Military Institute of Science and Technology

***Experiment 01*:** **Analyzing Audio Signals of Distinctive Words Using MATLAB**

Name: **Anindya Chanda Tirtha**

ID: **202216009**

Section: **A**

Course Code: **Digital Signal Processing I Laboratory [EECE-312]**

Date of Submission: 20/1/2025

1. Objective

- ❖ To use Fourier analysis to examine the frequency components of the Bangla words "chaar" and "tiin."
- ❖ To determine the unique qualities of the words' "aa" and "ii" vowel sounds.
- ❖ To distinguish between the terms "chaar" and "tiin" using their distinct correlation patterns and frequency signatures.
- ❖ To identify and match vowel sounds in audio sources using cross-correlation analysis.

2.Methodology

2.1 Audio Recording

Voices were taken from me, my mother, my sister and my brother. Additional clips containing isolated vowel sounds "aa" and "ii" were also recorded. Audio files were loaded into MATLAB using the “audioread” function. Visual inspection of the waveform was used to determine time intervals for isolating the words "chaar" and "tiin". Extracted segments were normalized by dividing them with their maximum amplitude for consistency in amplitude.

Step 1: Import Training Audio Files

```
[audio_tin, fs_tin] = audioread("G:\DSP assignment\EXP 01\voice\Male\MINE_TEEN.opus");
[audio_chaar, fs_chaar] = audioread("G:\DSP assignment\EXP 01\voice\Male\MINE_CHAAR.opus");

% Resample training audio if sampling rates don't match
if fs_tin ~= fs_chaar
    target_fs = max(fs_tin, fs_chaar);
    audio_tin = resample(audio_tin, target_fs, fs_tin);
    audio_chaar = resample(audio_chaar, target_fs, fs_chaar);
    fs = target_fs;
else
    fs = fs_tin;
end
```

Step 2: Preprocess Training Audio

```
audio_tin = audio_tin / max(abs(audio_tin));
audio_chaar = audio_chaar / max(abs(audio_chaar));
```

Step 3: Compute FFT for Training Audio

```
N_tin = length(audio_tin);
fft_tin = fft(audio_tin);
fft_tin = fft_tin(1:floor(N_tin/2)+1);
freq_tin = (0:floor(N_tin/2)) * (fs / N_tin);
amp_tin = abs(fft_tin / N_tin) * 2;

N_chaar = length(audio_chaar);
fft_chaar = fft(audio_chaar);
fft_chaar = fft_chaar(1:floor(N_chaar/2)+1);
```

```
freq_chaar = (0:floor(N_chaar/2)) * (fs / N_chaar);
amp_chaar = abs(fft_chaar / N_chaar) * 2;
```

Step 4: Plot Frequency Spectrum for Training Data

```
figure;

subplot(2, 1, 1);
plot(freq_tin, amp_tin);
title('Frequency Spectrum of "tiin"');
xlabel('Frequency (Hz)');
ylabel('Amplitude');
xlim([0 2000]);

subplot(2, 1, 2);
plot(freq_chaar, amp_chaar);
title('Frequency Spectrum of "chaar"');
xlabel('Frequency (Hz)');
ylabel('Amplitude');
xlim([0 2000]);
```

The audio signals were subjected to the Fast Fourier Transform (FFT). For my audio, frequency spectra for "chaar" and "tiin" were plotted. The main frequencies for "tiin" and "chaar" were then manually extracted from the plot. Using spectral peaks, frequencies associated with the vowel sounds "aa" and "ii" were determined. As seen in the image, they were **174.621,357.179** for "tiin" and **155.604,614.425,1230.53** for "chaar" (indicated in figure 2). Amplitude greater than 0.01 is the criterion for choosing the dominant frequency. As a result, unique frequency ranges for each vowel sound were highlighted to differentiate the words.

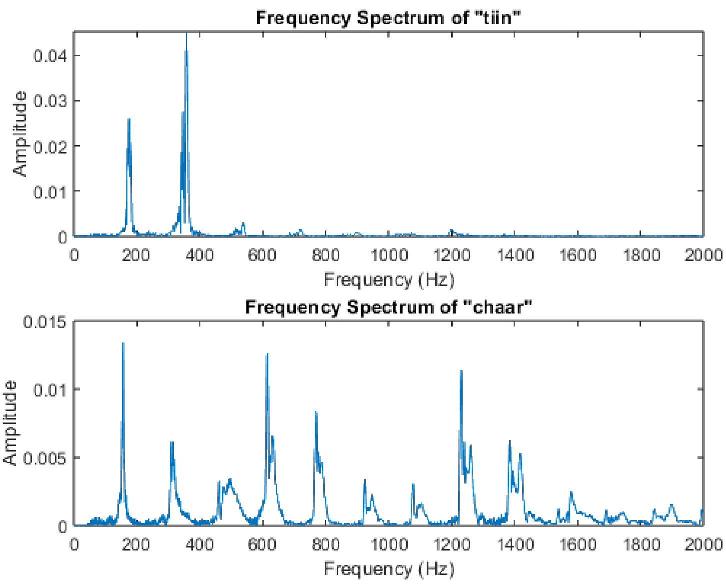

Figure 1: Frequency Spectrum For "tiin" and "chaar"

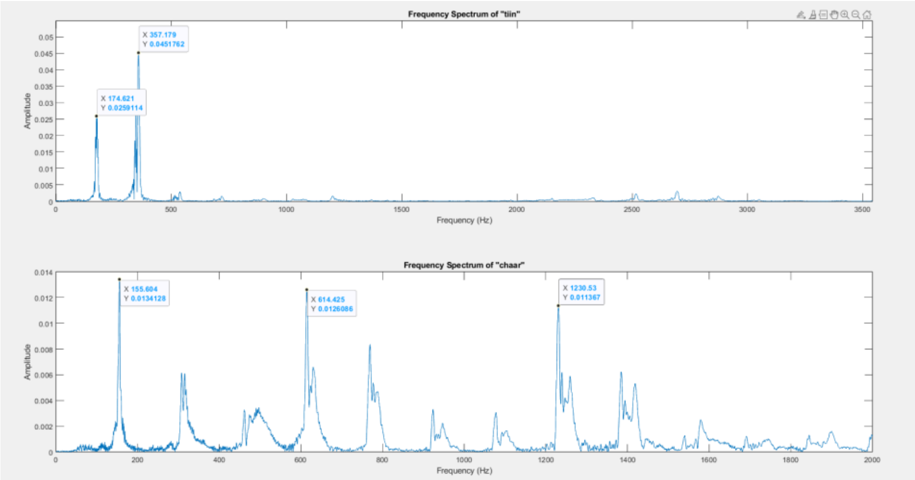

Figure 2: Dominant Frequencies for "tiin" and "chaar"

Step 5: Extract Dominant Frequencies for Training

```
dominant_freqs_tiin = [174.621,357.179];
dominant_freqs_chaar = [155.604,614.425,1230.53];
```

Step 6: Import and Preprocess Test Signal

```
[test_audio, fs_test] = audioread("G:\DSP assignment\EXP 01\voice\Male\bro_TEEN.m4a");

if fs_test ~= fs
    test_audio = resample(test_audio, fs, fs_test);
end

test_audio = test_audio / max(abs(test_audio));

N_test = length(test_audio);
fft_test = fft(test_audio);
fft_test = fft_test(1:floor(N_test/2)+1);
freq_test = (0:floor(N_test/2)) * (fs / N_test);
amp_test = abs(fft_test / N_test) * 2;
```

Step 7: Extract Dominant Frequencies for Test Signal

```
[pks_test, locs_test] = findpeaks(amp_test, 'MinPeakHeight', 0.01, 'MinPeakDistance', 20);
dominant_freqs_test = freq_test(locs_test);

disp('Dominant frequencies in test signal:');
disp(dominant_freqs_test);
```

```
Dominant frequencies in test signal:304.2641
```

Step 7.1: Plot Frequency Spectrum of Test Signal

```
figure;
plot(freq_test, amp_test);
title('Frequency Spectrum of Test Signal');
xlabel('Frequency (Hz)');
```

```
ylabel('Amplitude');
xlim([0 2000]);
```

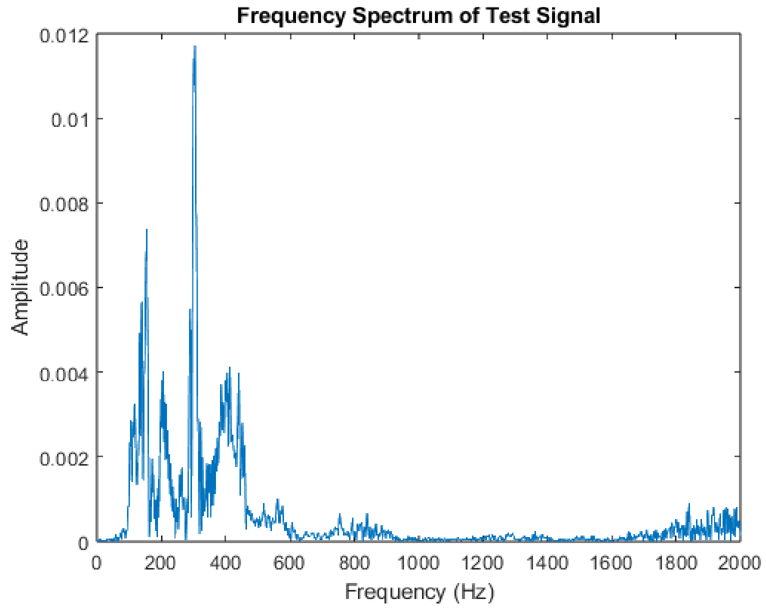

Figure 3:Frequency Spectrum for Test Signal

Step 8: Classify the Test Signal

```
distance_to_tiin = sum(min(abs(dominant_freqs_test' - dominant_freqs_tiin), [], 2));
distance_to_chaar = sum(min(abs(dominant_freqs_test' - dominant_freqs_chaar), [], 2));

if distance_to_tiin < distance_to_chaar
    disp('The test signal is classified as "tiin".');
else
    disp('The test signal is classified as "chaar".');
end
```

```
The test signal is classified as "tiin".
```

Following the taking of a test signal (I've taken this one for my brother).Using the findpeaks function, I will similarly extract the dominating frequencies.then to identify the test signal,I will merely calculate the distance from my refered tiin and chaar.Whether my test signal is tiin or chaar will be determined by the minimum distance.

2.3.Cross-Correlation Analysis

Audio segments containing "aa" and "ii" were used as templates. MATLAB's xcorr() function was applied to compute the cross-correlation between: "aa" and test signal and "ii" and test signal Peaks in correlation values were identified and analyzed to determine the stronger match for each vowel sound. For testing in this case , mom_teen audio file has been taken.

Step 1: Load Voice Files

```
% Load 'tin' audio file
EEE_file = "G:\DSP assignment\EXP 01\voice\'EE'sound.mp3";
```

```
[eee, Fs] = audioread(EEE_file);

% Load 'char' audio file
AAA_file = "G:\DSP assignment\EXP 01\voice\'AA'sound.mp3";
[aaa, ~] = audioread(AAA_file);

% Load test audio file
test_file ="G:\DSP assignment\EXP 01\voice\FEmale\mom_teen.opus";
[test_signal, ~] = audioread(test_file);
```

Step 2: Preprocessing

```
eee = eee / max(abs(eee));
aaa = aaa / max(abs(aaa));
test_signal = test_signal / max(abs(test_signal));
```

Step 3: Cross-Correlation

Pad signals to the same length.

```
max_length = max([length(test_signal), length(eee), length(aaa)]);
test_signal_padded = [test_signal; zeros(max_length - length(test_signal), 1)];
eee_padded = [eee; zeros(max_length - length(eee), 1)];
aaa_padded = [aaa; zeros(max_length - length(aaa), 1)];

% Compute cross-correlation with reference signals
[xcorr_eee, lags_eee] = xcorr(test_signal_padded, eee_padded, 'coeff'); % Normalized cross-
correlation
[xcorr_aaa, lags_aaa] = xcorr(test_signal_padded, aaa_padded, 'coeff');

% Find maximum correlation values
[max_corr_eee, idx_eee] = max(xcorr_eee);
[max_corr_aaa, idx_aaa] = max(xcorr_aaa);

% Display results
disp(['Maximum Correlation with eee: ', num2str(max_corr_eee)]);
disp(['Maximum Correlation with aaa: ', num2str(max_corr_aaa)]);
```

```
Maximum Correlation with eee: 0.050936
Maximum Correlation with aaa: 0.027603
```

Step 4: Plot Correlation Results

```
figure;
subplot(2, 1, 1);
plot(lags_eee, xcorr_eee);
title('Cross-Correlation with eee');
xlabel('Lag');
ylabel('Correlation');
grid on;

subplot(2, 1, 2);
plot(lags_aaa, xcorr_aaa);
title('Cross-Correlation with aaa');
xlabel('Lag');
```

```
ylabel('Correlation');
grid on;
```

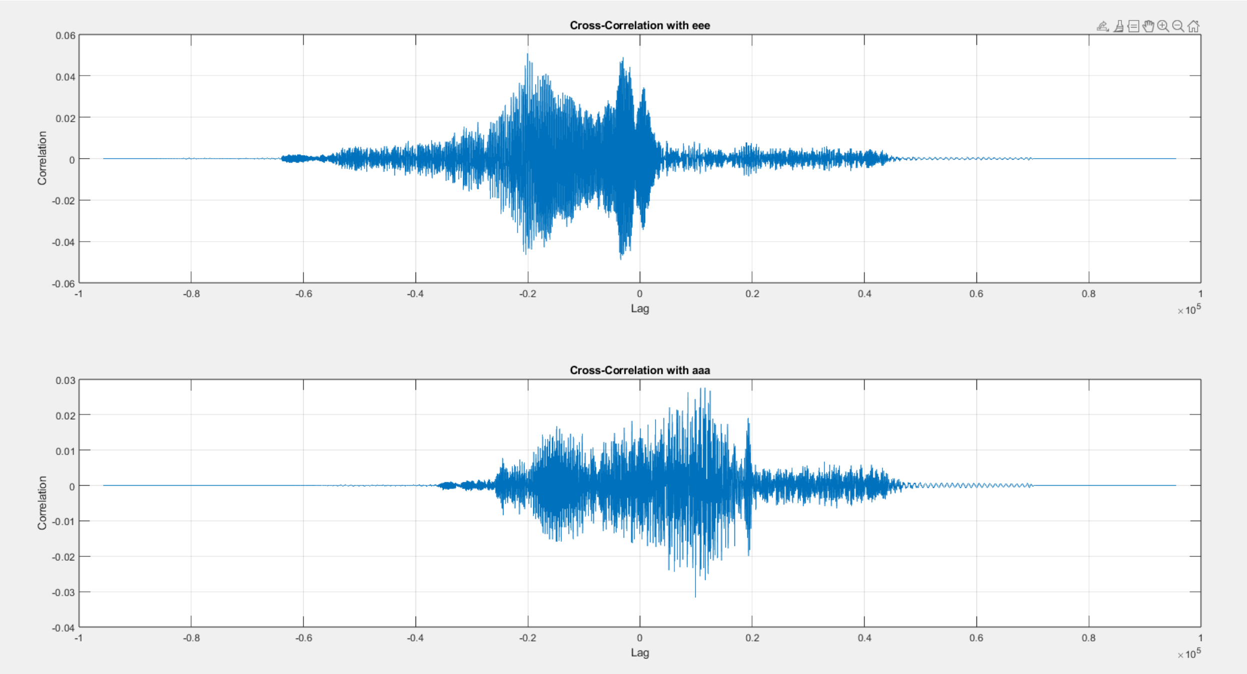

Figure 4:Crosscorelation plots with "eee" and "aaa"

Step 5: Classification

Classify based on the higher correlation value.

```
if max_corr_eee > max_corr_aaa
    disp('Predicted class: tin');
else
    disp('Predicted class: char');
end
```

Predicted class: tin.

3.Result analysis and Discussion

Fourier analysis and cross-correlation techniques were applied successfully in this project to differentiate between Bangla words "chaar" and "tiin." We could classify the audio samples based on their unique frequency characteristics and correlation patterns provided by the vowel sounds "aa" and "ii." Key insights/observations derived from the experiment are given below.

Frequency Characteristics:

- The "aa" vowel in the word "chaar" comprised dominant frequencies of 155.604 Hz, 614.425 Hz, and 1230.53 Hz, specifying that its energy spreads over both low and high-frequency ranges.
- The "ii" vowel in "tiin" dominated frequencies were 174.621 Hz and 357.179 Hz, concentrated energy was in the mid-frequency range.

- ❖ The unique spectral patterns for these vowels constitute a reliable guide to distinguishing the words "chaar" from "tiin.".

Fourier Analysis:

The Fast Fourier Transform allowed for a clear view of the frequency spectrum of both words. Peaks in the frequency domain pinpointed the characteristics of each vowel sound.

Normalizing the audio signals helped to minimize amplitude variations due to recording conditions, making the analysis more accurate.

Cross-Correlation Analysis:

Using the **xcorr()** function, the test signal was cross-correlated with reference signals containing isolated vowel sounds ("aa" and "ii").Where a larger correlation value between the test and reference signals was, there was a better match. The largest correlation of 0.050936 is "eee," while 0.027603 corresponds to "aaa," hence, classifying the test signal as "tiin.". Cross-correlation allowed for robust temporal alignment of the test signal and the reference ones, thus giving great results even in small variations in pronunciations or recording conditions.

Classification Accuracy:

The test signals were classified into "chaar" or "tiin" based on dominant frequencies and correlation values. For example, the test signal was classified as "tiin" because it was closer to the dominant frequencies and correlation patterns of "tiin."

Even though the extraction of dominant frequencies manually introduced subjectivity, the method proved effective for distinguishing the words within the dataset.

Limitations and Challenges:

<u>Background Noise:</u> Although minimized during preprocessing, residual noise may have affected the accuracy of the extracted frequency components and the results of the correlation analysis.

<u>Manual Frequency Selection:</u> The reliance on visual inspection for the identification of dominant frequencies introduces a degree of subjectivity. Automating this step could improve reproducibility and precision.

<u>Dataset Variability:</u> Variations in pronunciation, recording devices, and environmental conditions can introduce inconsistencies in the data.