English Vowel Phonemes Identification

Submitted to: Dr. Abualseoud Hanani

Department of Electrical and Computer Engineering

Birzeit University

Group members:
Nazeeh Rimawi 1170341
Nour Hallack , 1171088
Qamar Jabari, 1182522

***Abstract*— This project aims to design, implement and evaluate English vowel phonemes identification system by training sets of collected data for each vowel using three different signal processing methods: Correlation, Fourier transform and Filter bank.**

I. Introduction

Digital Signal Processing (DSP) is the processing of the real-word signals such as voice, audio, video, temperature pressure… etc., in the form of digital signal. Processing of these signals include performing several operations on them like “add”, “subtract”, “multiply” and many other operations.

The digital processing of a real-word signal includes digitalizing the signal (converting it to digital form). This can be done using an analog to digital converter (A/D). When applying this conversion, the signal turns into 1, 0 signal to be processed in computers. The digital signal processing then takes a place in processing these digitalized signals. After processing the digitalized signal by DSP systems, an analog signal can be then reconstructed using a digital to analog converter (D/A) so it can be ready for the world use.

A programmer can use many platforms to deal with/design/etc. digital signal processing systems. MATLAB is one of these platforms. “MATLAB is a programming platform designed specifically for engineers and scientists. The heart of MATLAB is the MATLAB language, a matrix-based language allowing the most natural expression of computational mathematics”. [1]

The existence of platforms such as MATLAB and the DSP being programmable have led to the use of DSP in many different applications. Moreover, some applications are purely built using concepts of DSP. This project is such a DSP application that made on MATLAB using the basic concepts of DSP such as filters, Fourier Transform.

The aim of the project is to determine the unknown vowel from a (.wav) file by applying one of the three methods used : Correlation, Fourier transform and Filter bank. This project is used in many everyday applications that ease human life, for example: Google translate, Siri to make and many automated systems.

II. Problem Specification

The main idea of this project is to build a relatively accurate model, by training a set of (.wav) files, that contains spelling of 13 different vowels in different accents and tones using different DSP techniques and functions, in order to represent each vowel by a vector then compare it with the unknown vowel signal according to the method used to decide which vowel it is.

The first phase in each method is to represent the training set of signals with one vector to each vowel (.wav) files through different techniques such as finding the vector closest to them all or just a function like taking their mean simply. The second phase is processing the unknown vowel (.wav) file, same as processing the training set before, in order to compare them next. The third phase is to compare each template signal of the vowels found in previous phases with the unknown vowel signal using one of the methods used. The fourth phase is to use the result from the previous part to decide which vowel the unknown signals are.

The three method used in this project : Cross Correlation , Fourier Transform and Filter Bank are not the ideal methods to build a high accurate model , error percentage is relatively high and caused by several factors as alignment where not all speakers of the voices starts at the same time , also the shortage of training data and the outlier existence where many accent with different pronunciation caused a misrepresentation, , also noticing some noise that surround the speaker that cannot be humanly heard but caused different power ranges then supposed to be.

III. Data

Inputs to system are (.wav) files, thirteen voice contains vowel (heed, hid, head, had, hard, hudd, hod, heard, hoard, hood, who’d, hade, hide) for each vowel we have ten (.wav) files for different English native speakers that vary in accent and tunes for both genders. We divide this data into two subsets (training and testing), seven for training and three for testing for each vowel in the three methods.

©2021 Birzeit University

Outputs of the systems are strings, which is a word that contains the target vowel in the unknown voice.

IV. Evaluation Criteria

The evaluation of the project is mainly based on the accuracy of the result. The accuracy of the results of our project has developed through working on the project. The first results were with accuracy less than required. Then, after working more on the project and studying more MATLAB and DSP techniques, the results were improved to give better accuracy. The results were also developed to be better by adding more features to the project. For example, we cut the signals off and keep the vowel only, so noise was reduced.

Staging in methods even make results more accurate throw working.At the end of the work on the project, the results were so satisfying comparing to the results and accuracy at earlier stages of working on the project.

V. Approach

The model was built in three different methods that vary in accuracy, efficiency and structure simplicity.

The First method “Cross Correlation” is applied by calculating a reference signal for each vowel using **mean( )** function in Matlab then taking the cross correlation between each vowel reference signal and the unknown vowel in .wav file signal using **xcorr( )** function, the vowel is then recognized by comparing between the results and choosing the highest correlation.

The Second method “Fourier Transform” is applied by converting the time-domain signal into frequency-domain using **ftt( )** function in Matlab. Then, we find the average of distances between the unknown signal and all set of signals for each vowel in the frequency domain. After that, the vowel is determined by finding the minimum average distance between the other averages. This method use the properties provided in the quantized spectrum to compare it with the unknown vowel signal.

The Third method “Filter Bank ” is applied by finding the reference signals for each vowel using **mean( )** function in Matlab then pass each of them through 5 Passband filters with different cut-off frequencies using **bandpass( )** function the result will be a power vector for each , after that the unknown signal from the (.wav) file is passed through the same 5 band pass filters and the dot product between unknown vowel signal power vector with each vowel reference power vector using **dot( )** function is found to recognize the right vowel .

VI. Results and Analysis

After evaluating each method separately we found that Fourier transform method gets the highest accuracy followed by filter bank method and finally cross correlation.

The results improved dramatically after cutting the signal off covering the vowel in each voice only and reducing noise to represent clearer signals.

We encountered some shortfalls, while determining the appropriate cutoff frequencies in each filter in filter bank method. Also the system was slow and tedious to run and results to appear. Dealing with signals with different number of samples was conflicting and cause and redundancy in code syntax.

After the training stage, we tested every method separately, we get different accuracies where the Fourier transform was 55% accurate while filter bank achieve 42% accuracy. On the other hand cross correlation didn’t exceeded 12% accuracy. The accuracy was calculated by summing the correct results divided by number of testing sets.

VII. Development

This system can be developed by using different techniques and more complicated ones such as pattern matching method, technologies involved include dynamic time warping (DTW), hidden Markov (HMM) and vector quantization (VQ). Phonetic pronunciation sounds are collected and pre-processed to meet the data stream requirements of speech recognition system. The feature data are extracted from the speech signal data stream, which are matched with the standard reference model. Results matched are then evaluated for the final output [2].

For the system we write can be developed by using more testing files in the three methods. For the Filter bank method, it can be more develop by using filters in frequency domain rather than time domain for more efficient. On other hand we can develop our system to auto detect the vowel frequency and cut the signal off and keep the vowel only.

VIII. Conclusion

The main operations of this project (English Vowel Phonemes Identification System) have included many concepts and basics of the DSP. The project has included dealing with filters and discrete Fourier transform as a main concept. And for sure many other concepts of DSP. Also, many MATLAB skills were gained after working on such a project. After finish working on the project, very satisfying results were obtained. But still, many more developments and features can be added to the project.

One of the great problem solving techniques we have learned is to divide the problem into phases in order to simplify solving it easily and simply. Also, gaining teamwork skill was enjoyable and empowering with the beautiful spirit that exists.

REFERENCES

[1] https://www.mathworks.com/discovery/what-is-matlab.html

[2] Research and Realization of Intelligent English Phonetic Pronunciation Training System. Hong Zheng

[3] Bandpass-filter signals - MATLAB bandpass (mathworks.com)