INITIAL PROJECT REPORT

DSP LAB (EC3093D)

IMPLEMENTATION OF MEL SPECTROGRAM FOR SPEECH RECOGNITION

under the instruction of

Dr Akhil P T

DEPARTMENT OF

ELECTRONICS AND COMMUNICATION ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY CALICUT

KERALA, INDIA 673601

Adwayith K S (B210664ec)
Adhyuth Narayan (B210650ec)
Aditya Tuppad (B210038ec)

1. MOTIVATION

A Mel spectrogram, or Mel-frequency spectrogram, is a representation of the frequency content of a signal over time, where the frequencies are converted into the Mel scale, which is a perceptual scale of pitches that approximates the human auditory system's response to sounds of different frequencies. The Mel scale is designed to mimic the non-linear response of human ears to different frequencies, where we are more sensitive to changes in pitch at lower frequencies compared to higher frequencies.

Feature extraction refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original data set. It yields better results than applying machine learning directly to the raw data.

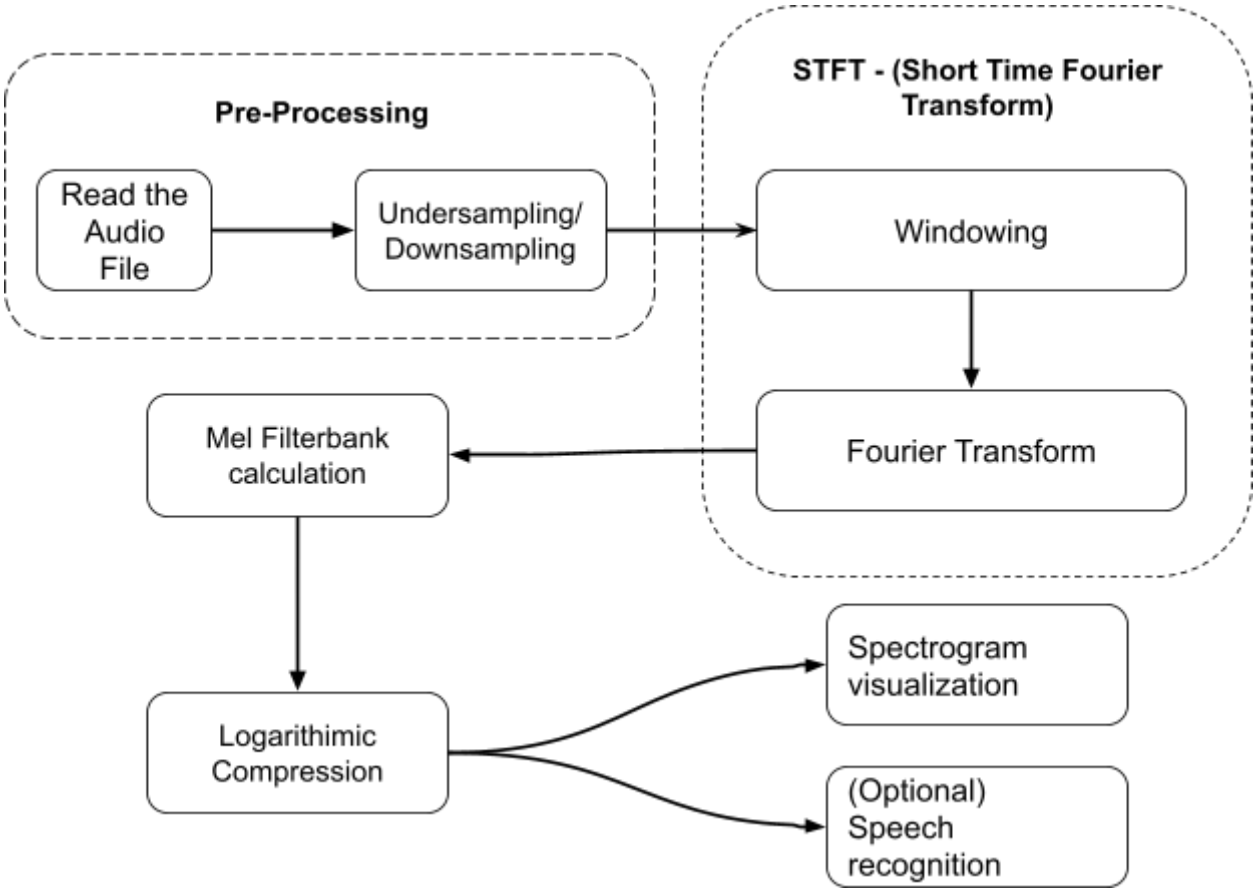
In fields like speech recognition, music analysis, and sound classification, Mel spectrograms are commonly used as features for machine learning algorithms. These spectrograms provide a compact representation of the audio signal that captures important perceptual features.

Mel spectrograms are widely used in speech processing tasks such as speech recognition, speaker identification, and emotion recognition. They are also used in audio processing tasks such as music genre classification, sound event detection, and audio scene analysis.

2. OBJECTIVES

Audio signals are one dimensional - amplitude over time. Our objective is to implement Mel spectrogram, a transformation that details the frequency composition of the signal over time.
The project involves manually implementing STFT in matlab and design of Mel Filterbanks.
Optionally, we will test our results on a pre-trained existing speech/sound recognition model.

3. METHODOLOGY



- *Pre - Processing*: Reads the audio file and samples to a optimum level the program can work on
- *Short Time Fourier Transform*: analyzes the frequency content of a signal over short, overlapping time intervals. By dividing the signal into frames and applying the Fourier Transform to each frame, the STFT provides a time-varying representation of the signal's frequency components. The resulting spectrogram visualizes the signal's time-frequency characteristics, making STFT essential in fields like audio processing, speech recognition, and music analysis for tasks such as feature extraction, event detection, and signal denoising.

- *Mel Filterbanks*: By applying Mel filterbanks to the magnitude spectrum obtained from the Short-Time Fourier Transform (STFT), Mel-frequency cepstral coefficients (MFCCs) can be extracted. MFCCs are a compact representation of the spectral envelope of an audio signal
- *(Optional) Speech recognition*: The features extracted can then be passed to an already existing trained model like YAMnet (contains about 512 sound classes for audio recognition) or Mozilla's Deepspeech. YAMnet has very good support in Matlab and we hope to use that.
