THE UNIVERSITY OF TEXAS AT ARLINGTON

COMPUTER SCIENCE AND ENGINEERIG

Discrete Fourier Transform Project

SIGNAL PROCESSING

Submitted by,

Servando Olvera

1001909287

Date 11/20/2023

Project A: Speech Recognition

Objective:

The goal of the project is to categorize two or more audio samples of spoken digits into their respective digit using DFT properties.

Data:

- Audio samples of spoken Digits from 0 to 9 are provided for the development of the project.
- There are 40 samples total for each digit.
- There are 10 samples for testing the accuracy of the project.

Suggestions:

- Use the DFT to find ‘features’ that are common to a particular digit.
    - For instance, “one” uses lower frequencies and “six” uses higher frequencies. The ratio of the average of lower frequency DFT coefficients to the average of higher frequency DFT coefficients can be used as a feature.
    - You can look at the DFTs for all speech samples of a digit and look for characteristics common to that digit that are different from other digits.

Problem

The problem at hand in this project is speech recognition specifically focused on categorizing two of the spoken digits from 0 to 9. The objective is to develop a system that can, to a certain degree, identify and categorize at least two spoken digits.

Objective: Automatically categorize two audio samples of spoken digits.

Data: Audio Samples of Spoken Digits from 0 to 9.

50 samples total of each digit.

40 samples of each digit for system development.

10 samples of each digit for evaluating the system's performance.

Key Challenges: Limited Data & Variability in Speech of Unknown Test Samples:

Limited Data: While 40 samples per digit provide a reasonable starting point, it might not capture the full variability present in natural speech. Limited data could lead to oversimplifying the system or adapting it to a single source, thus affecting the accuracy of it.

Different speakers pronounce digits differently. While the data provided appears to come from a single speaker there are variations in accents, pitches, speed, and emphasis, that this system does not account for.

Noise and Environment. Audio samples could have background noise, varying recording quality, or other environmental factors affecting the clarity of the spoken digits. An assumption is being made here about the clarity of the audio for this project.

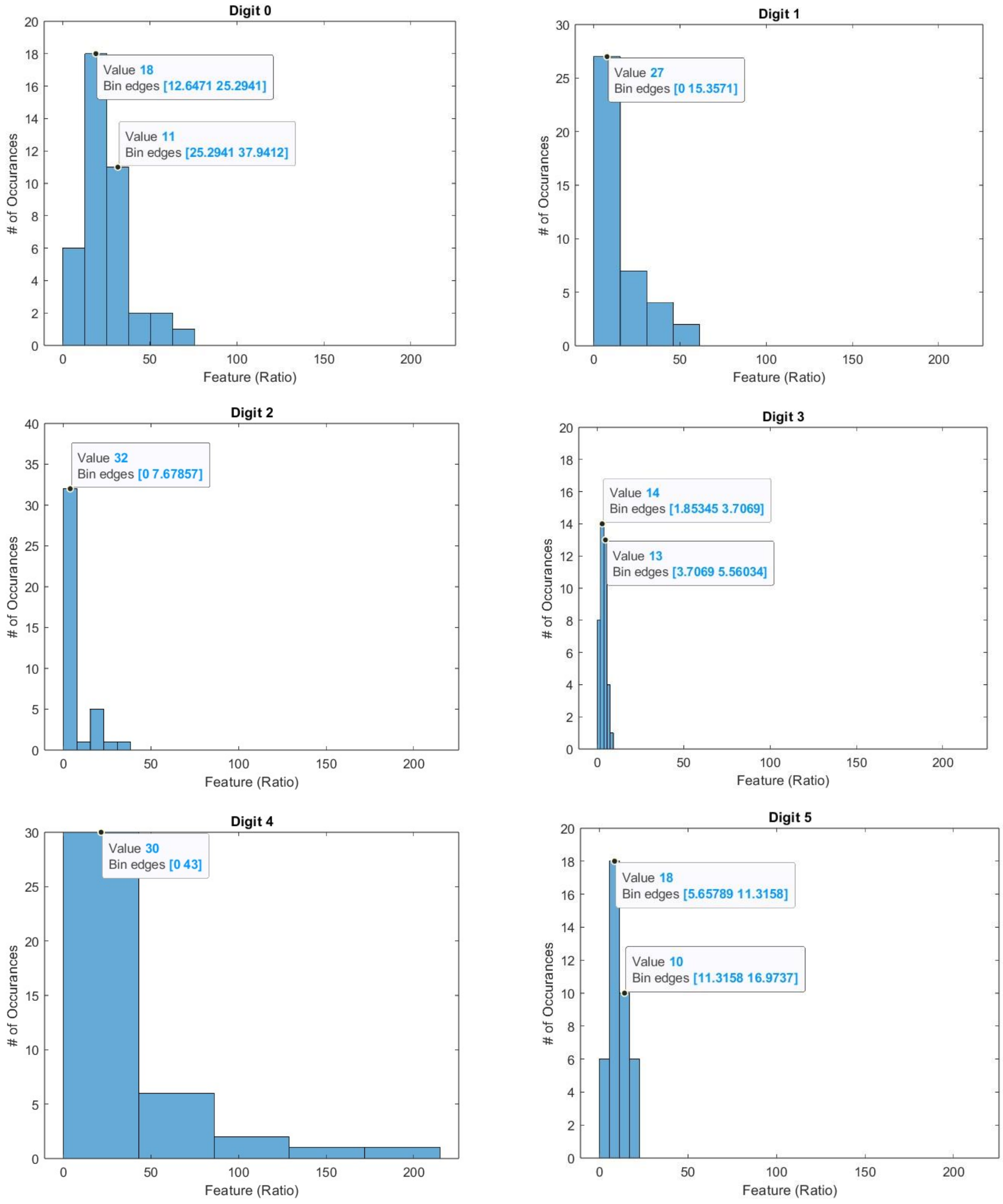
Methods & Approach

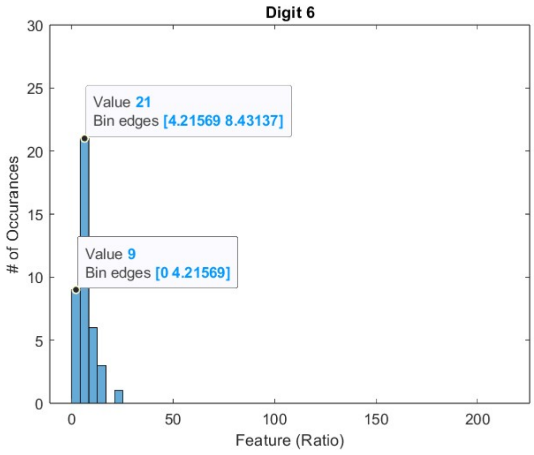
For this project, I decided to work with MATLAB. Although not too familiar with this language, it appears to be quite powerful when it comes to dealing with signal processing. Additionally, MATLAB has built in tools and functions that made it much simpler to approach the goal of this project.

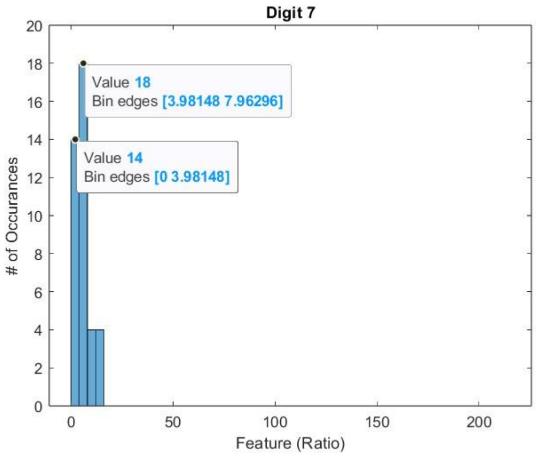
My approach to this project involved the first suggestion mentioned in the slides. In code I processed all 40 samples for each of the 10 digits: this entailed computing DFT coefficients of each audio sample for each digit, computing its low and high frequency ranges, then computing the average of lower and higher frequency DFT coefficients, and ultimately computing the ratio of low to high, and using that as a defining feature for each respective sample. I then plotted a histogram of the ratio feature over all samples for each digit to observe the range of these features and how often each one would occur. From there I was able to compare all the histograms of each digit with one another to find the two with the least overlap in ratio, that way I would be able to distinguish between two numbers or possibly more.

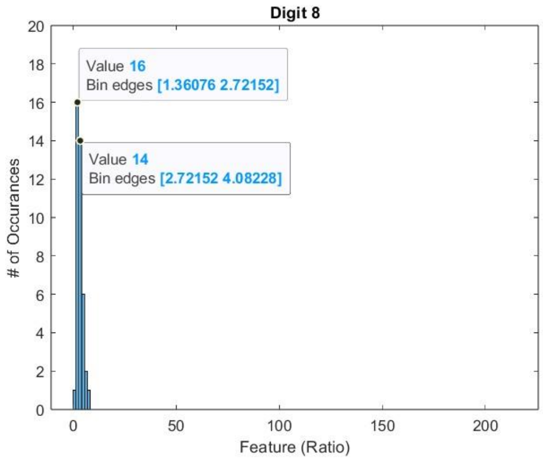
Results

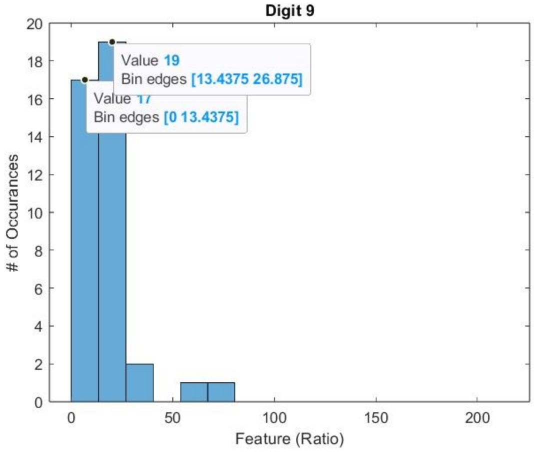
Histogram Of Each Digit



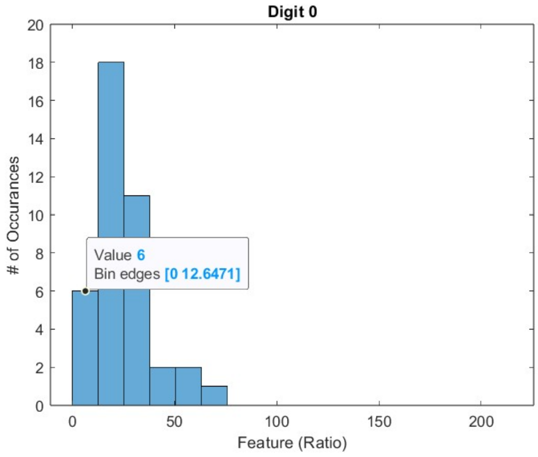


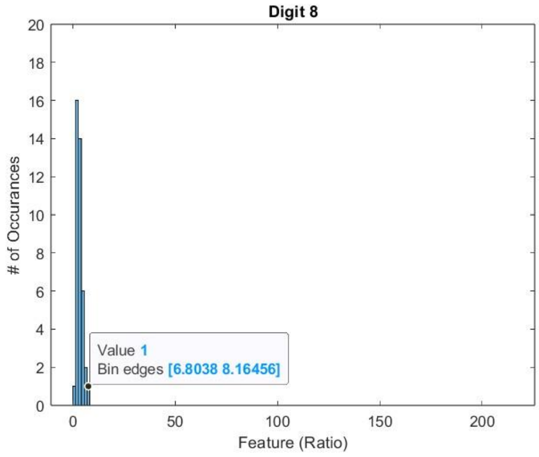






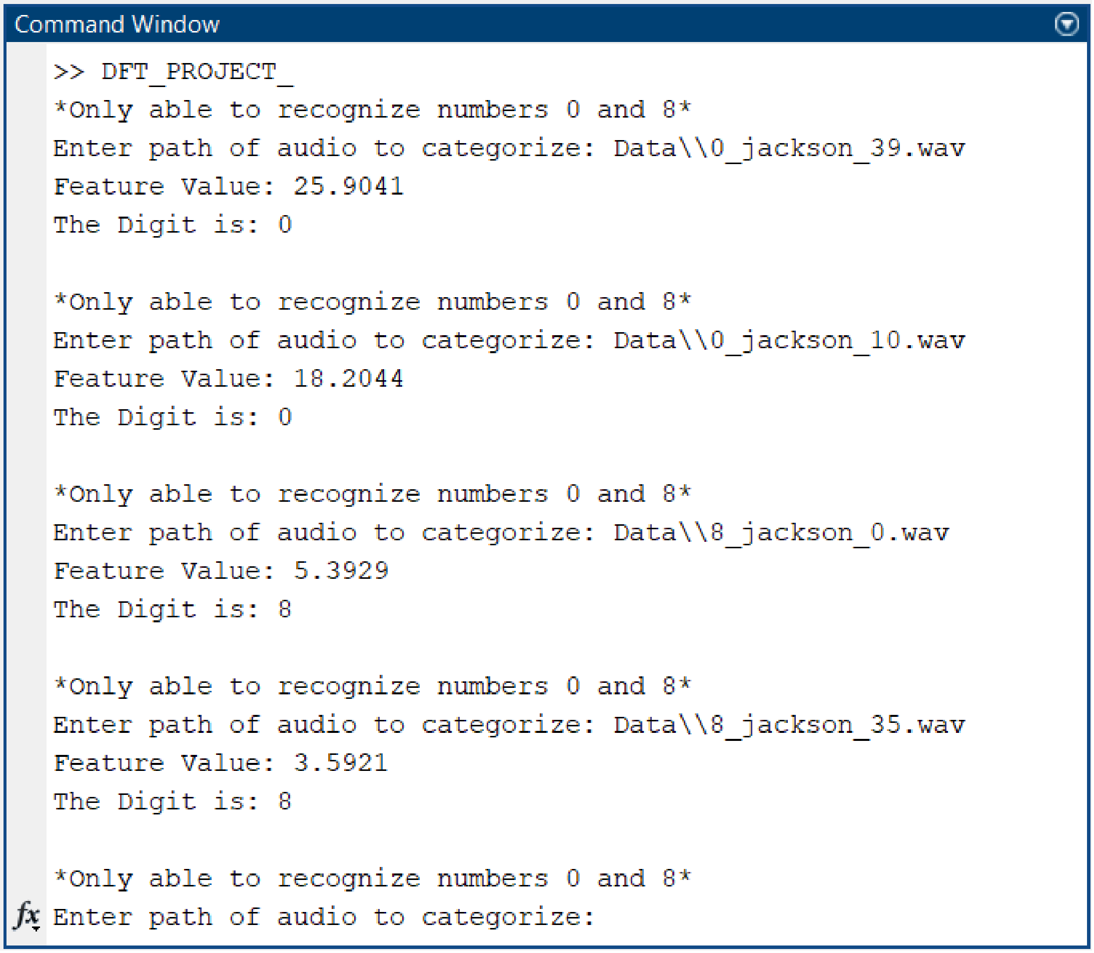
Histogram Of Digits with Least Overlap in Ration





From the histograms above, it can be observed that digits 0 & 8 have the least overlap in ratio. Digit 8 appears to have a range of 0-8.16, and that 8.16 value is a single instance from the 40 samples. Meanwhile, Digit 0 has a range of 0-75.88 with 6 instances out of the 40 being less than 12 and out of those 6 instances, 5 of them are bigger than 8. With this information it can be inferred that if the value of the feature is less than 8 then the audio sample is digit 8, otherwise, if the value of the feature is bigger than 8 then the digit is 0. These two numbers are the sound choice for the system to attempt to automatically categorize an audio sample of digits 0 or 8.

Output from Code



```
>> DFT_PROJECT_
*Only able to recognize numbers 0 and 8*
Enter path of audio to categorize: Data\\0_jackson_39.wav
Feature Value: 25.9041
The Digit is: 0

*Only able to recognize numbers 0 and 8*
Enter path of audio to categorize: Data\\0_jackson_10.wav
Feature Value: 18.2044
The Digit is: 0

*Only able to recognize numbers 0 and 8*
Enter path of audio to categorize: Data\\8_jackson_0.wav
Feature Value: 5.3929
The Digit is: 8

*Only able to recognize numbers 0 and 8*
Enter path of audio to categorize: Data\\8_jackson_35.wav
Feature Value: 3.5921
The Digit is: 8

*Only able to recognize numbers 0 and 8*
fx Enter path of audio to categorize:
```