A SUMMER INTERNSHIP REPORT ON
MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCC) FEATURE EXTRACTION FOR SPEECH SIGNALS

SPEECH LAB

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

IIT-GUWAHATI

MAY-JUNE, 2024

SUBMITTED BY:

Anika Mariam

(B.Tech CSE)

KIIT Deemed to be University

Bhubaneswar,Odisha

Acknowledgements

My sincerest gratitude is reserved for the Unseen Power that launched me through this path of doing the internship project with the strength and patience necessary for the work.

I would also like to express my heartfelt gratitude for my guide, Prof. Pradip Kumar Das, Department of Computer Science and Engineering, IIT-Guwahati, for his invaluable guidance and mentoring which helped me in completing my two months internship project.

My thanks also go to my supervisors, Mr. Himanshu Sharma, Research Scholar at Speech Lab, Department of Computer Science and Engineering , IIT- Guwahati; Ms. Ashita Batra, Research Scholar at Speech Lab, Department of Computer Science and Engineering, IIT- Guwahati; Mrs. Komal Bharti, Research Scholar at Speech Lab, Department of Computer Science and Engineering, IIT-Guwahati and Mr. Brajesh Kar, pursuing M.Tech in Computer Science and Engineering, IIT-Guwahati, for guiding and assisting me with all the basic concepts as well as for providing the necessary resources required for the internship project.

I would also like to thank my parents who motivated and supported me a lot during the period of doing the internship project.

Finally, my thanks are also there for all other people who directly or indirectly helped me while doing my project.

Contents

Title	Page Number
1. Introduction….	4
2. Objectives……	5
3. MFCC Algorithm……	6-8
4. Code Screenshot……..	9-11
5. Output……..	12-15
6. Results and Discussions……...	16-25
7. Conclusion……	26
8. Challenges Faced……..	27
9. Solutions Implemented…….	27
10. Reference…….	28-29
11. Appendices……..	30-42

Introduction

Speech recognition is a supervised learning task[1]. In speech recognition problem, the input is an audio signal that is provided to the system for recognition. A raw audio signal input cannot be provided directly to the system as it will contain a lot of noise. It is better to extract features from the audio signal and use those extracted features as input to the base model. Upon doing so, the model will show a better performance than directly considering raw audio signal as input. The most widely used technique for extracting features from audio signal is Mel Frequency Cepstral Coefficients[1].

MFCC (Mel Frequency Cepstral Coefficients) is a prominent feature extraction technique used in audio and speech processing. It takes an audio signal as input and generates a set of coefficients that encapsulate the frequency characteristics of the audio in a way that aligns with human auditory perception. The number of MFCC coefficients produced is defined by the analysis parameters, providing a concise representation of the audio signal's characteristics. These coefficients effectively capture the spectral properties of audio, making them invaluable for machine learning applications like speech recognition and music analysis.

Objectives

The study aims to:

1. Obtain Mel Frequency Cepstral Coefficients(MFCC) in the form of matrices having ‘n’ number of rows and ‘m’ number of columns for different input audio files.
2. Use the columns of each of the matrices obtained for the audio files to calculate and plot the graph of Euclidean Distance against total number of frames for comparison.

MFCC Algorithm

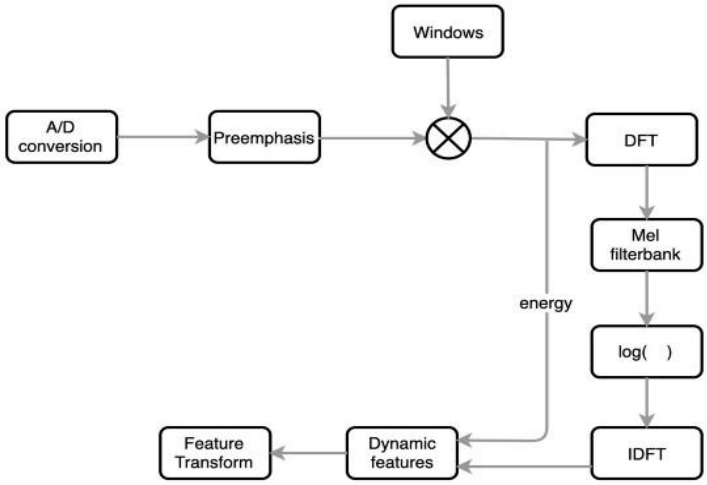


Fig: Steps involved in MFCC Algorithm[1]

1. A/D Conversion :

   The input audio signal is an analog signal. It is first converted into a digital signal with a sampling frequency of 8 kHz or 16 kHz.

2. Pre-emphasis :

   The digital signal obtained is then passed through a high-pass filter which amplifies the frequency component of the signal. Upon doing so, the energy distribution across the spectrum of the audio signal is balanced making it easier for subsequent processing steps to analyze the signal effectively. There is also a reduction in noise that makes the signal clearer and more effective. Pre-emphasis also enhances the signal-to-noise ratio, which is beneficial for more accurate speech recognition and other audio processing tasks. The formula used for pre-emphasis is as follows:

   S2(n)=S1(n)-a*S(n-1),

   Where, S1(n): input signal
   S2(n): Output signal
   a: pre-emphasis coefficient (0.9-1)

3. Windowing :

This is one of the most important step in MFCC feature extraction where an audio signal is divided into small, overlapping frames called windows. Windowing prevents spectral leakage. Hamming Window formula is popularly used for the purpose of windowing an audio signal.

$$w(n)=0.54-0.46\cos(2\pi n/N-1);\ (0)<=(n)<=(N-1)$$

Where, w(n): window value at sample n

N: total number of samples in the window

4. Discrete Fourier Transform(DFT) :

DFT is used to convert the time-domain signal into frequency-domain signal. The significance of DFT in MFCC feature extraction is that it provides a frequency spectrum that helps us to know that how much of each frequency is present in the input signal which is crucial for the further steps in MFCC feature extraction.

$$x(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{(-j2\pi n/N)kn},$$

Where, x(k): DFT of the input signal (n)

N: number of points in the DFT, which is equal to the length of the input signal frame

k: frequency bin index, ranging from 0 to N−1

x(n): input signal

j: imaginary unit

5. Mel-Filter Bank :

The power spectrum is passed through a set of filters of Mel scale that maps the actual frequency to the frequency that human ears can perceive.

$$Mel(f)=1127\ln(1+f/700)$$

6. Applying Log :

Log is applied to the output obtained using the Mel-filter formula to mimic human hearing system.

7. Inverse Discrete Fourier Transform(IDFT) :

   IDFT is applied to the log of Mel-Spectrum bank to convert the frequency-domain signal back into the time-domain signal. As a result, the cepstral coefficients are obtained that are essential in speech recognition . (Usually around 13 cepstral coefficients are taken)

8. Dynamic Features :

   Along with these 13 features, the MFCC technique will consider the first order derivative and second order derivatives of the features which constitute another 26 features.

   Derivatives are calculated by taking the difference of these coefficients between the samples of the audio signal and it will help in understanding how the transition is occurring.

The output of the above algorithm will be a set of MFCC represented in the form of a matrix having 13 rows. With the help of the MFCC matrix, graphs of Euclidean Distances against the total number of frames will be plotted with the help of the given formula:

$$\text{Euclidean Distance= |X-Y|} = \sum_{i=0}^{n} \sqrt[2]{(x_i - y_i)\text{^}2}$$

Code Screenshot

The Python code was obtained from online sources and 13 matrices were plotted for six different input audio files of the five English vowels(each audio files had samples in the range of 4500 to 5000). The code was written and compiled in Jupyter Environment. The libraries used in the code were: ‘Librosa’ for analysis of the audio files, ‘Matplotlib’ for visualization of data and ‘Numpy’ for high-level mathematical functions and arrays.

Phase I :

```
import librosa
import librosa.display
import matplotlib.pyplot as plt
import numpy as np
```

```
def extract_mfcc(audio_file,sr=None, n_mfcc=13):
    """
    Extract MFCC features from an audio file.

    Parameters:
    - audio_file: path to the audio file
    - n_mfcc: number of MFCC features to extract

    Returns:
    - mfcc: numpy array containing the MFCC features
    """
    try:
        y, sr = librosa.load(audio_file, sr=sr)  # Load audio file
        mfcc = librosa.feature.mfcc(y=y, sr=sr, n_mfcc=n_mfcc)  # Extract MFCC features
        return mfcc
    except FileNotFoundError:
        print(f"Audio file not found: {audio_file}")
        return None
```

```
def plot_mfcc(mfcc):
    """
    Plot the MFCC features.

    Parameters:
    - mfcc: numpy array containing the MFCC features
    """
    plt.figure(figsize=(4, 4))
    librosa.display.specshow(mfcc, x_axis='time')
    plt.colorbar()
    plt.title('MFCC')
    plt.tight_layout()
    plt.show()
```

```python
def calculate_euclidean_distance(column1, column2):
    """
    Calculate the Euclidean distance between two columns.

    Parameters:
    - column1: first numpy array column
    - column2: second numpy array column

    Returns:
    - distance: Euclidean distance between the columns
    """
    return np.linalg.norm(column1 - column2)

# Example usage:
audio_files = [' ']  # Replace with actual paths to your audio files

mfcc_matrices = []

for audio_file in audio_files:
    mfcc = extract_mfcc(audio_file)
    if mfcc is not None:
        print(f"MFCC features for {audio_file}:")
        print(mfcc)  # This will print the MFCC matrix
        plot_mfcc(mfcc)
        mfcc_matrices.append(mfcc)  # Append the first column of the MFCC matrix
```

```python
for i in range(mfcc.shape[1]):
        distance = calculate_euclidean_distance(mfcc[1:, 0], mfcc[1:, i])
        print(f"Euclidean distance between first column and column {i+1} of the same matrix: {distance}")
else:
    print("Failed to extract MFCC features.")
```

Phase II :

```python
import librosa
import librosa.display
import matplotlib.pyplot as plt
import numpy as np
```

```python
def extract_mfcc(audio_file,sr=None, n_mfcc=13):
    """
    Extract MFCC features from an audio file.

    Parameters:
    - audio_file: path to the audio file
    - n_mfcc: number of MFCC features to extract

    Returns:
    - mfcc: numpy array containing the MFCC features
    """
    try:
        y, sr = librosa.load(audio_file, sr=sr)  # Load audio file
        mfcc = librosa.feature.mfcc(y=y, sr=sr, n_mfcc=n_mfcc)  # Extract MFCC features
        return mfcc
    except FileNotFoundError:
        print(f"Audio file not found: {audio_file}")
        return None
```

```python
def plot_mfcc(mfcc):
    """
    Plot the MFCC features.

    Parameters:
    - mfcc: numpy array containing the MFCC features
    """
    plt.figure(figsize=(4, 4))
    librosa.display.specshow(mfcc, x_axis='time')
    plt.colorbar()
    plt.title('MFCC')
    plt.tight_layout()
    plt.show()
```

```python
def calculate_euclidean_distance(column1, column2):
    """
    Calculate the Euclidean distance between two columns.

    Parameters:
    - column1: first numpy array column
    - column2: second numpy array column

    Returns:
    - distance: Euclidean distance between the columns
    """
    return np.linalg.norm(column1 - column2)

# Example usage:
audio_files = [' ', ' ']  # Replace with actual paths to your audio files

mfcc_matrices = []

for audio_file in audio_files:
    mfcc = extract_mfcc(audio_file)
    if mfcc is not None:
        print(f"MFCC features for {audio_file}:")
        print(mfcc)  # This will print the MFCC matrix
        plot_mfcc(mfcc)
        mfcc_matrices.append(mfcc)  # Append the first column of the MFCC matrix
```

```python
if len(mfcc_matrices) == 2:
    for i in range(min(mfcc_matrices[0].shape[1], mfcc_matrices[1].shape[1])):
        column1 = mfcc_matrices[0][1:, i]  # Exclude the first element
        column2 = mfcc_matrices[1][1:, i]  # Exclude the first element
        distance = calculate_euclidean_distance(column1, column2)
        print(f"Euclidean distance between column {i+1} of the first matrix and column {i+1} of the second matrix: {distance}
else:
    print("Failed to extract MFCC features.")
```

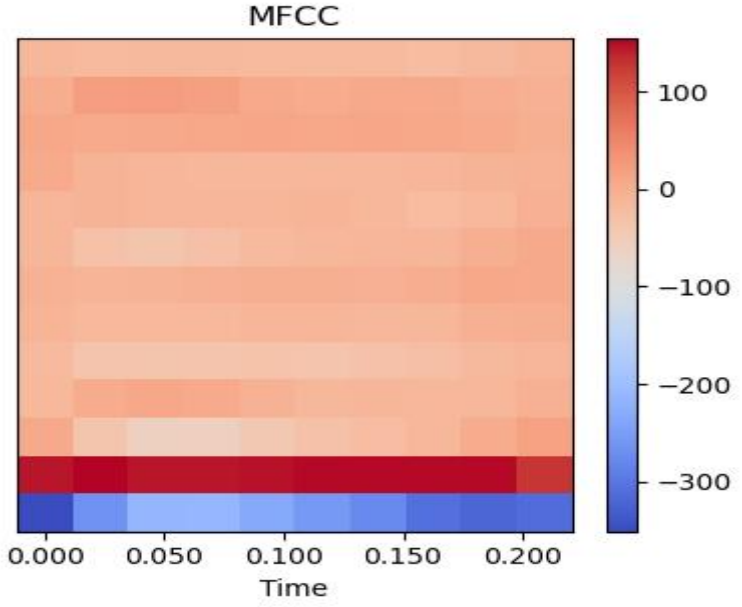
Output

PHASE I:

Vowel |a| :

MFCC features for C:\IIT-G Internship\ASCII Waves Recording\234101011_a_11(2)-2.wav:

```
[[-351.00757   -267.04715   -213.45972   -211.7793    -234.43544
  -256.72998   -277.63052   -308.67694   -319.8311    -312.05426  ]
 [ 146.03682    154.03574    145.97983    145.63773    147.69489
   150.95065    150.5588     151.30782    150.57013    126.746704 ]
 [   6.092407   -35.51602    -59.464455   -57.227127   -42.361214
   -29.684784   -19.752018   -12.907835     3.8814592   16.698416 ]
 [ -14.169696     3.638988     8.18647      4.651394    -4.6651945
   -12.31867    -11.24423    -12.123945   -12.138529    -2.1274855 ]
 [ -18.280289   -33.66287    -36.25567    -33.876747   -33.135895
   -33.551727   -30.873142   -24.729717   -17.103405   -10.319367  ]
 [  -9.357389   -13.701714   -15.181549   -14.764006   -11.554573
    -9.762973   -13.44471    -12.12282     -3.6347733   -0.88035524]
 [  -4.485645    -8.859351    -7.415781    -2.751775    -0.9473416
    -0.92675334  -2.6860862    1.2148435    8.40486      6.176793  ]
 [ -11.215554   -30.257336   -36.328255   -28.979748   -18.66191
   -13.073076   -11.20605    -10.066433    -0.35809547   6.5718055 ]
 [  -9.80196     -7.5694876  -10.092711   -10.673542    -9.716137
    -9.359627   -11.735107   -21.278118   -14.871116    -2.44669   ]
 [   5.5019083   -6.805078   -10.635617   -12.30052    -12.995281
   -12.916488   -13.10021    -11.492619    -6.8954062   -5.2659607 ]
 [   9.028929     5.619988     7.1044736    8.174194    10.10865
    9.862011    11.450119     9.328226     5.8070135   -1.8140237 ]
 [   0.75368744  20.640318    22.453506    19.0405      7.842724
    3.692804     7.473935     7.5740213    1.7157669   -3.2051663 ]
 [ -12.643644   -18.89718    -17.196636   -17.005      -19.208546
   -18.203127   -18.697468   -22.201405   -15.7650385   -8.03163   ]]
```



Euclidean distance between first column and column 1 of the same matrix: 0.0

Euclidean distance between first column and column 2 of the same matrix: 57.89870834350586

Euclidean distance between first column and column 3 of the same matrix: 80.92955017089844

Euclidean distance between first column and column 4 of the same matrix: 75.01103973388672

Euclidean distance between first column and column 5 of the same matrix: 56.311580657958984

Euclidean distance between first column and column 6 of the same matrix: 44.008522033691406

Euclidean distance between first column and column 7 of the same matrix: 36.23415756225586

Euclidean distance between first column and column 8 of the same matrix: 32.16379928588867

Euclidean distance between first column and column 9 of the same matrix: 23.41339111328125

Euclidean distance between first column and column 10 of the same matrix: 38.980567932128906

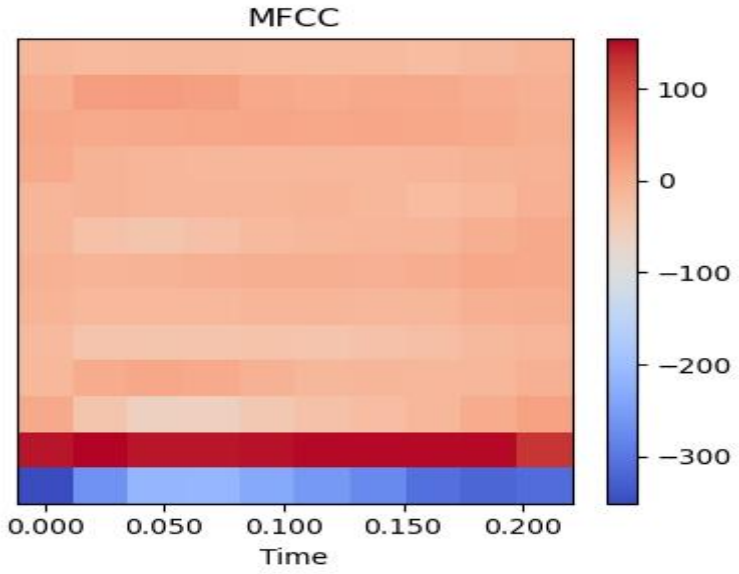
The rest of the outputs are presented in 'Appendix-A Section'.

PHASE II:

Vowel |a| and |o| :

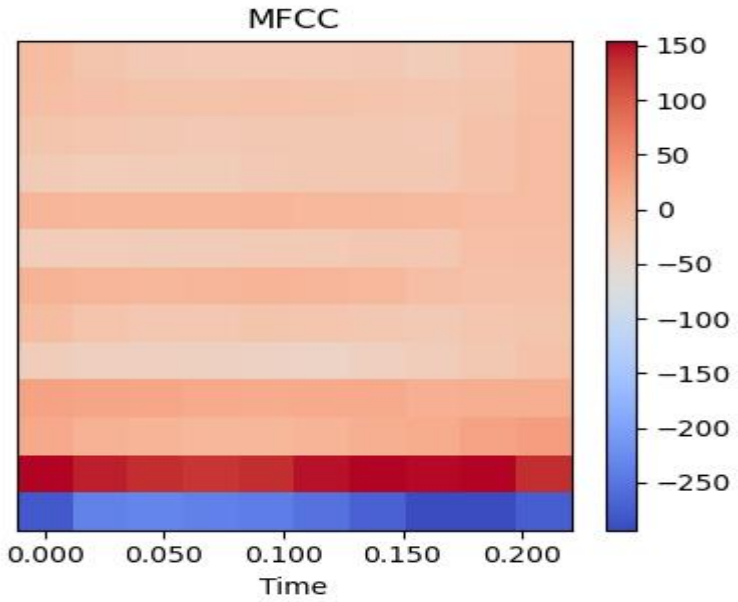
MFCC features for C:\IIT-G Internship\ASCII Waves Recording\Used\234101011_a_11(2)-2.wav:

```
[[-351.00757   -267.04715   -213.45972   -211.7793    -234.43544
  -256.72998   -277.63052   -308.67694   -319.8311    -312.05426  ]
 [ 146.03682    154.03574    145.97983    145.63773    147.69489
   150.95065    150.5588     151.30782    150.57013    126.746704 ]
 [   6.092407   -35.51602    -59.464455   -57.227127   -42.361214
   -29.684784   -19.752018   -12.907835     3.8814592   16.698416 ]
 [ -14.169696     3.638988     8.18647      4.651394    -4.6651945
   -12.31867    -11.24423    -12.123945   -12.138529    -2.1274855 ]
 [ -18.280289   -33.66287    -36.25567    -33.876747   -33.135895
   -33.551727   -30.873142   -24.729717   -17.103405   -10.319367  ]
 [  -9.357389   -13.701714   -15.181549   -14.764006   -11.554573
    -9.762973   -13.44471    -12.12282     -3.6347733   -0.88035524]
 [  -4.485645    -8.859351    -7.415781    -2.751775    -0.9473416
    -0.92675334  -2.6860862    1.2148435    8.40486      6.176793  ]
 [ -11.215554   -30.257336   -36.328255   -28.979748   -18.66191
   -13.073076   -11.20605    -10.066433    -0.35809547   6.5718055 ]
 [  -9.80196    -7.5694876  -10.092711   -10.673542    -9.716137
    -9.359627   -11.735107   -21.278118   -14.871116    -2.44669   ]
 [   5.5019083   -6.805078   -10.635617   -12.30052    -12.995281
   -12.916488   -13.10021    -11.492619    -6.8954062   -5.2659607 ]
 [   9.028929     5.619988     7.1044736    8.174194    10.10865
    9.862011    11.450119     9.328226     5.8070135   -1.8140237 ]
 [   0.75368744  20.640318    22.453506    19.0405      7.842724
    3.692804     7.473935     7.5740213    1.7157669   -3.2051663 ]
 [ -12.643644   -18.89718    -17.196636   -17.005      -19.208546
   -18.203127   -18.697468   -22.201405   -15.7650385   -8.03163   ]]
```



MFCC features for C:\IIT-G Internship\ASCII Waves Recording\Used\234101011_o_11-5.wav:

```
[[-2.77891052e+02 -2.36564941e+02 -2.33022842e+02 -2.37977554e+02
  -2.42510239e+02 -2.54679947e+02 -2.72556335e+02 -2.93521423e+02
  -2.91807556e+02 -2.75917877e+02]
 [ 1.52155869e+02  1.43099854e+02  1.34740280e+02  1.30787125e+02
   1.34744476e+02  1.47401703e+02  1.52282806e+02  1.50421295e+02
   1.53620316e+02  1.35219696e+02]
 [ 2.35752449e+01  1.28966541e+01  8.82866859e+00  6.66573143e+00
   5.87618494e+00  9.40012741e+00  1.67778778e+01  2.00155582e+01
   3.09033737e+01  3.74443970e+01]
 [ 2.98087425e+01  2.81266117e+01  2.61585770e+01  2.22503357e+01
   1.99513168e+01  2.22090607e+01  2.13529320e+01  1.53626375e+01
   1.70073013e+01  1.64527683e+01]
 [-2.99601059e+01 -3.39167595e+01 -3.34620209e+01 -3.57017517e+01
  -3.88289680e+01 -4.04606476e+01 -3.47825851e+01 -3.05867653e+01
  -1.93724060e+01 -9.91511631e+00]
 [-3.34251213e+00 -1.23432436e+01 -1.86077499e+01 -1.77866402e+01
  -1.43893061e+01 -1.58622351e+01 -1.96974220e+01 -2.13809090e+01
  -1.72866096e+01 -1.45359364e+01]
 [ 1.06158562e+01  7.23377132e+00  5.89317703e+00  7.47908020e+00
   8.95137215e+00  7.97829485e+00  4.11840820e+00 -5.08348560e+00
  -9.00300980e+00 -9.68267059e+00]
 [-2.94869499e+01 -2.95495224e+01 -3.05528030e+01 -2.66225090e+01
  -2.39171410e+01 -2.28534355e+01 -2.03238373e+01 -1.75650368e+01
  -6.11445522e+00 -3.81086445e+00]
 [ 7.94421959e+00  6.28949594e+00  5.37427807e+00  6.44421291e+00
   7.58687687e+00  4.89404964e+00  2.39408898e+00  7.33923316e-02
  -3.48442721e+00 -2.18477964e+00]
 [-2.47457962e+01 -2.84333553e+01 -2.72338047e+01 -2.62895794e+01
  -2.20281258e+01 -2.01934013e+01 -1.98237667e+01 -1.96356506e+01
  -1.04977579e+01 -2.28154969e+00]
 [-1.45225563e+01 -1.68399105e+01 -1.95427017e+01 -2.12964611e+01
  -1.99442348e+01 -1.95838089e+01 -2.06745377e+01 -2.13720016e+01
  -1.00486555e+01 -2.20870042e+00]
 [-4.52103806e+00 -7.10537148e+00 -1.08751869e+01 -1.08146305e+01
  -1.04984360e+01 -1.22525082e+01 -1.24595814e+01 -1.74234447e+01
  -1.46701221e+01 -5.18068218e+00]
 [-2.86115384e+00 -1.45820847e+01 -2.24075203e+01 -2.28117294e+01
  -2.28710766e+01 -2.36260681e+01 -2.19751015e+01 -2.72275772e+01
  -1.84330006e+01 -5.99584389e+00]]
```



Euclidean distance between column 1 of the first matrix and column 1 of the second matrix: 70.14081573486328
Euclidean distance between column 2 of the first matrix and column 2 of the second matrix: 72.64411926269531
Euclidean distance between column 3 of the first matrix and column 3 of the second matrix: 87.78414154052734
Euclidean distance between column 4 of the first matrix and column 4 of the second matrix: 83.77072143554688
Euclidean distance between column 5 of the first matrix and column 5 of the second matrix: 70.00396728515625
Euclidean distance between column 6 of the first matrix and column 6 of the second matrix: 66.32584381103516
Euclidean distance between column 7 of the first matrix and column 7 of the second matrix: 65.33541870117188
Euclidean distance between column 8 of the first matrix and column 8 of the second matrix: 64.5832748413086
Euclidean distance between column 9 of the first matrix and column 9 of the second matrix: 52.78549575805664
Euclidean distance between column 10 of the first matrix and column 10 of the second matrix: 37.557315826416016

The rest of the outputs are presented in ‘Appendix-B Section’.

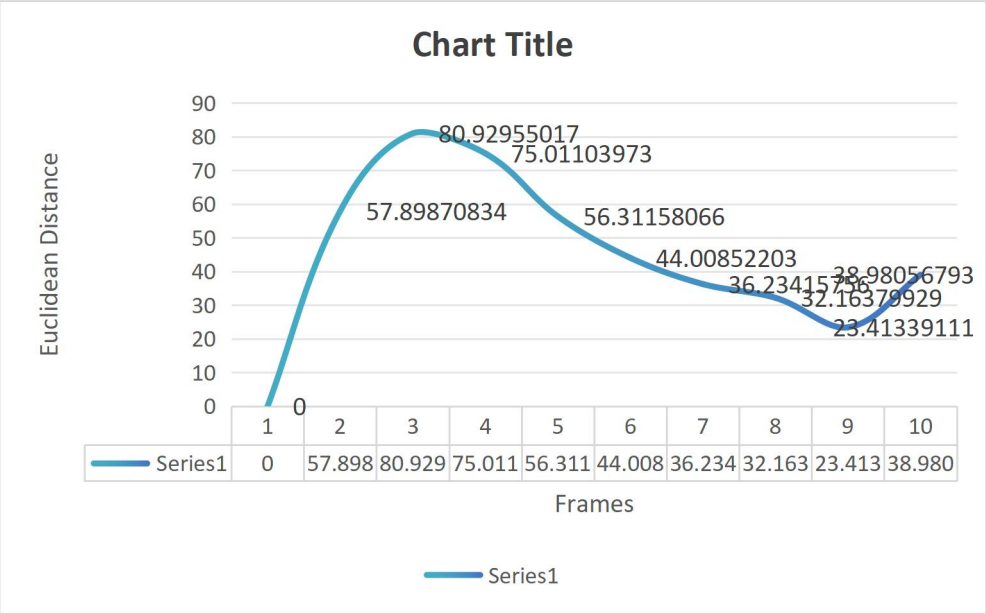
Results and Discussions

Ten graphs are obtained for six different audio files of the five English vowels for comparison. The Euclidean Distance for the five English vowels is plotted against the total number of frames(columns of the matrices). The total number of frames taken for each of the English vowel is 10. The results are obtained in 'Phase I' and 'Phase II'.

PHASE I:

x-axis: Number of Frames

y-axis: Euclidean Distance

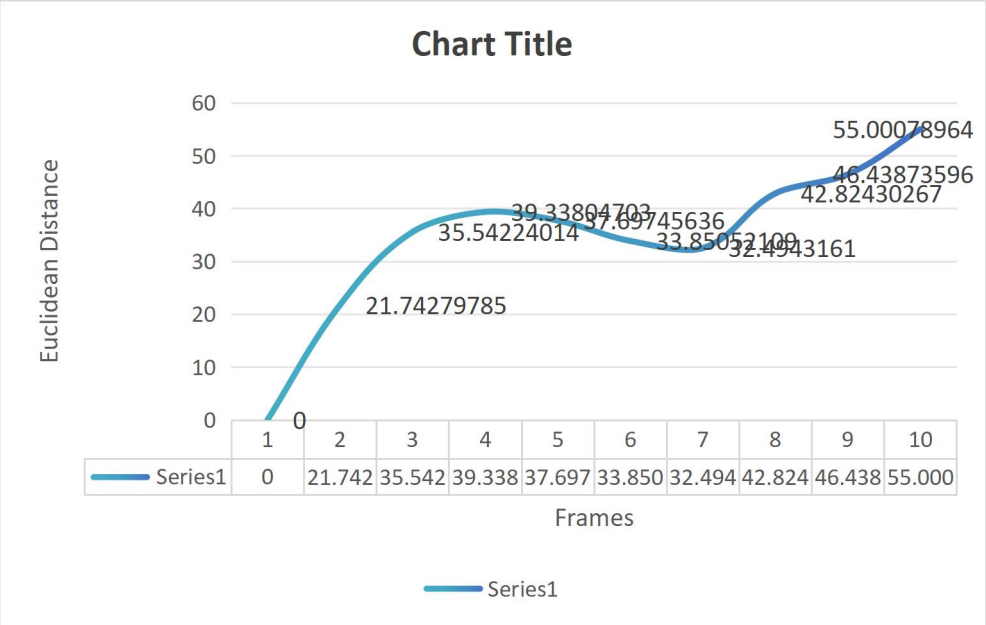


Graph 1: Euclidean Distance of Vowel |a|

Analysis of Graph 1:

1. The smallest distance in the graph after the first frame is in frame 9; the distance is 23.413. This indicates the similarity among the spectral features of frame 1 and frame 9.

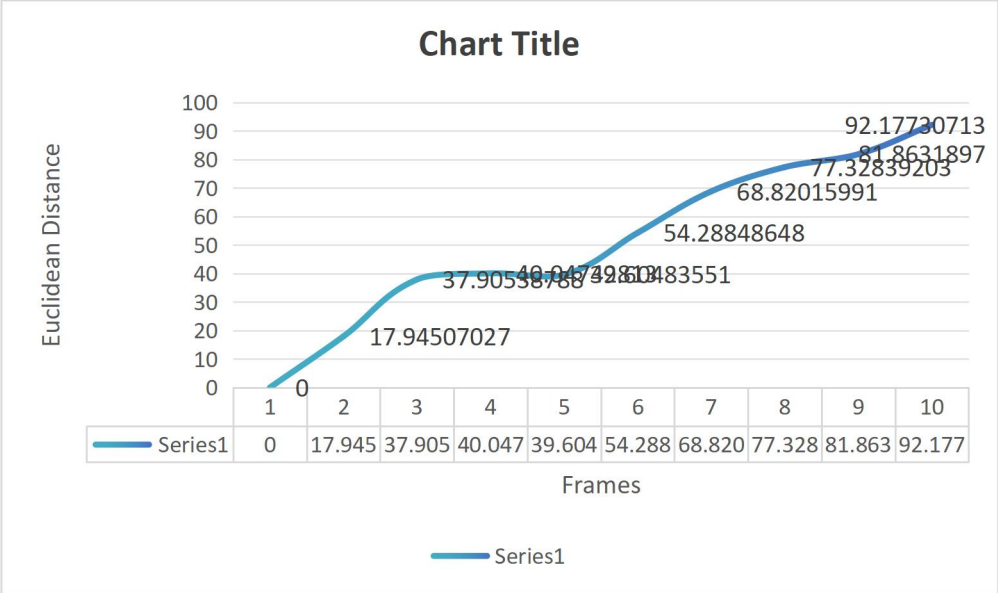
2. The distance in the graph generally increases from frame 1 to frame 3. The distance in frame 3 is 80.929, that is the peak value. After reaching the peak value, the distance is fluctuating. This indicates a gradual change in the spectral features, possibly due to the dynamic changes of the vowel that might have occurred during recording.
3. A stable phase can be observed from frame 6 to frame 8 indicating that the spectral features are consistent and hence the pronunciation is stable.



Graph 2: Euclidean Distance of Vowel |o|

Analysis of Graph 2 :

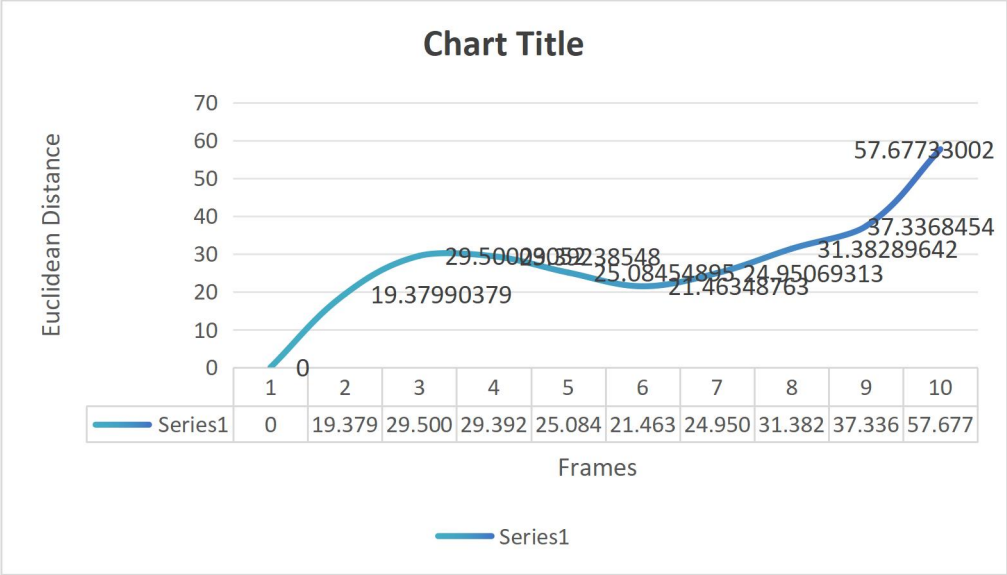
1. The smallest distance in the graph after the first frame is in frame 2; the distance is 21.742. This indicates the similarity among the spectral features of frame 1 and frame 2.
2. The distance increases from frame 1 to frame 4 and starts fluctuating from frame 5. This indicates a gradual change in the spectral features, possibly due to the dynamic changes of the vowel that might have occurred during recording.
3. The graph reaches its peak value at frame 5 where the distance is 55.000 indicating greater changes in spectral features as time progresses.



Graph 3: Euclidean Distance of Vowel |e|

Analysis of Graph 3 :

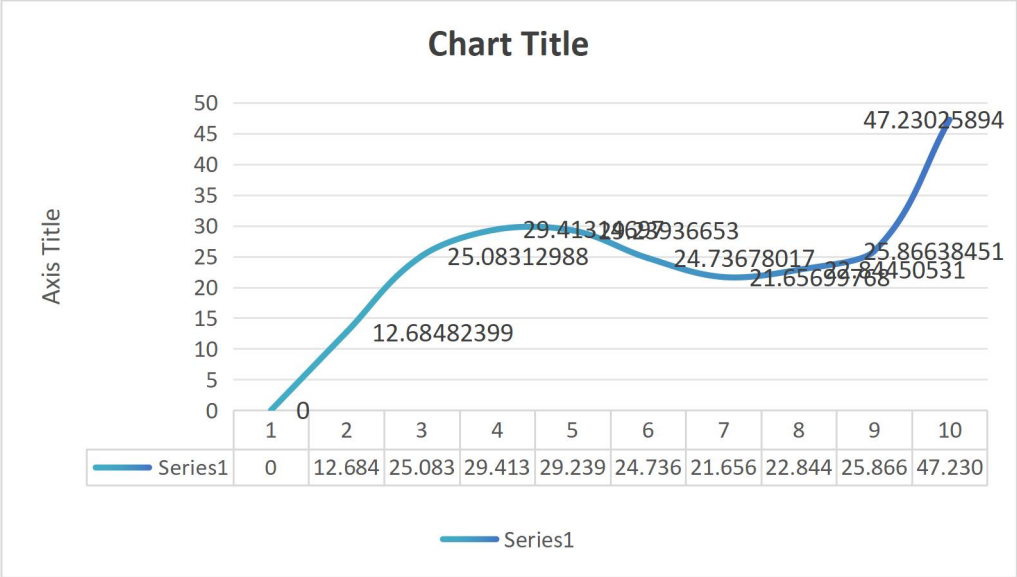
1. The smallest distance in the graph after the first frame is in frame 2; the distance is 17.945. This indicates the similarity among the spectral features of frame 1 and frame 2.
2. The distance increases from frame 1 to frame 4 and decreases in frame 5. This indicates a gradual change in the spectral features, possibly due to the dynamic changes of the vowel that might have occurred during recording.
3. The distance again increases from frame 6 and reaches the peak value at frame 10 where the distance is 92.177. This indicates greater changes in spectral features as time progresses.



Graph 4: Euclidean Distance of Vowel |i|

Analysis of Graph 4 :

1. The smallest distance in the graph after the first frame is in frame 2; the distance is 19.379. This indicates the similarity among the spectral features of frame 1 and frame 2.
2. The distance increases from frame 1 to frame 4 and starts fluctuating from frame 5. This indicates a gradual change in the spectral features, possibly due to the dynamic changes of the vowel that might have occurred during recording.
3. The graph reaches its peak value at frame 10 where the distance is 57.66. This indicates greater changes in spectral features as time progresses.



Graph 5: Euclidean Distance of Vowel |u|

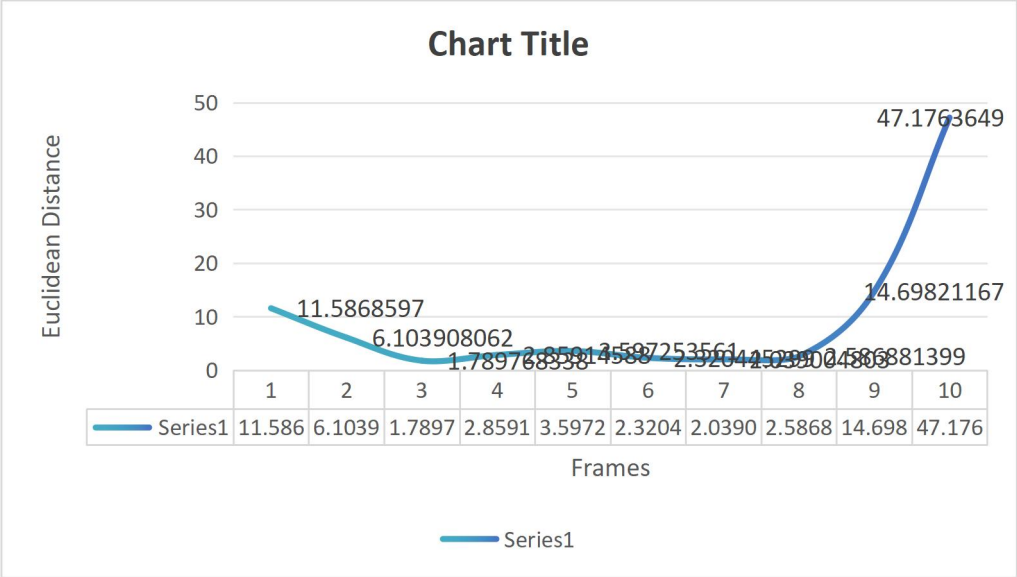
Analysis of Graph 5 :

1. The smallest distance in the graph after the first frame is in frame 2; the distance is 12.684. This indicates the similarity among the spectral features of frame 1 and frame 2.
2. The distance increases from frame 1 to frame 4 and starts fluctuating from frame 5. This indicates a gradual change in the spectral features, possibly due to the dynamic changes of the vowel that might have occurred during recording.
3. The graph reaches its peak value at frame 10 where the distance is 47.230. This indicates greater changes in spectral features as time progresses.

PHASE II:

x-axis: Euclidean Distance

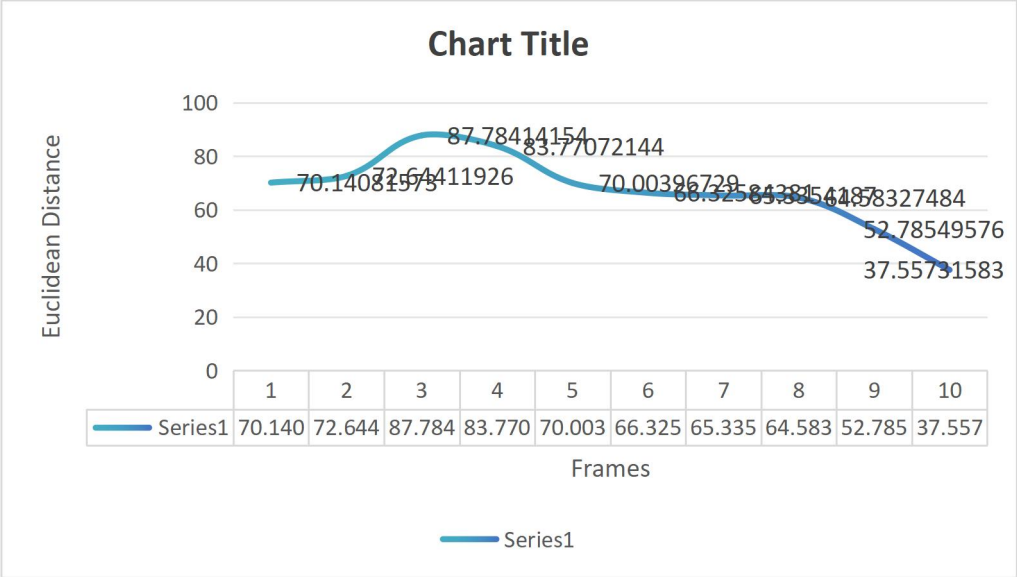
y-axis: Number of Frames



Graph 1: Euclidean Distance between Vowel |a| and |a|

Analysis of Graph 2 :

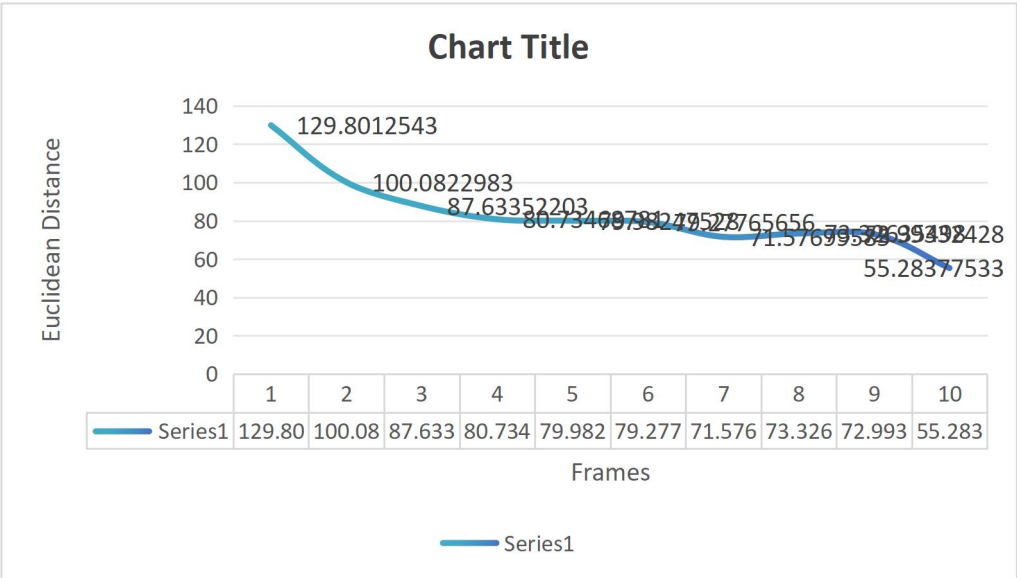
1. The distance in the first frame is 11.586 which indicates moderate difference between the two vowels |a| and |a|.This might be due to variations in the articulatory movements required to produce or pronounce the vowels.
2. The graph reaches its peak at frame 10 where the distance is 47.176 indicating greater distinction in the spectral features of the two vowels.
3. The lowest distance is observed in frame 7 where the distance is 2.0390 indicating greater similarities among the spectral features of the two vowels.



Graph 2: Euclidean Distance between Vowel |a| and |o|

Analysis of Graph 2 :

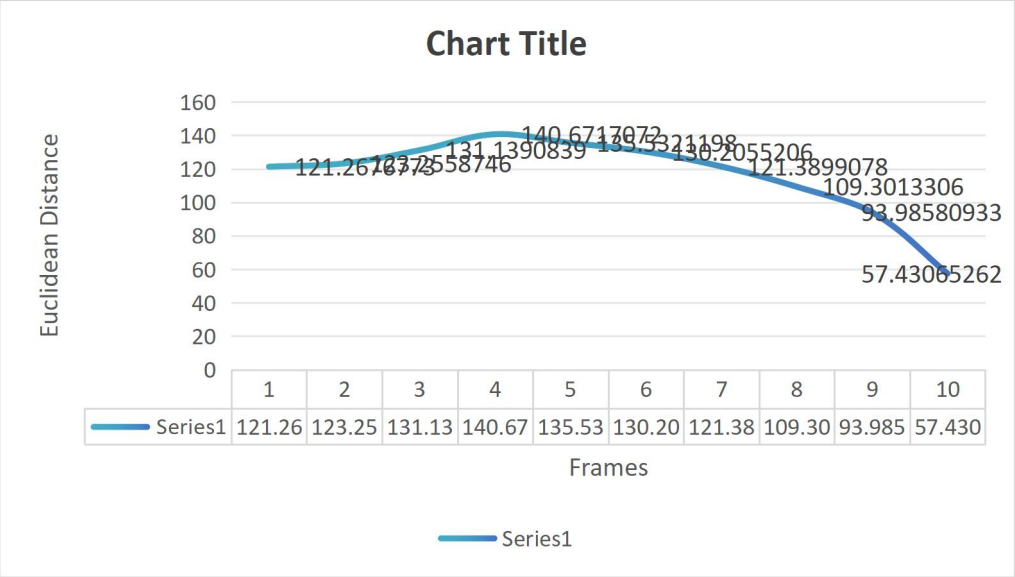
4. The distance in the first frame is 70.140 which indicates moderate difference between the two vowels |a| and |o|.This might be due to variations in the articulatory movements required to produce or pronounce the vowels.
5. The graph reaches its peak at frame 3 where the distance is 87.784 indicating greater distinction among the spectral features of the two vowels.
6. The lowest distance is observed in frame 10 where the distance is 37.557 indicating greater similarities among the spectral features of the two vowels.



Graph 3: Euclidean Distance between Vowel |a| and |e|

Analysis of Graph 3 :

1. The distance in the first frame is 129.80 which is the peak value indicating greater difference between the spectral features of the two vowels |a| and |e|.This might be due to variations in the articulatory movements required to produce or pronounce the vowels.
2. The distance decreases from frame 1 onwards up-to frame 7 and starts fluctuating from frame 8. This indicates a gradual change in the spectral features, possibly due to the dynamic changes of the vowel that might have occurred during recording.
3. The lowest distance is observed in frame 10 where the distance is 55.283 indicating greater similarities among the spectral features of the two vowels.

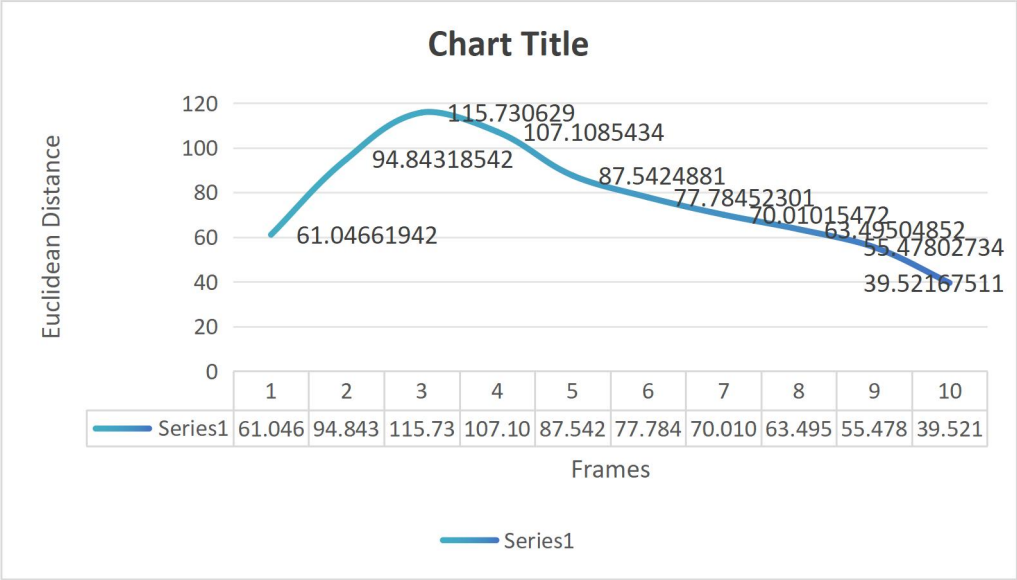


Graph 4: Euclidean Distance between Vowel |a| and |i|

Analysis of Graph 4 :

1. The distance in the first frame is 121.26 which indicates moderate difference between the two vowels |a| and |i|.This might be due to variations in the articulatory movements required to produce or pronounce the vowels. Also, the distance continues to increases from frame 1 to frame 4 and decreases from frame 5 onwards up-to frame 10. This indicates a gradual change in the spectral features, possibly due to the dynamic changes of the vowel that might have occurred during recording.
2. The graph reaches its peak at frame 4 where the distance is 140.67 indicating greater distinction among the spectral features of the two vowels.

3. The lowest distance is observed at frame 10 where the distance is 57.430 indicating greater similarities among the spectral features of the two vowels.



Graph 5: Euclidean Distance between Vowel |a| and |u|

Analysis of Graph 5 :

1. The distance in the first frame is 61.046 which indicates moderate difference between the two vowels |a| and |u|.This might be due to variations in the articulatory movements required to produce or pronounce the vowels.Also, the distance increases from frame 1 onwards up-to frame 3 and continues to decrease from frame 4 to frame 10. This indicates a gradual change in the spectral features, possibly due to the dynamic changes of the vowel that might have occurred during recording
2. The graph reaches its peak at frame 3 where the distance is 115.73 indicating greater distinction between the spectral features of the two vowels.
3. The lowest distance is observed in frame 10 which is 39.521 indicating greater similarities among the spectral features of the two vowels.

Therefore, the analysis of Euclidean distances in MFCC matrices for the five different English vowels in 'Phase I' helps in the better understanding of the dynamic and phonetic features of the different English vowel sounds. In addition, the analysis of Euclidean distances among different columns of MFCC matrices of five different English vowels in 'Phase II' highlights both the distinct and overlapping spectral features of the vowels, which helps in the better understanding of the phonetic and acoustic features of the vowel sounds.

Hence, all these variations present in the dynamic, phonetic and acoustic features of the five different English vowels play an important role in speech recognition and speaker identification.

Conclusion

To conclude, the Python code for MFCC Feature Extraction was obtained from online sources. Using that code, 13 matrices were plotted for 6 different audio files of the five English vowels. With the help of the matrices, 10 graphs of Euclidean Distance against the number of frames were plotted to analyse the similarities and distinctions among the spectral features of the five English vowels.

Challenges Faced

One of the major challenges faced while doing the project was the lack of resources. There were only few sites and tutorial videos that would provide a clear explanation to the concept of MFCC Feature extraction. Initially it was planned that the codes would be written in C programming language but later the idea was dropped due to the complexity of C and its time consuming nature.

Solutions Implemented

Due to the lack of resources and complexity in the C programming language, Python was used to calculate the MFCC for the input audio files. The reason behind using Python was that it is an easy programming language with in-built libraries that helps to solve any complex problem in a short period of time. Hence, the Python code of MFCC Feature Extraction was obtained from relevant online sources and with the help of the code, the MFCC matrices for different input audio files were plotted.

References

Articles:

1. Kiran, U. (2023, August 14). MFCC Technique for Speech Recognition. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/06/mfcc-technique-for-speech-recognition/#

2. Fayek, H. (2016, April 21). Speech Processing for Machine Learning: Filter banks, Mel-Frequency Cepstral Coefficients (MFCCs) and What's In-Between. Haytham Fayek. https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html

3. Tanuwidjaja, O. (2022, January 22). Get to know audio feature extraction in Python - towards data science. Medium. https://towardsdatascience.com/get-to-know-audio-feature-extraction-in-python-a499fdaefe42

Research Papers:

1. Shehzen, M. (2022b, January 6). How to generate MFCC from Audio. — ML for Lazy 2021. Medium. https://medium.com/analytics-vidhya/how-to-generate-mfcc-from-audio-ml-for-lazy-2021-42c2fdfa208

Websites:

1. Hossan, Md & Memon, Sheeraz & Gregory, Mark. (2011). A novel approach for MFCC feature extraction. 1 - 5. 10.1109/ICSPCS.2010.5709752.

2.Issues importing pandas tool scatter_matrix. (n.d.). Stack Overflow. https://stackoverflow.com/questions/49191740/issues-importing-pandas-tool-scatter-matrix

3. pandas.read_csv FileNotFoundError: File b'\xe2\x80\xaa' despite correct path. (n.d.). Stack Overflow. https://stackoverflow.com/questions/42165649/pandas-read-csv-filenotfounderror-file-b-xe2-x80-xaaetc-despite-correct-pat

4. Python Data Analysis with Pandas and Matplotlib. (n.d.-b). https://ourcodingclub.github.io/tutorials/pandas-python-intro/

5. FileNotFoundError what locating the CSV. (n.d.). 365 Data Science. https://365datascience.com/question/filenotfounderror-what-locating-the-csv

7.GeeksforGeeks. (2024, February 29). How to plot MFCC in Python using Matplotlib? GeeksforGeeks. https://www.geeksforgeeks.org/how-to-plot-mfcc-in-python-using-matplotlib/

8. librosa.feature.mfcc — librosa 0.10.2.post1 documentation. (n.d.). https://librosa.org/doc/main/generated/librosa.feature.mfcc.html

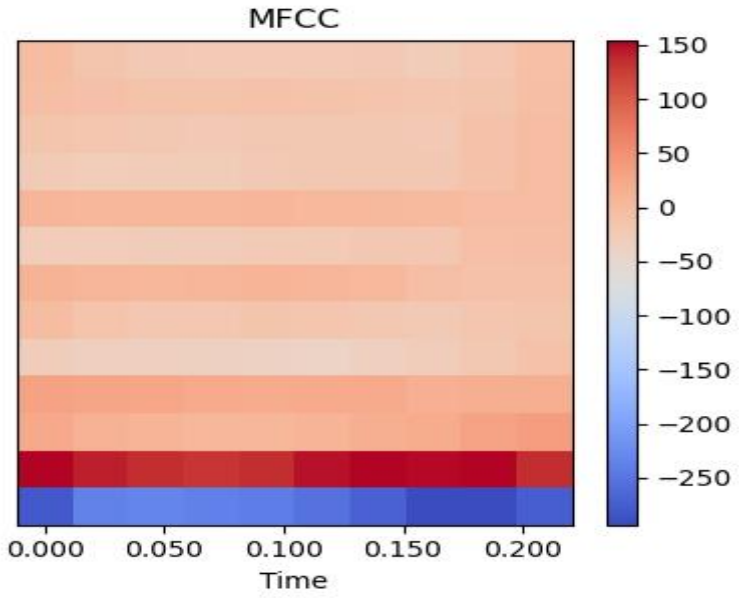
9.Audio comparison using MFCC and DTW. (n.d.). TestDevLab Blog. https://www.testdevlab.com/blog/audio-comparison-using-mfcc-and-dtw

Appendix-A

Vowel |o| :

MFCC features for C:\IIT-G Internship\ASCII Waves Recording\234101011_o_11-5.wav:

```
[[-2.77891052e+02 -2.36564941e+02 -2.33022842e+02 -2.37977554e+02
  -2.42510239e+02 -2.54679947e+02 -2.72556335e+02 -2.93521423e+02
  -2.91807556e+02 -2.75917877e+02]
 [ 1.52155869e+02  1.43099854e+02  1.34740280e+02  1.30787125e+02
   1.34744476e+02  1.47401703e+02  1.52282806e+02  1.50421295e+02
   1.53620316e+02  1.35219696e+02]
 [ 2.35752449e+01  1.28966541e+01  8.82866859e+00  6.66573143e+00
   5.87618494e+00  9.40012741e+00  1.67778778e+01  2.00155582e+01
   3.09033737e+01  3.74443970e+01]
 [ 2.98087425e+01  2.81266117e+01  2.61585770e+01  2.22503357e+01
   1.99513168e+01  2.22090607e+01  2.13529320e+01  1.53626375e+01
   1.70073013e+01  1.64527683e+01]
 [-2.99601059e+01 -3.39167595e+01 -3.34620209e+01 -3.57017517e+01
  -3.88289680e+01 -4.04606476e+01 -3.47825851e+01 -3.05867653e+01
  -1.93724060e+01 -9.91511631e+00]
 [-3.34251213e+00 -1.23432436e+01 -1.86077499e+01 -1.77866402e+01
  -1.43893061e+01 -1.58622351e+01 -1.96974220e+01 -2.13809090e+01
  -1.72866096e+01 -1.45359364e+01]
 [ 1.06158562e+01  7.23377132e+00  5.89317703e+00  7.47908020e+00
   8.95137215e+00  7.97829485e+00  4.11840820e+00 -5.08348560e+00
  -9.00300980e+00 -9.68267059e+00]
 [-2.94869499e+01 -2.95495224e+01 -3.05528030e+01 -2.66225090e+01
  -2.39171410e+01 -2.28534355e+01 -2.03238373e+01 -1.75650368e+01
  -6.11445522e+00 -3.81086445e+00]
 [ 7.94421959e+00  6.28949594e+00  5.37427807e+00  6.44421291e+00
   7.58687687e+00  4.89404964e+00  2.39408898e+00  7.33923316e-02
  -3.48442721e+00 -2.18477964e+00]
 [-2.47457962e+01 -2.84333553e+01 -2.72338047e+01 -2.62895794e+01
  -2.20281258e+01 -2.01934013e+01 -1.98237667e+01 -1.96356506e+01
  -1.04977579e+01 -2.28154969e+00]
 [-1.45225563e+01 -1.68399105e+01 -1.95427017e+01 -2.12964611e+01
  -1.99442348e+01 -1.95838089e+01 -2.06745377e+01 -2.13720016e+01
  -1.00486555e+01 -2.20870042e+00]
 [-4.52103806e+00 -7.10537148e+00 -1.08751869e+01 -1.08146305e+01
  -1.04984360e+01 -1.22525082e+01 -1.24595814e+01 -1.74234447e+01
  -1.46701221e+01 -5.18068218e+00]
 [-2.86115384e+00 -1.45820847e+01 -2.24075203e+01 -2.28117294e+01
  -2.28710766e+01 -2.36260681e+01 -2.19751015e+01 -2.72275772e+01
  -1.84330006e+01 -5.99584389e+00]]
```



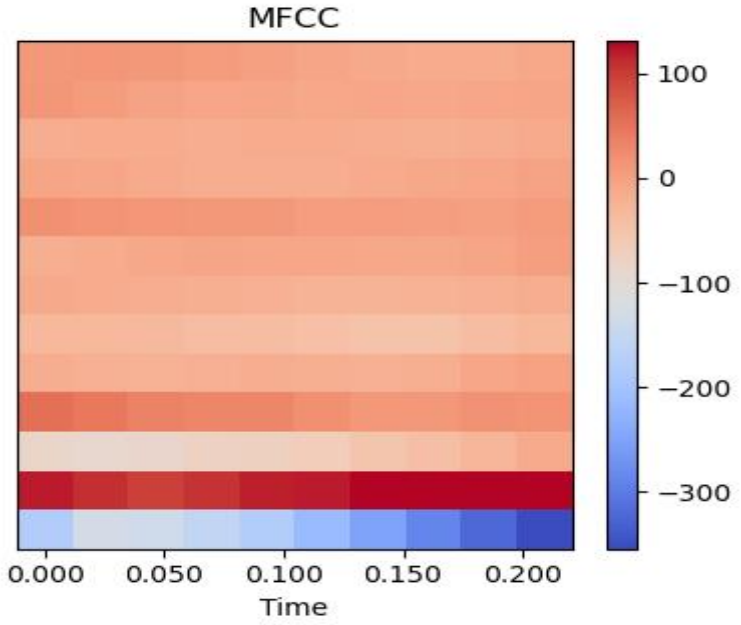
Euclidean distance between first column and column 1 of the same matrix: 0.0
Euclidean distance between first column and column 2 of the same matrix: 21.7427978515625
Euclidean distance between first column and column 3 of the same matrix: 35.542240142822266
Euclidean distance between first column and column 4 of the same matrix: 39.33804702758789
Euclidean distance between first column and column 5 of the same matrix: 37.69745635986328
Euclidean distance between first column and column 6 of the same matrix: 33.850521087646484
Euclidean distance between first column and column 7 of the same matrix: 32.49431610107422
Euclidean distance between first column and column 8 of the same matrix: 42.824302673339844
Euclidean distance between first column and column 9 of the same matrix: 46.43873596191406
Euclidean distance between first column and column 10 of the same matrix: 55.000789642333984

Vowel |e| :

MFCC features for C:\IIT-G Internship\ASCII Waves Recording\234101011_e_10-4.wav:

```
[[-181.66394   -130.52516   -135.31253   -155.55383   -179.64703
  -215.27823   -250.7267    -290.00143   -323.54022   -354.83057  ]
 [ 121.084656   109.21268    98.38544    107.57034    119.22588
   121.07956    129.7507    131.07277    130.14883    130.9653   ]
 [ -86.94835    -91.757645   -88.11727    -77.68785    -74.905045
   -66.36754    -53.01963    -40.72782    -27.756775   -12.872986 ]
 [  54.10563    46.83529     34.447002    31.470982    31.406414
    20.138264   10.706554    11.020617    17.392914    17.058363 ]
 [ -16.464523   -19.666431   -21.568855   -18.957024   -16.043291
   -17.317417   -18.976555   -18.835724    -6.4380627   -0.41792202]
 [ -31.748924   -30.66806    -33.865395   -38.07168    -39.295765
   -42.02412    -47.828484   -49.105385   -39.4506     -31.790142 ]
 [ -11.257023   -14.310413   -16.88364    -18.076218   -19.845913
   -23.40499    -24.25985    -22.60356    -20.33097    -16.90108  ]
 [ -18.19958    -13.214086    -7.525191    -4.611625    -5.9021416
    -7.160363    -8.890835    -7.9082985   -4.662279     2.8612022 ]
 [  20.02372    17.059067    12.948431    11.424824    10.4334955
     4.9399076    5.6345243    3.4401748    1.1876383    6.9955883 ]
 [  -4.476014   -7.4480414   -12.416617   -15.401159   -16.186893
   -16.185116   -12.78858     -8.78071    -6.1768656   -2.1935005 ]
 [ -16.079887   -14.381432   -14.696789   -16.257153   -14.613634
   -14.691807   -15.298426   -17.293304   -16.088732   -11.662342  ]
 [  12.081397     5.9020243   -3.4584785   -6.3440585   -5.4189157
    -7.8524446   -7.3111124   -9.153239    -5.9080706   -3.9139862 ]
```

[11.030154 11.926878 9.987385 7.435226 1.8032324
-4.143328 -10.255065 -13.460906 -14.683739 -7.841857]]



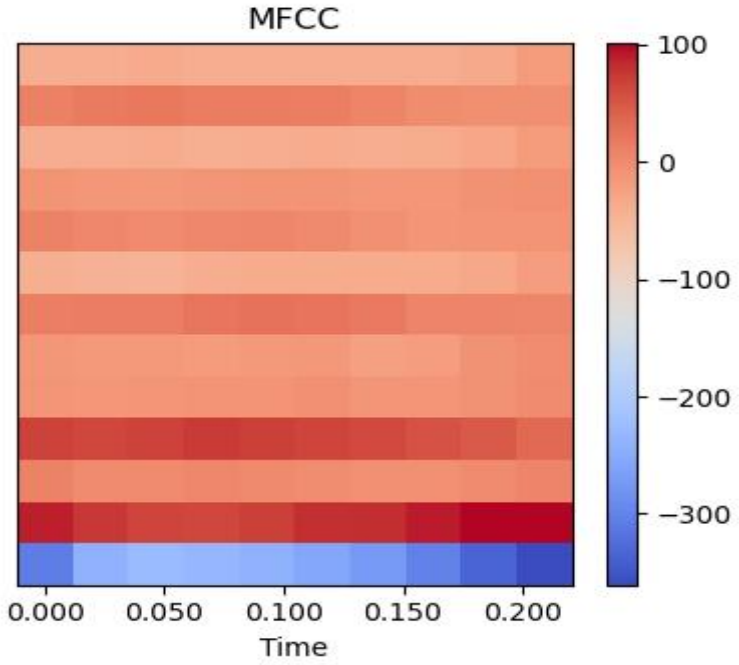
Euclidean distance between first column and column 1 of the same matrix: 0.0
Euclidean distance between first column and column 2 of the same matrix: 17.945070266723633
Euclidean distance between first column and column 3 of the same matrix: 37.90538787841797
Euclidean distance between first column and column 4 of the same matrix: 40.047428131103516
Euclidean distance between first column and column 5 of the same matrix: 39.604835510253906
Euclidean distance between first column and column 6 of the same matrix: 54.28848648071289
Euclidean distance between first column and column 7 of the same matrix: 68.82015991210938
Euclidean distance between first column and column 8 of the same matrix: 77.3283920288086
Euclidean distance between first column and column 9 of the same matrix: 81.86318969726562
Euclidean distance between first column and column 10 of the same matrix: 92.17730712890625

Vowel |i| :

MFCC features for C:\IIT-G Internship\ASCII Waves Recording\234101011_i_10-3.wav:

[[-3.08098419e+02 -2.42914963e+02 -2.29027481e+02 -2.34396484e+02
-2.42652573e+02 -2.56968353e+02 -2.73420624e+02 -3.02090820e+02
-3.34272858e+02 -3.60551910e+02]
[8.85430908e+01 7.45720215e+01 6.52996674e+01 6.29460678e+01
6.98274841e+01 7.92403412e+01 8.22811279e+01 9.15700989e+01
1.00603439e+02 1.00109520e+02]
[9.81128883e+00 9.96046901e-01 5.26072383e-01 4.97005367e+00
2.35574436e+00 -2.07060218e+00 -4.56171846e+00 -5.66147327e+00
2.15101004e-01 7.37822628e+00]
[6.64969177e+01 6.40833664e+01 6.69744415e+01 7.25142975e+01
6.92227631e+01 6.56073380e+01 6.11063080e+01 5.53971786e+01
4.77620773e+01 3.46474533e+01]
[-1.07678356e+01 -1.13269234e+01 -1.06191149e+01 -8.95854568e+00
-8.22668839e+00 -4.82553196e+00 -1.04702635e+01 -9.77834892e+00
-5.83545971e+00 1.39636898e+00]
[-1.22446156e+01 -1.58539810e+01 -1.64261341e+01 -1.65014381e+01
-1.53967781e+01 -1.46843357e+01 -2.25404358e+01 -2.00193863e+01
-6.71064568e+00 -1.78175247e+00]
[1.38454437e+01 1.53841429e+01 1.43084755e+01 2.21953735e+01
2.63100815e+01 2.49391403e+01 1.88168812e+01 8.28676891e+00
7.03014469e+00 5.43453360e+00]

[-4.04679108e+01 -4.45331421e+01 -4.55216408e+01 -3.91666794e+01
 -3.64719009e+01 -3.67824974e+01 -3.63787384e+01 -3.69905205e+01
 -3.20944176e+01 -1.88585720e+01]
[8.93927288e+00 4.68952322e+00 -6.78360462e-02 3.47247672e+00
  5.13178539e+00 1.78841376e+00 -4.52753162e+00 -9.42070770e+00
 -8.30786705e+00 -7.58317804e+00]
[-8.39308548e+00 -1.27835093e+01 -1.28873615e+01 -1.00768566e+01
 -8.58474541e+00 -8.64002609e+00 -1.14795141e+01 -1.18646088e+01
 -6.33909225e+00 -4.30468750e+00]
[-3.88346481e+01 -3.93081894e+01 -3.70739594e+01 -4.09201202e+01
 -3.97349129e+01 -3.75865936e+01 -3.88943939e+01 -3.74935341e+01
 -2.91999779e+01 -1.70590286e+01]
[1.16118183e+01 1.68092289e+01 2.11088142e+01 1.59026585e+01
  1.51996002e+01 1.28302050e+01 8.65713882e+00 -9.06635463e-01
 -2.87313533e+00 -2.59533739e+00]
[-3.85233841e+01 -3.83022842e+01 -3.61420898e+01 -3.93999481e+01
 -3.85159111e+01 -3.86501160e+01 -3.92577667e+01 -3.95455933e+01
 -3.30766640e+01 -1.82327347e+01]]

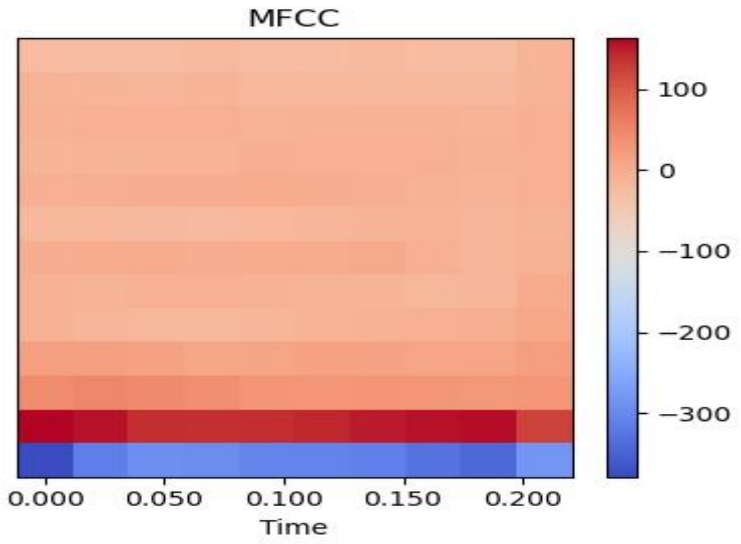


Euclidean distance between first column and column 1 of the same matrix: 0.0
Euclidean distance between first column and column 2 of the same matrix: 19.37990379333496
Euclidean distance between first column and column 3 of the same matrix: 29.500030517578125
Euclidean distance between first column and column 4 of the same matrix: 29.392385482788086
Euclidean distance between first column and column 5 of the same matrix: 25.084548950195312
Euclidean distance between first column and column 6 of the same matrix: 21.46348762512207
Euclidean distance between first column and column 7 of the same matrix: 24.950693130493164
Euclidean distance between first column and column 8 of the same matrix: 31.382896423339844
Euclidean distance between first column and column 9 of the same matrix: 37.33684539794922
Euclidean distance between first column and column 10 of the same matrix: 57.677330017089844

Vowel |u| :

MFCC features for C:\IIT-G Internship\ASCII Waves Recording\234101011_u_10-4.wav:

```
[[-3.7858109e+02 -3.1448361e+02 -2.9271432e+02 -2.9530002e+02
  -3.0740341e+02 -3.0825293e+02 -3.1120959e+02 -3.2904172e+02
  -3.4268735e+02 -2.8337692e+02]
 [ 1.6205362e+02  1.5484944e+02  1.4055774e+02  1.3688129e+02
    1.3954395e+02  1.4494392e+02  1.5061316e+02  1.5410245e+02
    1.5643163e+02  1.2333443e+02]
 [ 4.3544350e+01  4.8024513e+01  4.5224499e+01  4.0735672e+01
    3.3009666e+01  3.1050348e+01  3.1853561e+01  2.9547695e+01
    2.7112839e+01  3.0398342e+01]
 [ 1.6800835e+01  1.7683708e+01  1.2723030e+01  8.4760914e+00
    1.1397938e+01  1.5590961e+01  1.2469809e+01  1.1900520e+01
    1.1003634e+01  1.9682003e+01]
 [-7.2328682e+00 -1.4340352e+01 -1.5699909e+01 -1.6773586e+01
  -1.3612335e+01 -9.4266748e+00 -7.9935198e+00 -6.3709717e+00
  -3.7294080e+00  6.4038277e+00]
 [-8.1520090e+00 -1.0588312e+01 -5.2021055e+00 -5.7250929e+00
  -8.8817825e+00 -1.0962483e+01 -1.1227865e+01 -1.5979319e+01
  -1.4199347e+01 -4.7783607e-01]
 [-2.5435574e+00  4.4776380e-01 -2.7156448e-01 -1.8763576e+00
    4.3479797e-01  7.7334237e-01  3.6876616e+00 -5.1784511e+00
  -1.4015841e+01 -8.4973297e+00]
 [-1.8062908e+01 -1.9471325e+01 -2.1668053e+01 -2.1679798e+01
  -1.8348808e+01 -1.4088489e+01 -1.1378168e+01 -1.0326197e+01
  -1.3973052e+01 -1.0318947e+01]
 [-5.6269674e+00 -3.9618385e+00 -1.2796469e+00 -1.5720322e+00
  -1.6694391e-01 -8.3226490e-01 -4.4264174e+00 -7.3065319e+00
  -9.0141039e+00 -6.7043605e+00]
 [-1.1873879e+01 -1.0412034e+01 -9.4323149e+00 -1.0377415e+01
  -4.0052471e+00 -4.9384089e+00 -6.8853436e+00 -6.2080803e+00
  -8.5512781e+00 -6.0510044e+00]
 [-8.7014027e+00 -4.9236984e+00 -5.1230760e+00 -6.7690778e+00
  -9.5824108e+00 -8.7650194e+00 -7.5714846e+00 -8.6608324e+00
  -1.2188867e+01 -6.8884211e+00]
 [-1.0483044e+01 -1.1571942e+01 -1.3717555e+01 -1.2928986e+01
  -1.8346426e+01 -1.7360872e+01 -1.7739218e+01 -1.8916870e+01
  -2.0446266e+01 -9.4940453e+00]
 [-2.5841997e+01 -2.6477064e+01 -2.4861927e+01 -2.1910555e+01
  -2.3953743e+01 -2.3919773e+01 -2.2858673e+01 -2.3875130e+01
  -2.7495975e+01 -1.2772858e+01]]
```



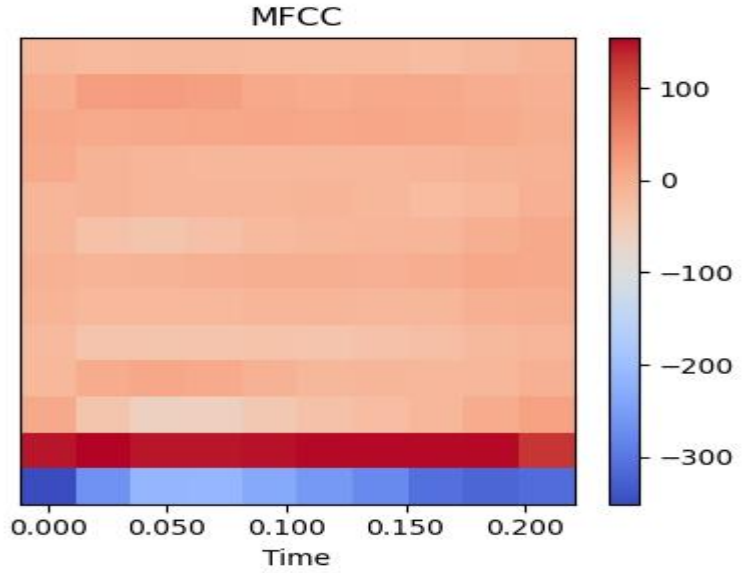
Euclidean distance between first column and column 1 of the same matrix: 0.0
Euclidean distance between first column and column 2 of the same matrix: 12.684823989868164
Euclidean distance between first column and column 3 of the same matrix: 25.0831298828125
Euclidean distance between first column and column 4 of the same matrix: 29.41314697265625
Euclidean distance between first column and column 5 of the same matrix: 29.23936653137207
Euclidean distance between first column and column 6 of the same matrix: 24.736780166625977
Euclidean distance between first column and column 7 of the same matrix: 21.656997680664062
Euclidean distance between first column and column 8 of the same matrix: 22.844505310058594
Euclidean distance between first column and column 9 of the same matrix: 25.866384506225586
Euclidean distance between first column and column 10 of the same matrix: 47.23025894165039

Appendix-B

Vowel |a| and |e| :

MFCC features for C:\IIT-G Internship\ASCII Waves Recording\Used\234101011_a_11(2)-2.wav:

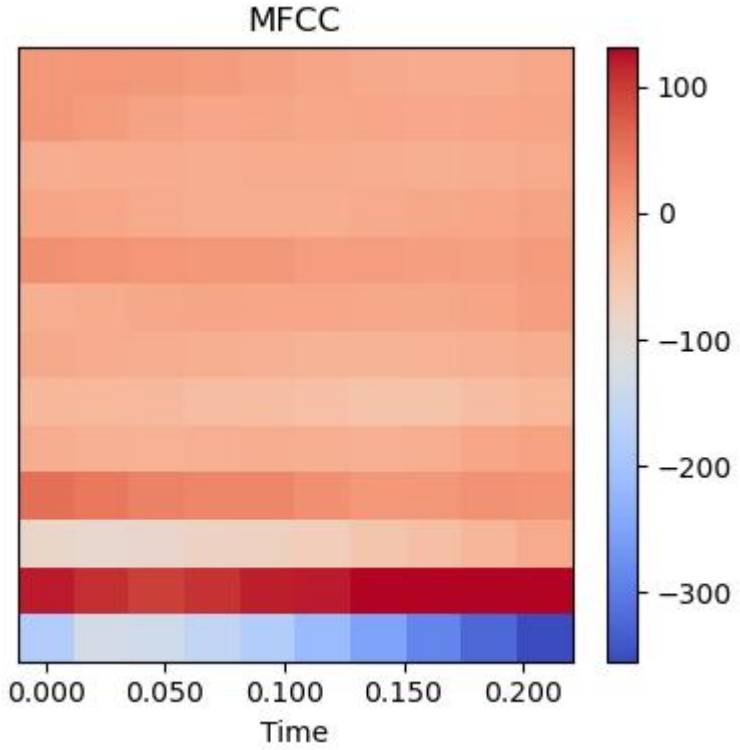
```
[[-351.00757   -267.04715   -213.45972   -211.7793    -234.43544
 -256.72998   -277.63052   -308.67694   -319.8311    -312.05426  ]
 [ 146.03682   154.03574   145.97983   145.63773   147.69489
  150.95065   150.5588    151.30782   150.57013   126.746704  ]
 [   6.092407  -35.51602   -59.464455  -57.227127  -42.361214
 -29.684784  -19.752018  -12.907835    3.8814592  16.698416  ]
 [ -14.169696    3.638988    8.18647     4.651394   -4.6651945
 -12.31867   -11.24423   -12.123945  -12.138529   -2.1274855 ]
 [ -18.280289  -33.66287   -36.25567   -33.876747  -33.135895
 -33.551727  -30.873142  -24.729717  -17.103405  -10.319367  ]
 [  -9.357389  -13.701714  -15.181549  -14.764006  -11.554573
  -9.762973  -13.44471   -12.12282    -3.6347733  -0.88035524]
 [  -4.485645   -8.859351   -7.415781   -2.751775   -0.9473416
  -0.92675334  -2.6860862   1.2148435    8.40486     6.176793  ]
 [ -11.215554  -30.257336  -36.328255  -28.979748  -18.66191
 -13.073076  -11.20605   -10.066433   -0.35809547   6.5718055 ]
 [  -9.80196    -7.5694876  -10.092711  -10.673542   -9.716137
  -9.359627  -11.735107  -21.278118  -14.871116   -2.44669   ]
 [   5.5019083  -6.805078  -10.635617  -12.30052   -12.995281
 -12.916488  -13.10021   -11.492619   -6.8954062  -5.2659607 ]
 [   9.028929    5.619988    7.1044736   8.174194   10.10865
   9.862011   11.450119    9.328226    5.8070135  -1.8140237 ]
 [   0.75368744 20.640318   22.453506   19.0405     7.842724
   3.692804    7.473935    7.5740213   1.7157669  -3.2051663 ]
 [ -12.643644  -18.89718   -17.196636  -17.005     -19.208546
 -18.203127  -18.697468  -22.201405  -15.7650385   -8.03163   ]]
```



MFCC features for C:\IIT-G Internship\ASCII Waves Recording\Used\234101011_e_10-4.wav:

```
[[-181.66394   -130.52516   -135.31253   -155.55383   -179.64703
 -215.27823   -250.7267    -290.00143   -323.54022   -354.83057  ]
```

[121.084656 109.21268   98.38544 107.57034 119.22588
 121.07956 129.7507 131.07277 130.14883 130.9653]
 [-86.94835 -91.757645 -88.11727 -77.68785 -74.905045
 -66.36754 -53.01963 -40.72782 -27.756775 -12.872986]
 [54.10563 46.83529 34.447002 31.470982 31.406414
 20.138264 10.706554 11.020617 17.392914 17.058363]
 [-16.464523 -19.666431 -21.568855 -18.957024 -16.043291
 -17.317417 -18.976555 -18.835724 -6.4380627 -0.41792202]
 [-31.748924 -30.66806 -33.865395 -38.07168 -39.295765
 -42.02412 -47.828484 -49.105385 -39.4506 -31.790142]
 [-11.257023 -14.310413 -16.88364 -18.076218 -19.845913
 -23.40499 -24.25985 -22.60356 -20.33097 -16.90108]
 [-18.19958 -13.214086 -7.525191 -4.611625 -5.9021416
 -7.160363 -8.890835 -7.9082985 -4.662279 2.8612022]
 [20.02372 17.059067 12.948431 11.424824 10.4334955
 4.9399076 5.6345243 3.4401748 1.1876383 6.9955883]
 [-4.476014 -7.4480414 -12.416617 -15.401159 -16.186893
 -16.185116 -12.78858 -8.78071 -6.1768656 -2.1935005]
 [-16.079887 -14.381432 -14.696789 -16.257153 -14.613634
 -14.691807 -15.298426 -17.293304 -16.088732 -11.662342]
 [12.081397 5.9020243 -3.4584785 -6.3440585 -5.4189157
 -7.8524446 -7.3111124 -9.153239 -5.9080706 -3.9139862]
 [11.030154 11.926878 9.987385 7.435226 1.8032324
 -4.143328 -10.255065 -13.460906 -14.683739 -7.841857]]

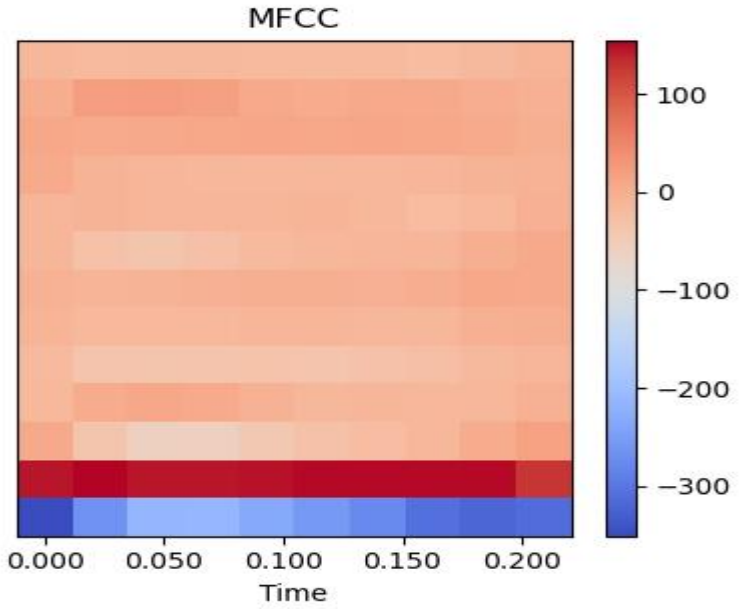


Euclidean distance between column 1 of the first matrix and column 1 of the second matrix: 129.80125427246094
Euclidean distance between column 2 of the first matrix and column 2 of the second matrix: 100.0822982788086
Euclidean distance between column 3 of the first matrix and column 3 of the second matrix: 87.6335220336914
Euclidean distance between column 4 of the first matrix and column 4 of the second matrix: 80.73468780517578
Euclidean distance between column 5 of the first matrix and column 5 of the second matrix: 79.98247528076172
Euclidean distance between column 6 of the first matrix and column 6 of the second matrix: 79.27765655517578
Euclidean distance between column 7 of the first matrix and column 7 of the second matrix: 71.57699584960938
Euclidean distance between column 8 of the first matrix and column 8 of the second matrix: 73.32635498046875
Euclidean distance between column 9 of the first matrix and column 9 of the second matrix: 72.99332427978516
Euclidean distance between column 10 of the first matrix and column 10 of the second matrix: 55.283775329589844

Vowel |a| and |i| :

MFCC features for C:\IIT-G Internship\ASCII Waves Recording\Used\234101011_a_11(2)-2.wav:

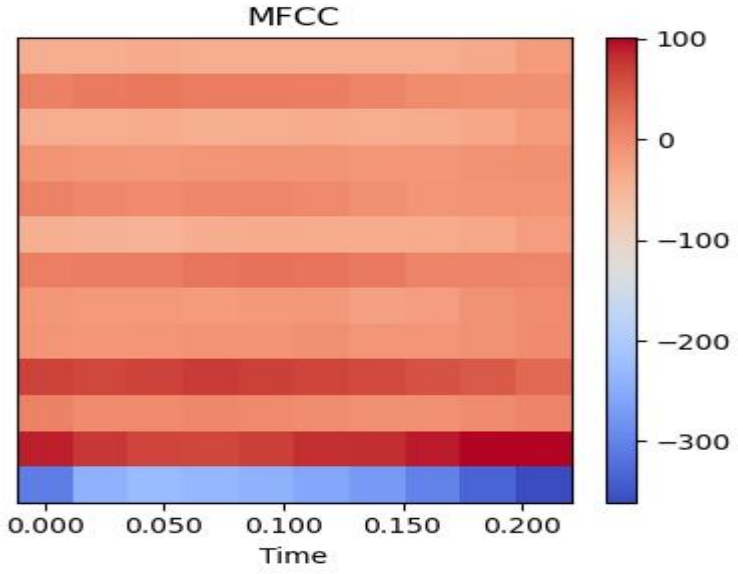
```
[[-351.00757   -267.04715   -213.45972   -211.7793    -234.43544
 -256.72998   -277.63052   -308.67694   -319.8311    -312.05426  ]
 [ 146.03682   154.03574   145.97983   145.63773   147.69489
  150.95065   150.5588    151.30782   150.57013   126.746704 ]
 [  6.092407  -35.51602   -59.464455  -57.227127  -42.361214
 -29.684784  -19.752018  -12.907835    3.8814592  16.698416 ]
 [ -14.169696    3.638988    8.18647     4.651394   -4.6651945
 -12.31867   -11.24423   -12.123945  -12.138529   -2.1274855 ]
 [ -18.280289  -33.66287   -36.25567   -33.876747  -33.135895
 -33.551727  -30.873142  -24.729717  -17.103405  -10.319367 ]
 [  -9.357389  -13.701714  -15.181549  -14.764006  -11.554573
  -9.762973  -13.44471   -12.12282    -3.6347733  -0.88035524]
 [  -4.485645   -8.859351   -7.415781   -2.751775   -0.9473416
  -0.92675334  -2.6860862   1.2148435   8.40486     6.176793  ]
 [ -11.215554  -30.257336  -36.328255  -28.979748  -18.66191
 -13.073076  -11.20605   -10.066433   -0.35809547   6.5718055 ]
 [  -9.80196    -7.5694876  -10.092711  -10.673542   -9.716137
  -9.359627  -11.735107  -21.278118  -14.871116   -2.44669   ]
 [   5.5019083  -6.805078  -10.635617  -12.30052   -12.995281
 -12.916488  -13.10021   -11.492619   -6.8954062  -5.2659607 ]
 [   9.028929    5.619988    7.1044736   8.174194   10.10865
   9.862011   11.450119    9.328226    5.8070135  -1.8140237 ]
 [   0.75368744  20.640318   22.453506   19.0405     7.842724
   3.692804    7.473935    7.5740213   1.7157669  -3.2051663 ]
 [ -12.643644  -18.89718   -17.196636  -17.005     -19.208546
 -18.203127  -18.697468  -22.201405  -15.7650385   -8.03163   ]]
```



MFCC features for C:\IIT-G Internship\ASCII Waves Recording\Used\234101011_i_10-3.wav:

```
[[-3.08098419e+02 -2.42914963e+02 -2.29027481e+02 -2.34396484e+02
 -2.42652573e+02 -2.56968353e+02 -2.73420624e+02 -3.02090820e+02
 -3.34272858e+02 -3.60551910e+02]
 [ 8.85430908e+01  7.45720215e+01  6.52996674e+01  6.29460678e+01
  6.98274841e+01  7.92403412e+01  8.22811279e+01  9.15700989e+01
  1.00603439e+02  1.00109520e+02]
```

[9.81128883e+00 9.96046901e-01 5.26072383e-01 4.97005367e+00
2.35574436e+00 -2.07060218e+00 -4.56171846e+00 -5.66147327e+00
2.15101004e-01 7.37822628e+00]
[6.64969177e+01 6.40833664e+01 6.69744415e+01 7.25142975e+01
6.92227631e+01 6.56073380e+01 6.11063080e+01 5.53971786e+01
4.77620773e+01 3.46474533e+01]
[-1.07678356e+01 -1.13269234e+01 -1.06191149e+01 -8.95854568e+00
-8.22668839e+00 -4.82553196e+00 -1.04702635e+01 -9.77834892e+00
-5.83545971e+00 1.39636898e+00]
[-1.22446156e+01 -1.58539810e+01 -1.64261341e+01 -1.65014381e+01
-1.53967781e+01 -1.46843357e+01 -2.25404358e+01 -2.00193863e+01
-6.71064568e+00 -1.78175247e+00]
[1.38454437e+01 1.53841429e+01 1.43084755e+01 2.21953735e+01
2.63100815e+01 2.49391403e+01 1.88168812e+01 8.28676891e+00
7.03014469e+00 5.43453360e+00]
[-4.04679108e+01 -4.45331421e+01 -4.55216408e+01 -3.91666794e+01
-3.64719009e+01 -3.67824974e+01 -3.63787384e+01 -3.69905205e+01
-3.20944176e+01 -1.88585720e+01]
[8.93927288e+00 4.68952322e+00 -6.78360462e-02 3.47247672e+00
5.13178539e+00 1.78841376e+00 -4.52753162e+00 -9.42070770e+00
-8.30786705e+00 -7.58317804e+00]
[-8.39308548e+00 -1.27835093e+01 -1.28873615e+01 -1.00768566e+01
-8.58474541e+00 -8.64002609e+00 -1.14795141e+01 -1.18646088e+01
-6.33909225e+00 -4.30468750e+00]
[-3.88346481e+01 -3.93081894e+01 -3.70739594e+01 -4.09201202e+01
-3.97349129e+01 -3.75865936e+01 -3.88943939e+01 -3.74935341e+01
-2.91999779e+01 -1.70590286e+01]
[1.16118183e+01 1.68092289e+01 2.11088142e+01 1.59026585e+01
1.51996002e+01 1.28302050e+01 8.65713882e+00 -9.06635463e-01
-2.87313533e+00 -2.59533739e+00]
[-3.85233841e+01 -3.83022842e+01 -3.61420898e+01 -3.93999481e+01
-3.85159111e+01 -3.86501160e+01 -3.92577667e+01 -3.95455933e+01
-3.30766640e+01 -1.82327347e+01]]



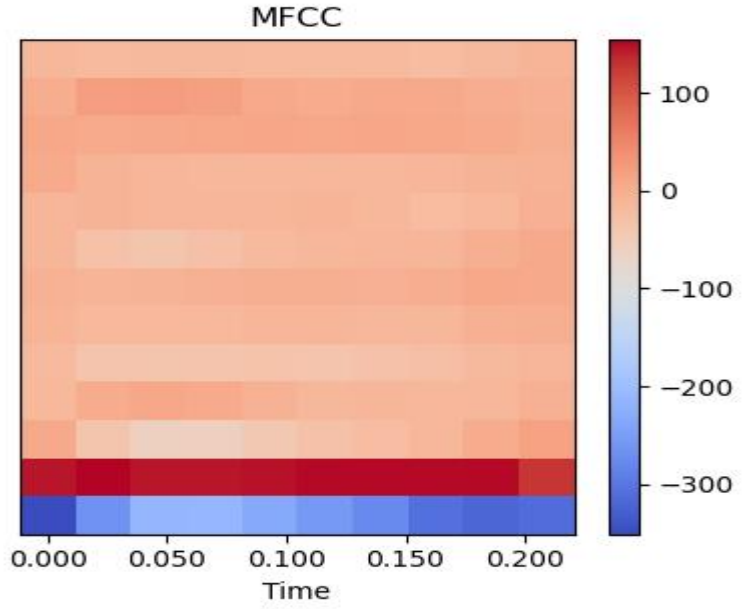
Euclidean distance between column 1 of the first matrix and column 1 of the second matrix: 121.2676773071289
Euclidean distance between column 2 of the first matrix and column 2 of the second matrix: 123.25587463378906
Euclidean distance between column 3 of the first matrix and column 3 of the second matrix: 131.1390838623047
Euclidean distance between column 4 of the first matrix and column 4 of the second matrix: 140.6717071533203
Euclidean distance between column 5 of the first matrix and column 5 of the second matrix: 135.53211975097656
Euclidean distance between column 6 of the first matrix and column 6 of the second matrix: 130.2055206298828
Euclidean distance between column 7 of the first matrix and column 7 of the second matrix: 121.38990783691406
Euclidean distance between column 8 of the first matrix and column 8 of the second matrix: 109.30133056640625

Euclidean distance between column 9 of the first matrix and column 9 of the second matrix: 93.98580932617188
Euclidean distance between column 10 of the first matrix and column 10 of the second matrix: 57.4306526184082

Vowel |a| and |u| :

MFCC features for C:\IIT-G Internship\ASCII Waves Recording\Used\234101011_a_11(2)-2.wav:

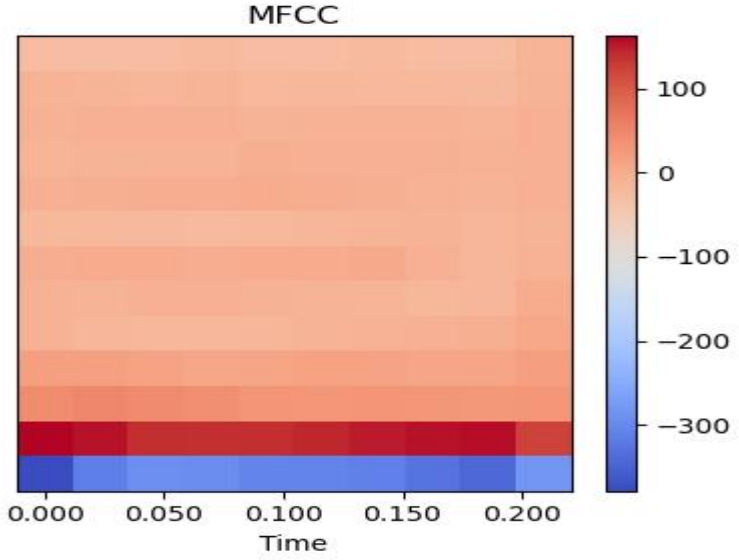
```
[[-351.00757   -267.04715   -213.45972   -211.7793    -234.43544
 -256.72998   -277.63052   -308.67694   -319.8311    -312.05426  ]
 [ 146.03682   154.03574   145.97983   145.63773   147.69489
  150.95065   150.5588    151.30782   150.57013   126.746704  ]
 [   6.092407  -35.51602   -59.464455  -57.227127  -42.361214
 -29.684784   -19.752018  -12.907835    3.8814592  16.698416  ]
 [ -14.169696    3.638988    8.18647     4.651394   -4.6651945
 -12.31867    -11.24423   -12.123945  -12.138529   -2.1274855 ]
 [ -18.280289  -33.66287   -36.25567   -33.876747  -33.135895
 -33.551727   -30.873142  -24.729717  -17.103405  -10.319367  ]
 [  -9.357389  -13.701714  -15.181549  -14.764006  -11.554573
  -9.762973   -13.44471   -12.12282    -3.6347733  -0.88035524]
 [  -4.485645   -8.859351   -7.415781   -2.751775   -0.9473416
  -0.92675334  -2.6860862   1.2148435    8.40486     6.176793  ]
 [ -11.215554  -30.257336  -36.328255  -28.979748  -18.66191
 -13.073076   -11.20605   -10.066433   -0.35809547   6.5718055 ]
 [  -9.80196    -7.5694876  -10.092711  -10.673542   -9.716137
  -9.359627   -11.735107  -21.278118  -14.871116   -2.44669   ]
 [   5.5019083  -6.805078  -10.635617  -12.30052   -12.995281
 -12.916488   -13.10021   -11.492619   -6.8954062  -5.2659607 ]
 [   9.028929    5.619988    7.1044736   8.174194   10.10865
   9.862011   11.450119    9.328226    5.8070135  -1.8140237 ]
 [   0.75368744 20.640318   22.453506   19.0405     7.842724
   3.692804    7.473935    7.5740213   1.7157669  -3.2051663 ]
 [ -12.643644  -18.89718   -17.196636  -17.005     -19.208546
 -18.203127   -18.697468  -22.201405  -15.7650385  -8.03163   ]]
```



MFCC features for C:\IIT-G Internship\ASCII Waves Recording\Used\234101011_u_10-4.wav:

```
[[-3.7858109e+02 -3.1448361e+02 -2.9271432e+02 -2.9530002e+02
 -3.0740341e+02 -3.0825293e+02 -3.1120959e+02 -3.2904172e+02
  -3.4268735e+02 -2.8337692e+02]
 [ 1.6205362e+02  1.5484944e+02  1.4055774e+02  1.3688129e+02
  1.3954395e+02  1.4494392e+02  1.5061316e+02  1.5410245e+02
```

1.5643163e+02 1.2333443e+02]
[4.3544350e+01 4.8024513e+01 4.5224499e+01 4.0735672e+01
3.3009666e+01 3.1050348e+01 3.1853561e+01 2.9547695e+01
2.7112839e+01 3.0398342e+01]
[1.6800835e+01 1.7683708e+01 1.2723030e+01 8.4760914e+00
1.1397938e+01 1.5590961e+01 1.2469809e+01 1.1900520e+01
1.1003634e+01 1.9682003e+01]
[-7.2328682e+00 -1.4340352e+01 -1.5699909e+01 -1.6773586e+01
-1.3612335e+01 -9.4266748e+00 -7.9935198e+00 -6.3709717e+00
-3.7294080e+00 6.4038277e+00]
[-8.1520090e+00 -1.0588312e+01 -5.2021055e+00 -5.7250929e+00
-8.8817825e+00 -1.0962483e+01 -1.1227865e+01 -1.5979319e+01
-1.4199347e+01 -4.7783607e-01]
[-2.5435574e+00 4.4776380e-01 -2.7156448e-01 -1.8763576e+00
4.3479797e-01 7.7334237e-01 3.6876616e+00 -5.1784511e+00
-1.4015841e+01 -8.4973297e+00]
[-1.8062908e+01 -1.9471325e+01 -2.1668053e+01 -2.1679798e+01
-1.8348808e+01 -1.4088489e+01 -1.1378168e+01 -1.0326197e+01
-1.3973052e+01 -1.0318947e+01]
[-5.6269674e+00 -3.9618385e+00 -1.2796469e+00 -1.5720322e+00
-1.6694391e-01 -8.3226490e-01 -4.4264174e+00 -7.3065319e+00
-9.0141039e+00 -6.7043605e+00]
[-1.1873879e+01 -1.0412034e+01 -9.4323149e+00 -1.0377415e+01
-4.0052471e+00 -4.9384089e+00 -6.8853436e+00 -6.2080803e+00
-8.5512781e+00 -6.0510044e+00]
[-8.7014027e+00 -4.9236984e+00 -5.1230760e+00 -6.7690778e+00
-9.5824108e+00 -8.7650194e+00 -7.5714846e+00 -8.6608324e+00
-1.2188867e+01 -6.8884211e+00]
[-1.0483044e+01 -1.1571942e+01 -1.3717555e+01 -1.2928986e+01
-1.8346426e+01 -1.7360872e+01 -1.7739218e+01 -1.8916870e+01
-2.0446266e+01 -9.4940453e+00]
[-2.5841997e+01 -2.6477064e+01 -2.4861927e+01 -2.1910555e+01
-2.3953743e+01 -2.3919773e+01 -2.2858673e+01 -2.3875130e+01
-2.7495975e+01 -1.2772858e+01]]



Euclidean distance between column 1 of the first matrix and column 1 of the second matrix: 61.0466194152832
Euclidean distance between column 2 of the first matrix and column 2 of the second matrix: 94.84318542480469
Euclidean distance between column 3 of the first matrix and column 3 of the second matrix: 115.73062896728516
Euclidean distance between column 4 of the first matrix and column 4 of the second matrix: 107.1085433959961
Euclidean distance between column 5 of the first matrix and column 5 of the second matrix: 87.54248809814453
Euclidean distance between column 6 of the first matrix and column 6 of the second matrix: 77.7845230102539
Euclidean distance between column 7 of the first matrix and column 7 of the second matrix: 70.0101547241211

Euclidean distance between column 8 of the first matrix and column 8 of the second matrix: 63.49504852294922
Euclidean distance between column 9 of the first matrix and column 9 of the second matrix: 55.47802734375
Euclidean distance between column 10 of the first matrix and column 10 of the second matrix: 39.52167510986328