BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING

Course No. : EEE 312
Course Title: Digital Signal Processing(I) laboratory

Group No : 1
Section : B1

Project Report

Title:

Keyword Identification from Speech

Submitted By:

Student ID	Name
1906066	Sheikh Abu Al Raihan
1906067	Kazi Abid Hasan
1906068	Md. Hasnat Karim Rupak
1906069	Md. Nahid Mustafa

Date of Submission: 08/03/2023

Problem Statement

Title: Keyword Identification from Speech

"Identify which keyword was uttered."

Bonus:

- Identify a foreign keyword (Not in the dataset)
- Identify keyword from sentence

Set of keywords-

["Chips", "Coffee", "Juice", "Candy"]

Proposed Methodology

First of all, we started to collect audio samples of our selected keywords from people of different age and different gender in different accent, and created an audio library. Then we have started the coding part. We executed the whole process in 4 steps-

1. Extraction 'MFCC Coefficients' of the collected audio signals.
2. Taking input audio signal as 'Test Signal' from users and extraction MFCC coefficients of the Test Signal too.
3. Matching the MFCC coefficients of the Test Signal with the MFCC coefficients of the audio signals from our audio library using 'DTW ALGORITHM'.
4. Comparing the DTW Algorithm's output and decide which sample from our audio library matches the Test Signal most.

MFCC (Mel Frequency Cepstral Coefficients) of Audio Signals:

MFCC algorithm extracts 13 features of an audio Signal. In this methodology a set of work have been performed by the compiler.

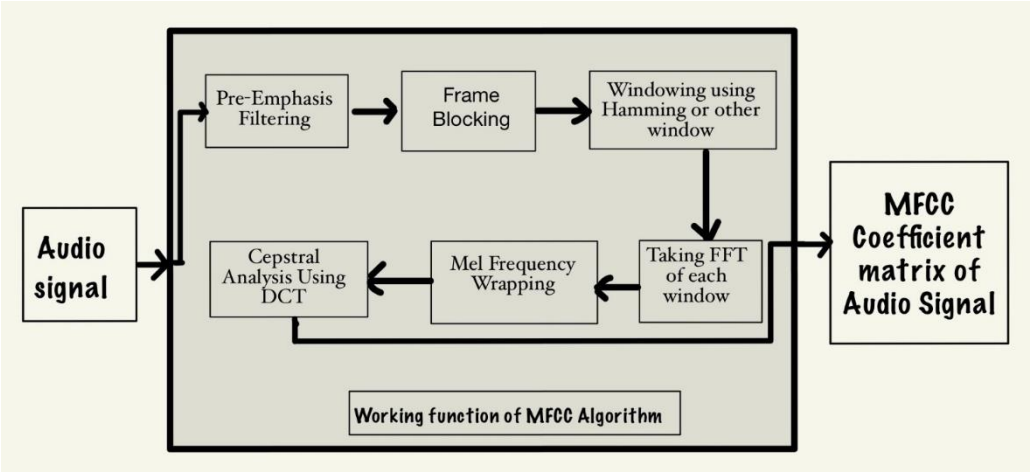

Fig: Working Function of MFCC Algorithm (Block Diagram)

DTW Algorithm:

The full form of the term DTW is 'Dynamic Time Wrapping' algorithm. This is a technique of matching two dataset's equivalent points. Executing this method we can distinguish the most similar dataset among all the datasets provided using defining a 'cost matrix'.

Here we discuss the overall idea of this algorithm with the help of an example.

Let's take 2 datasets

W = [0, 2, 0, 2, 0, 2, 1, 0] & X = [0, 2, 0, 1, 0, 0]

And a test dataset , Y = [0, 0, 0.5, 2, 0, 1, 0]

Now, we have to match this dataset Y with W and X and have to determine which dataset between W & X is more similar with the test dataset Y. Here we have to perform a calculation using the cost matrix.

How does the Cost Matrix work!

Firstly, Cost matrix takes 2 datasets and then help us to determine which data point(s) of one dataset is representing the same data point(s) of the other dataset. Beside finding the similar data points of each dataset, cost matrix determines the difference between similar data points of the datasets. Then assign a 'cost value' for each modification. Cost value denotes how much dissimilarities are there present in the two datasets. After determining all the datasets cost matrix with the testing dataset, we can identify which dataset is more similar to the given dataset comparing the cost values.

➢ Cost Matrix:

Calculation of Cost Matrix:

$D_{ij} = d(x_i, y_j) + \min \{ D_{i-1, j-1}\ D_{i-1, j}\ D_{i, j-1} \}$

Where,

D_{ij} = value of the matrix in i^{th} row and j^{th} coloumn

$d(x_i, y_j)$ = difference between x_i & y_j

Red colored arrows tell us which value from $D_{i-1, j-1}\ D_{i-1, j}\ D_{i, j-1}$ is contributing in the equation of 'D_{ij}'.

➢ Cost Matrix of X & Y:

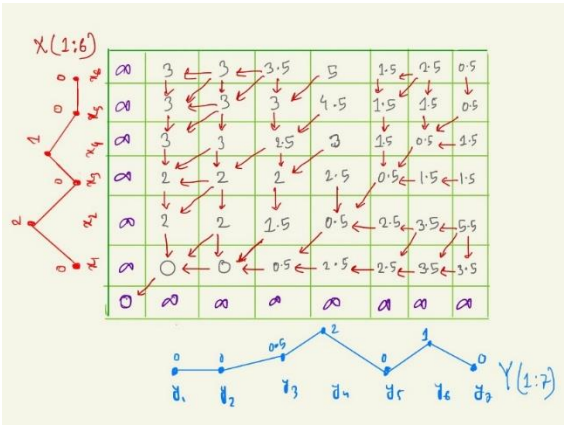

Fig: Cost Matrix having cost values

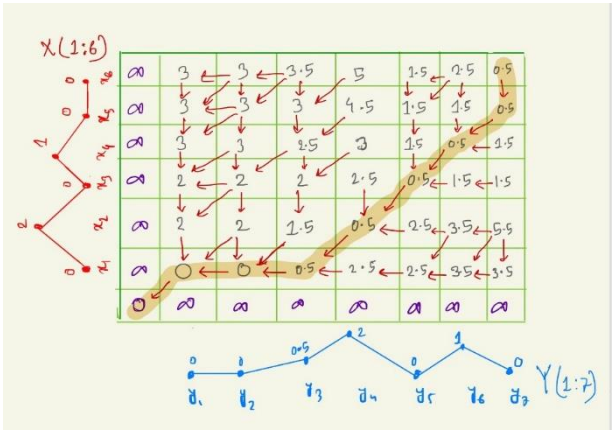

Fig: Cost Matrix of minimum cost path while matching 2 datasets

Minimum cost path in the cost matrix tells us which point of X is similar with which point of Y resulting minimum cost.

- If the D_{ij} comes from the $D_{i-1, j-1}$ then the i^{th} & j^{th} value of the 2 datasets matches.
- If the D_{ij} comes from the $D_{i, j-1}$ then the i^{th} & j^{th} value of the 2 datasets doesn't match, and we need to delete the j^{th} datapoint from the Y dataset.
- If the D_{ij} comes from the $D_{i-1, j}$ then the i^{th} & j^{th} value of the 2 datasets doesn't match, and we need to insert a similar datapoint in the Y dataset.

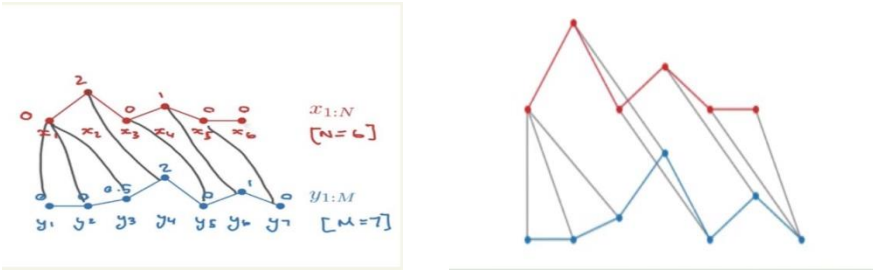

Fig: Similar points in both Dataset (X & Y)

The $D_{M,N}$ (where M is the last datapoint of X and N is the last datapoint of Y) value of the Cost Matrix shows the minimal cost of matching the 2 Dataset. For example, in the X & Y Dataset we have created a modified Y_1 dataset which carries the inserted data point(s) and doesn't carries the deleted datapoint(s) of Y. And the overall cost of this matching is **0.5**.

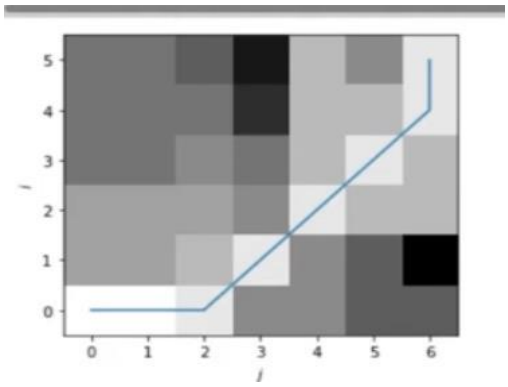

Fig: Cost Matrix (Gradient Graph)

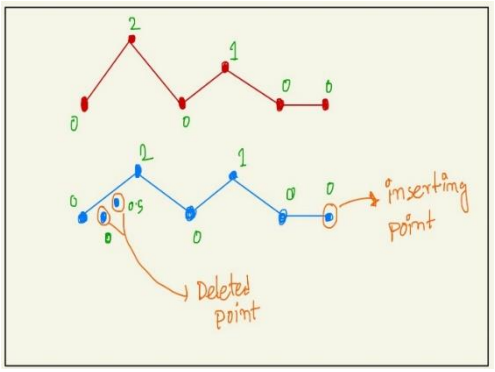

Fig: Matched Dataset after insertion & deletion of datapoints

Now, If we take the Cost matrix of Dataset W & Y, we will get that the cost of matching W & Y will be higher than the cost of X & Y.

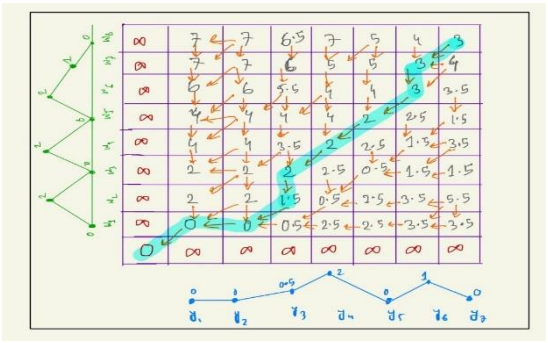

Fig: Cost Matrix having cost values & minimum cost path while matching 2 datasets

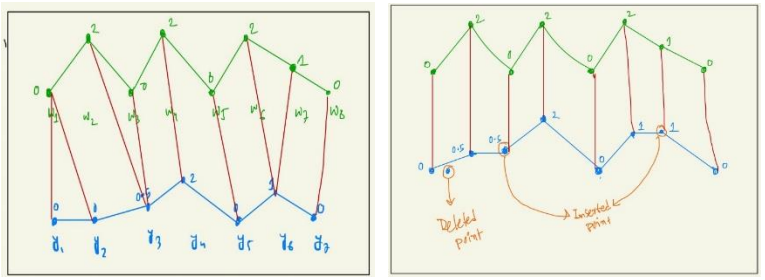

Fig: Similar points in both Dataset

After insertion and deletion, we can calculate value by value the cost of the modification in the process and then adding the cost of each value, we can find out which cost matrix has minimal cost and point it out as the most similar dataset with the given dataset.

Thus, DTW determines which given dataset is more closer to the Testing Dataset considering the overall cost.

Results

Code Snaps:

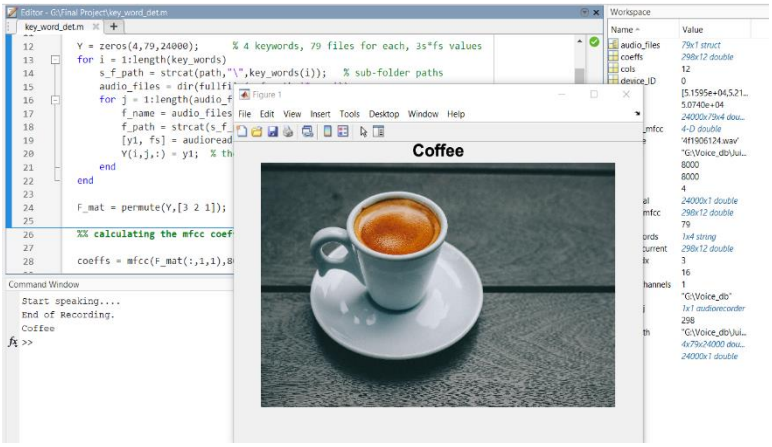

Fig: Output of uttered word 'coffee'

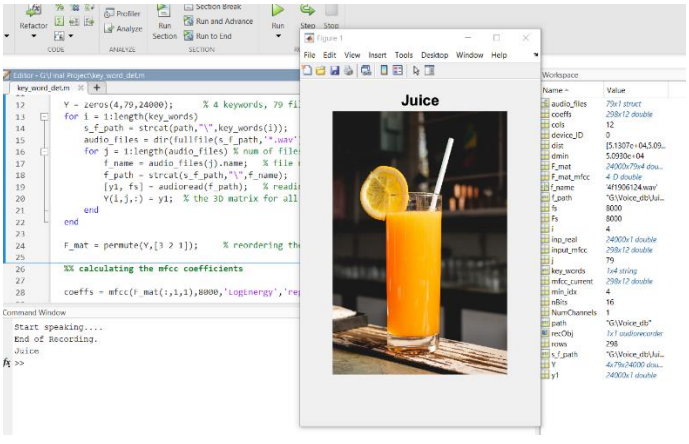

Fig: Output of uttered word 'juice'

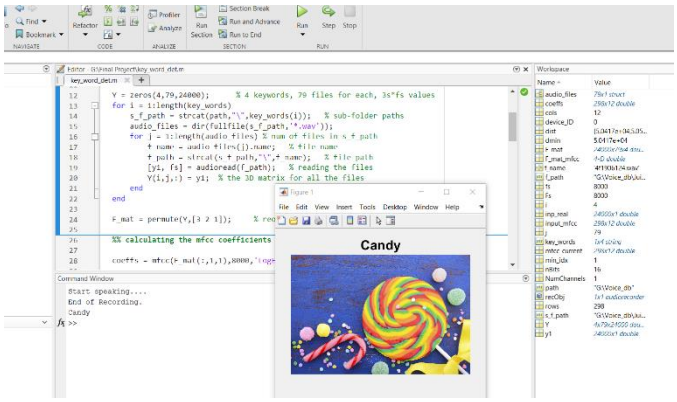

Fig: Output of uttered word ‘candy’

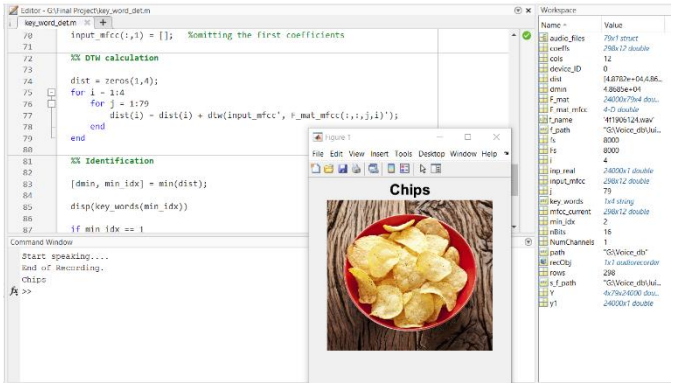

Fig: Output of uttered word ‘chips’

- **We have made a video demonstrating how does the project works**.

Drive link: https://drive.google.com/drive/folders/1x3f5xXSpWBjmhHcEfzAgwMn_ij_PjZXE

Discussions

We have successfully executed the main task and the outcome is able to find out which word of its directory has been uttered by the user. DTW Algorithm is a very useful method for such kind of challenges. The outcome of our project can be largely used in commercial sectors like establishing “Vending Machines”, “Robot-science”, “Library Assistant” etc. The accuracy of our project is more than 95% but this is not the optimum state. We had collected almost 250 sample/word We can increase the accuracy by taking more and more samples from different aged and different nationalities people.

If anyone wants to develop the concept of our project, we appreciate it.

List of References

[1] Youtube Channel: Herman Kamper, Dynamic Time Wrapping

https://www.youtube.com/@HermanKamperML

**** We express our heartiest gratitude to those who had given their voice samples to create our audio library****

Thanksgiving,

Student Id: 1906001, 1906009, 1906066, 1906067, 1906068, 1906070, 1906071, 1906072, 1906074, 1906079, 1906081, 1906084, 1906085, 1906092, 1906093, 1906096, 1906101, 1906103, 1906106, 1906109, 1906123, 1906124, 1906125, 1906142, 1906147