PROJECT 16 - EC5.408

Spectral Features in Vowel Regions for Language Identification

Speech Signal Processing - Mid Project Presentation

Kowshik Sai	**2019101073**
Dheekshith Akula	**2019112022**
Rahothvarman P	**2020114008**

Agenda

- Introduction
- Motivation
- Approach
- Work Done So Far
- Contribution
- References

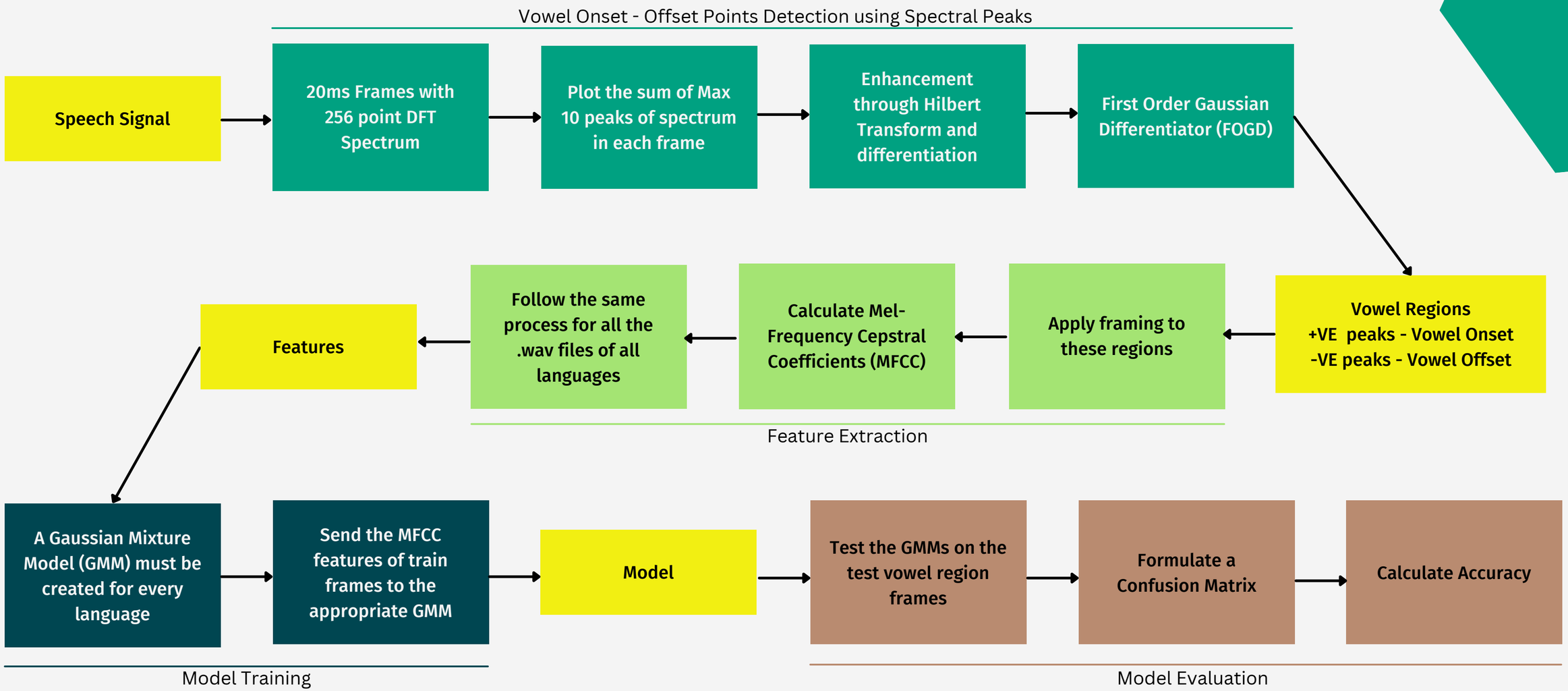
Introduction

- A speech signal comprises a voiced signal, an unvoiced signal and noise. vowels, along with a few consonants are considered voiced, while others as unvoiced.
- Voiced speech has high energy compared to unvoiced speech, therefore voiced signals tend to be louder like the vowels /a/, /e/, /i/, /u/, /o/. unvoiced signals, on the other hand, tend to be more abrupt, like the stop consonants /p/, /t/, /k/.
- Vowels are produced by keeping the vocal tract in an open position with minimum obstruction along the length and using glottal vibration as the excitation.
- The vowel onset point is the instant at which the beginning of a vowel takes place during speech production.
- The vowel offset point is the instant at which the ending of a vowel takes place during speech production.
- Whenever a vowel is produced, there must be a significant change in energy. thus, to identify, we will analyze the energy of the speech signal and process it to find vowel onset and offset points.

Motivation

- Language Identification(LID) systems classify spoken language from the audio sample by an unknown speaker.
- Automatic spoken language identification serves as a pre-processing step for a number of speech applications like multi-lingual automatic speech recognition, speech-to-speech translation, and customer routing in call centres.
- To create a LID system, we need the features to train the GMM model, and spectral features extracted from the vop regions are one of the best features for it.
- There are many existing methods, like zcr and pitch, whose performance depend on the amount of evidence available about the VOP in the selected speech feature, which is dependent on the method employed for the extraction of the feature.
- Hence, there may always be a limitation on the performance achieved using only one feature.
- Once we have the vops, we can find extract the spectral features and use them for training a GMM model for the LID detection

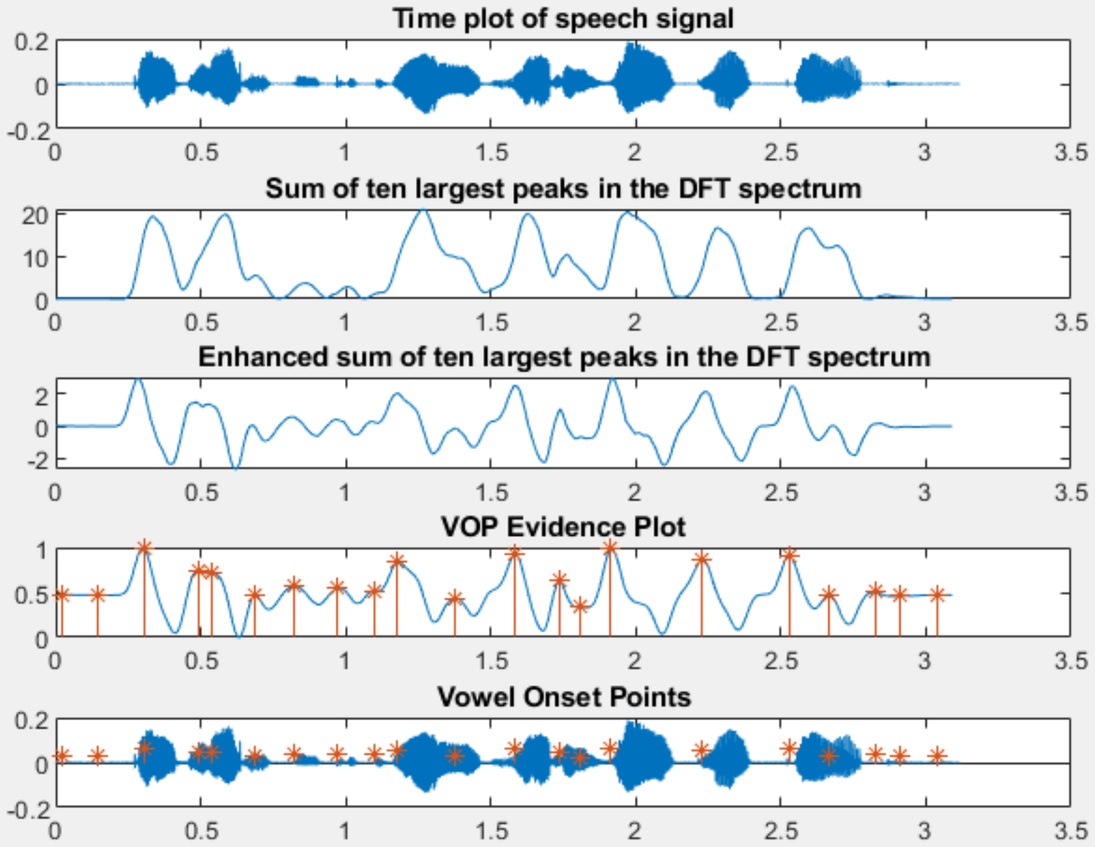
Approach


Vowel Onset - Offset Points Detection using Spectral Peaks
Speech Signal
20ms Frames with 256 point DFT Spectrum
Plot the sum of Max 10 peaks of spectrum in each frame
Enhancement through Hilbert Transform and differentiation
First Order Gaussian Differentiator (FOGD)
Vowel Regions
+VE peaks - Vowel Onset
-VE peaks - Vowel Offset
Apply framing to these regions
Calculate Mel-Frequency Cepstral Coefficients (MFCC)
Follow the same process for all the .wav files of all languages
Features
Feature Extraction
A Gaussian Mixture Model (GMM) must be created for every language
Send the MFCC features of train frames to the appropriate GMM
Model
Model Training
Test the GMMs on the test vowel region frames
Formulate a Confusion Matrix
Calculate Accuracy
Model Evaluation

Work Done So Far ...

Vowel Onset Points Detection

Successfully implemented the IEEE Transactions Paper on Vowel Onset Point Detection using Spectral Features. Using these peaks, the vowel regions in the speech were identified.



Timeline

Vowel Onset Points Detection

Successfully implemented the IEEE Transactions Paper on Vowel Onset point Detection using Spectral Features. Using these peaks, the vowel regions in the speech was identified.

Present

Mid Project Presentations. Here's a link to the github repository which contains the code implemented till now ...

LINK

Extract MFCC Features

Calculate MFCCs for each frame. Use these values as features for the classifier model

Train and Test the Gaussian Mixture Model

Train the GMMs on the MFCC features on Language Identification Task. Report the results.

Contribution

Kowshik Sai :

- Reading the Research Papers
- Enhancement of VOP evidence

Dheekshith Akula :

- Reading the Research Papers
- Extract MFCC features

Rahothvarman P :

- Reading the Research Papers
- Implementing Vowel Onset Points Detection

References

1. Vowel Onset Point Detection Using Source, Spectral Peaks, and Modulation Spectrum Energies

2. Vowel Onset Point Detection for Low Bit Rate Coded Speech


THANK
YOU!