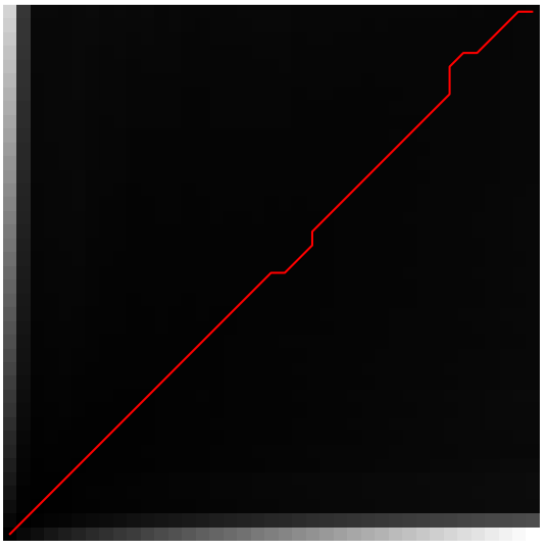
[a] and [i]:8326.598957061768



[a] and [o]:8309.292694091797

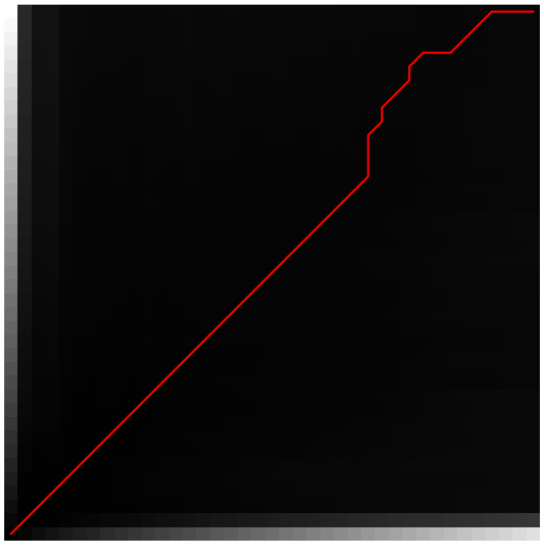


[a] and [u]:8051.348201751709

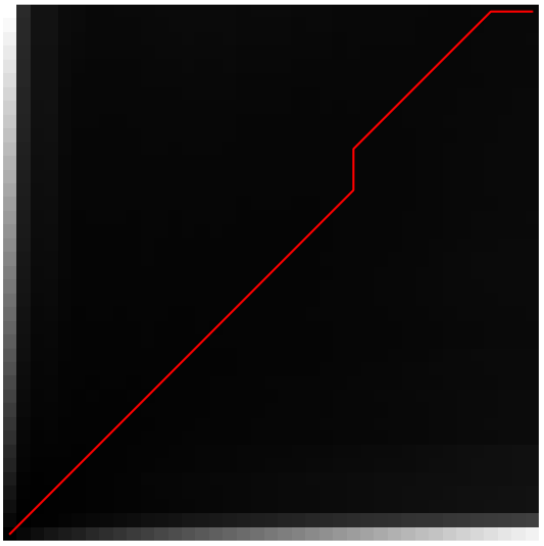


for [e] and different items:

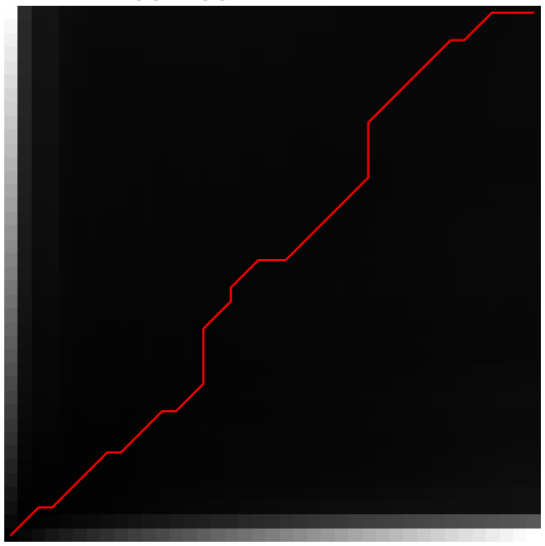
[e] and [a]:7524.409202575684



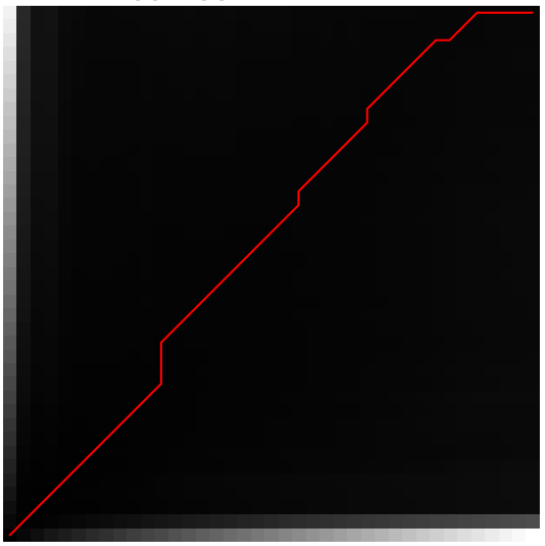
[e] and [i]:8723.512069702148



[e] and [o]:9016.326694488525

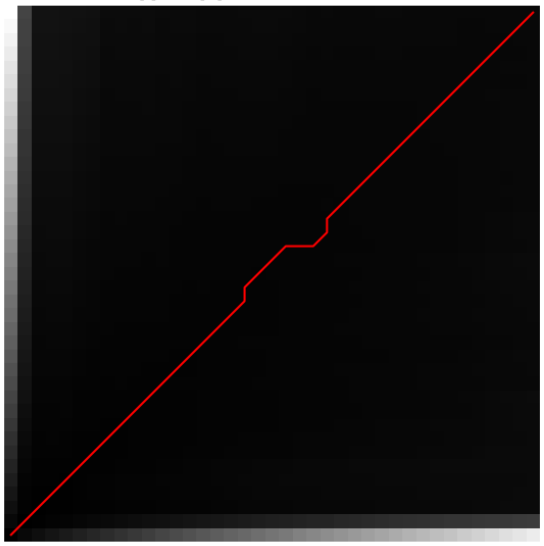


[e] and [u]:8079.996753692627

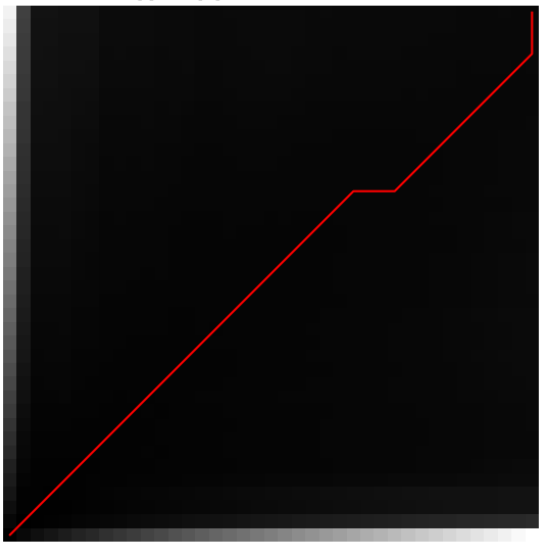


for [i] and different items:

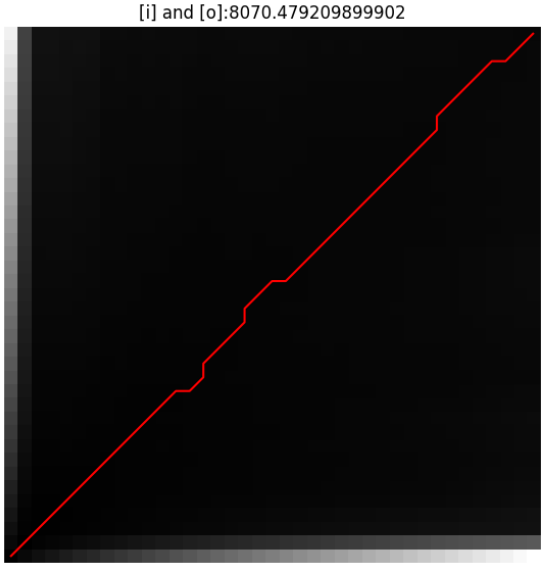
[i] and [a]:8326.598957061768



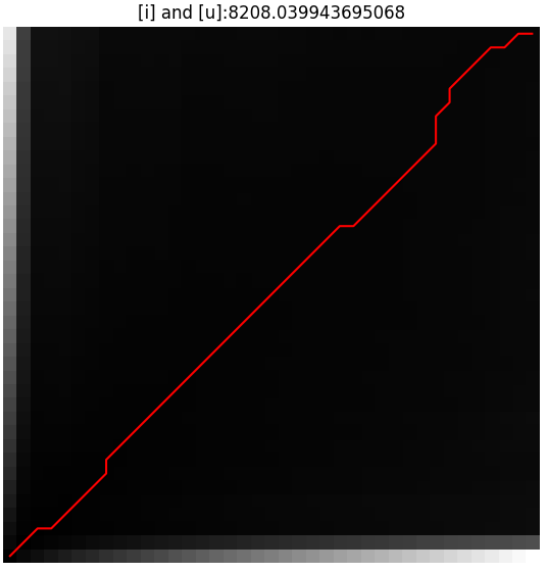
[i] and [e]:8723.512069702148



[i] and [o]:8070.479209899902



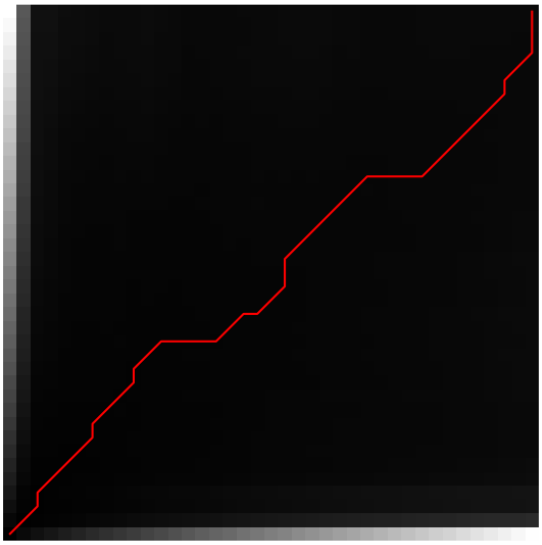
[i] and [u]:8208.039943695068



for [o] and different items:

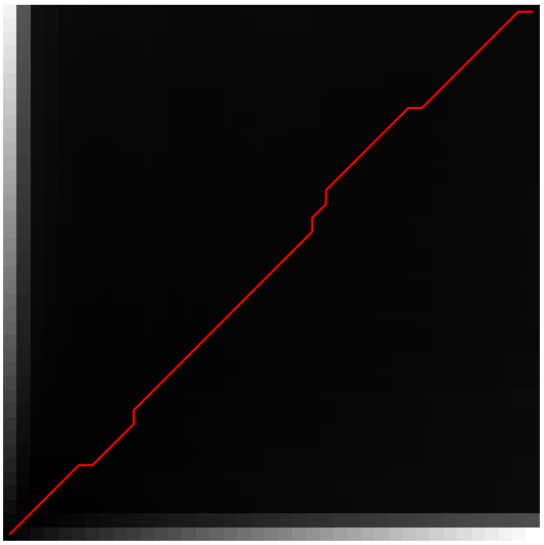
[o] and [a]:8309.292694091797

[o] and [e]:9016.326694488525



[o] and [i]:8070.479209899902

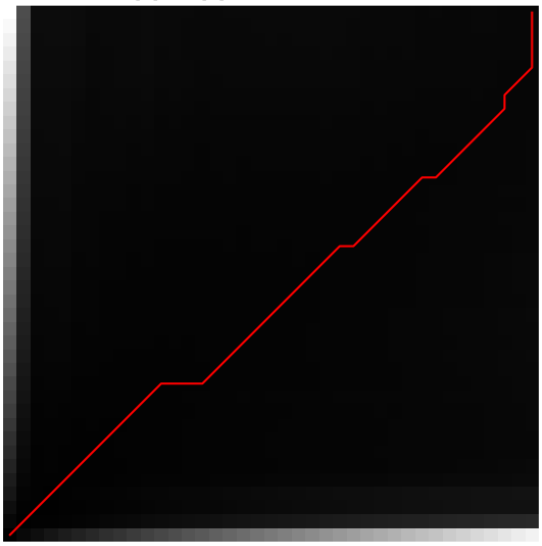
[o] and [u]:7063.869972229004



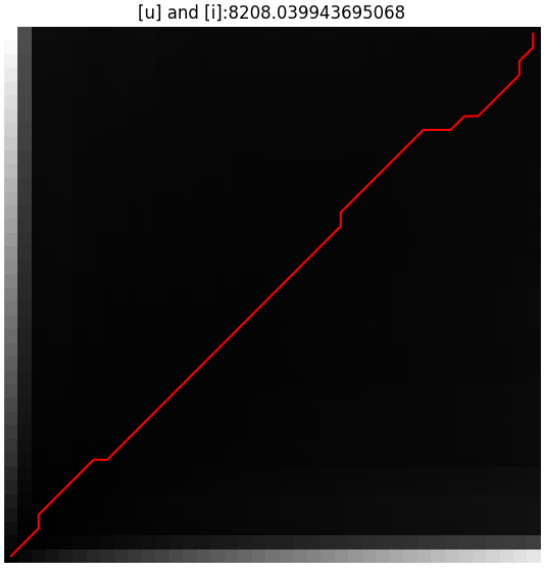
for [u] and different items:

[u] and [a]:8051.348201751709

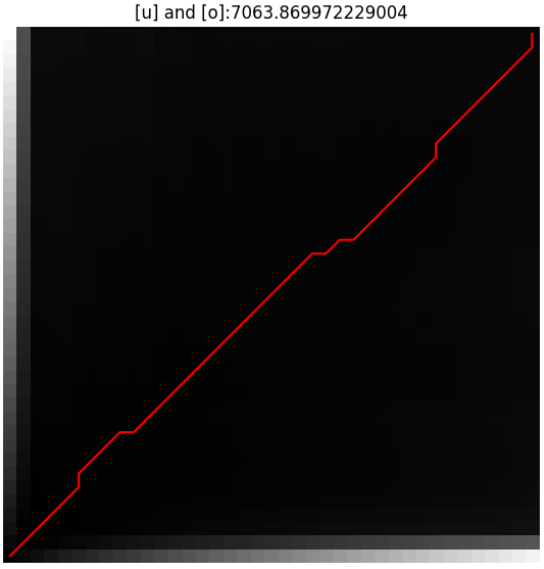
[u] and [e]:8079.996753692627



[u] and [i]:8208.039943695068



[u] and [o]:7063.869972229004



(2)

For 2nd question, the data used is present in the folder named data, each digit is present in its own respective folder.

Code working:

The **project2.ipynb** ,5 samples were taken from each case and mfcc features were calculated, after calculating these mfcc features, 32 component GMM model is trained for each digit, after that 20 samples were used as test data, and confusion matrix for final result is this:

```
[[20  0  0  0  0  0  0  0  0  0]
 [ 0 16  0  0  0  0  0  4  0  0]
 [ 0  0 20  0  0  0  0  0  0  0]
 [ 0  0  0 20  0  0  0  0  0  0]
 [ 0  0  0  0 20  0  0  0  0  0]
 [ 0  0  0  0  0 18  0  2  0  0]
 [ 0  0  0  0  0  0 20  0  0  0]
 [ 0  0  0  0  0  0  0 20  0  0]
 [ 0  0  0  0  0  0  0  0 20  0]
 [ 0  0  0  0  0  0  0  0  0 20]]
```

Observations:

Here, for [1], 4 samples were misplaced as [7]

And for [5], 2 samples were misplaced as [7]

We can improve the efficiency of the model by increasing the training set and increasing number of components in GMM model.

4.

For 4th question, the data used is present in the folder named data4, total 5 speakers data were used with speech containing **“Hello world, how are you”** of 2sec.

Per each speaker, 3 samples were used as training and 8 samples as testing data.

Code working:

The **project4.ipynb** , first it contains the code running speaker identification based on dtw score of normalized mfcc vectors, the confusion matrix for this part is:

```
[[8 0 0 0 0]
 [0 8 0 0 0]
 [6 0 2 0 0]
 [0 0 0 8 0]
 [0 0 0 0 8]]
```

And for second part, we calculated Gaussian posteriograms of testing data, whose GMM’s are trained from labeled training data present in each speaker folder. The confusion matrix obtained here are:

```
[[8 0 0 0 0]
 [0 8 0 0 0]
 [0 1 7 0 0]
 [0 0 0 8 0]
 [0 0 0 0 8]]
```

Clearly, we can see that the model accuracy increased by using GMM and gaussian posteriorgrams for dtw score method, instead of mfcc features.