

# Design of Automatic Speaker Recognition (Partial System) using MFCC Feature Extraction in Python

**Digital Signal Processing Lab (ECE229)**

---

MAULANA AZAD NATIONAL INSTITUTE OF  
TECHNOLOGY, BHOPAL

**Submitted By:**

Name: Harshvardhan Patidar

Scholar Number: 2311401217

Department: Electronics and Communication Engineering

Project Title: MFCC-based Speaker Recognition

## 1. Introduction

This mini-project focuses on developing a speaker recognition system using MFCC (Mel-Frequency Cepstral Coefficients). Speaker recognition involves identifying or verifying a person's identity using their voice. This report presents a modular Python implementation and a Streamlit-based application that visualizes the feature extraction process.

## 2. Objective

Design and implement an MFCC-based front-end for a speaker recognition system. This includes frame blocking, windowing, FFT, mel filtering, and DCT to extract MFCCs from audio files, with visualization through an interactive web app.

## 3. Methodology

The MFCC pipeline consists of the following stages:

1. Frame Blocking
2. Windowing (Hamming Window)
3. FFT - Conversion to Frequency Domain
4. Mel Filter Bank Application
5. Log Compression and DCT to obtain MFCCs

## 4. Streamlit Application

The Streamlit web app enables users to upload .wav files, adjust processing parameters (frame size, overlap, sample rate, mel filters, etc.), and visualize various components such as waveform, power spectrum, mel filters, and MFCC heatmaps.

## 5. Exam Problem Statement Analysis

### 5.1 Spectrogram vs MFCC Interpretation

The app provides side-by-side plots of the spectrogram and MFCC heatmap, showing how MFCCs extract perceptually important features.

### 5.2 Frame Size and Overlap Analysis

Interactive controls allow observing how varying frame duration and overlap affect MFCC resolution and time granularity.

### 5.3 Mel Filter Bank Customization

Users can select the number of mel filters to examine how granularity changes both the filter shapes and MFCC features.

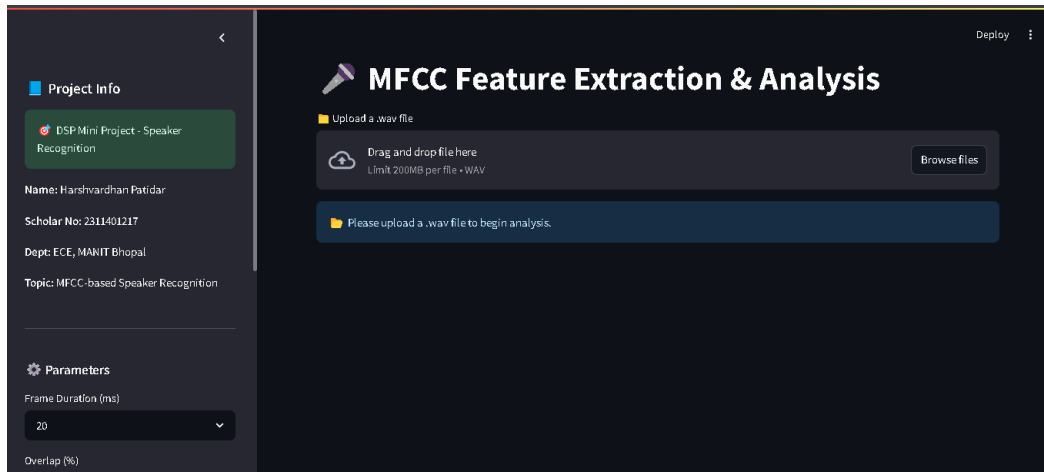
## 5.4 Sampling Rate Degradation Study

Resampling options let users explore how downsampling affects signal fidelity, power spectrum, and MFCC quality.

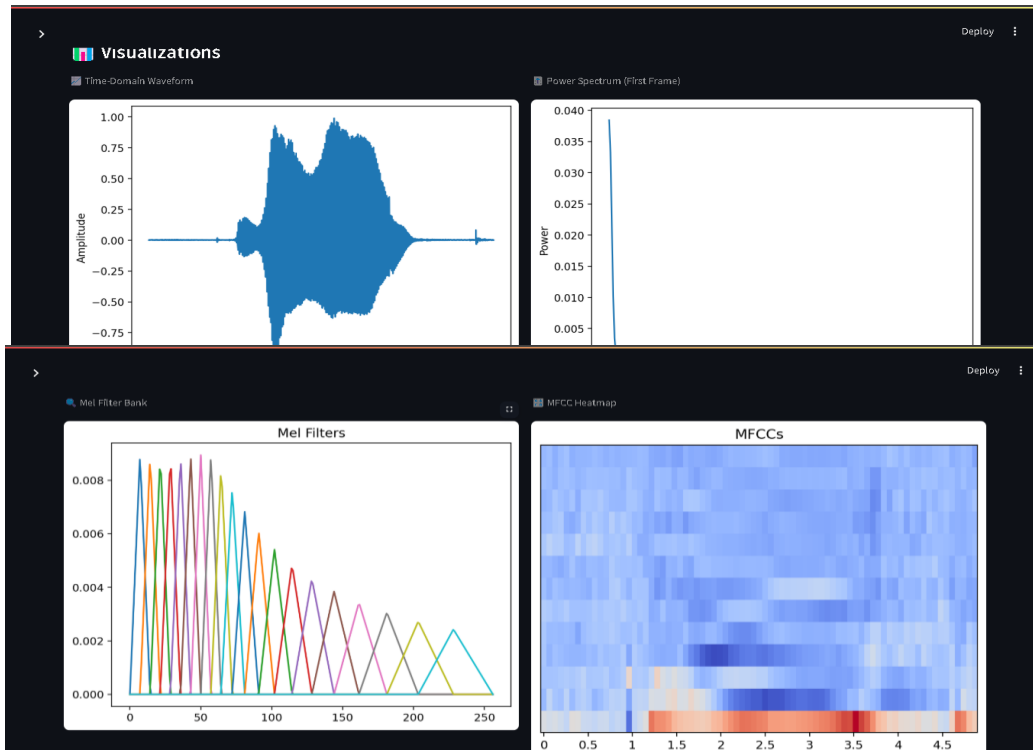
## 6. Screenshots

These are the screenshots of the app required to ensure that it is running properly and covers all the problems stated in the problem document: -


### 1. App running with MY name and Scholar ID



### 2. Waveform Plot



### 3. Parameter Menu

 Parameters

Frame Duration (ms)  

20

Overlap (%)  

25

Number of Mel filters  

20

MFCC coefficients  

8

13

20

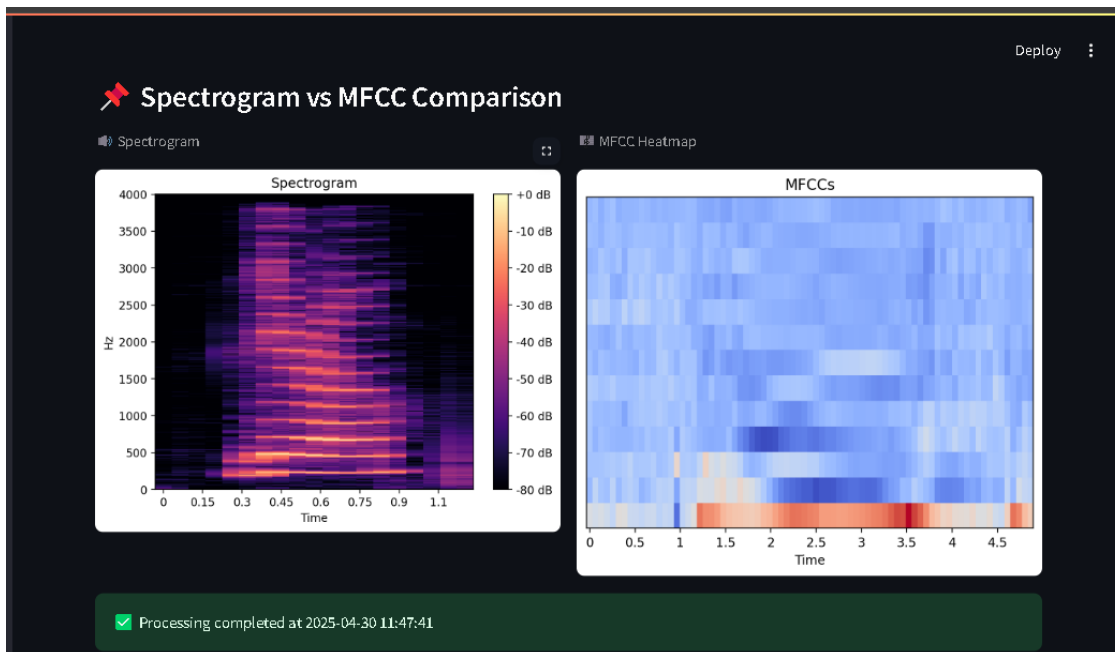
Resample Audio Rate (Hz)  

8000

Run ID: 20250430\_114735\_a9c9e7

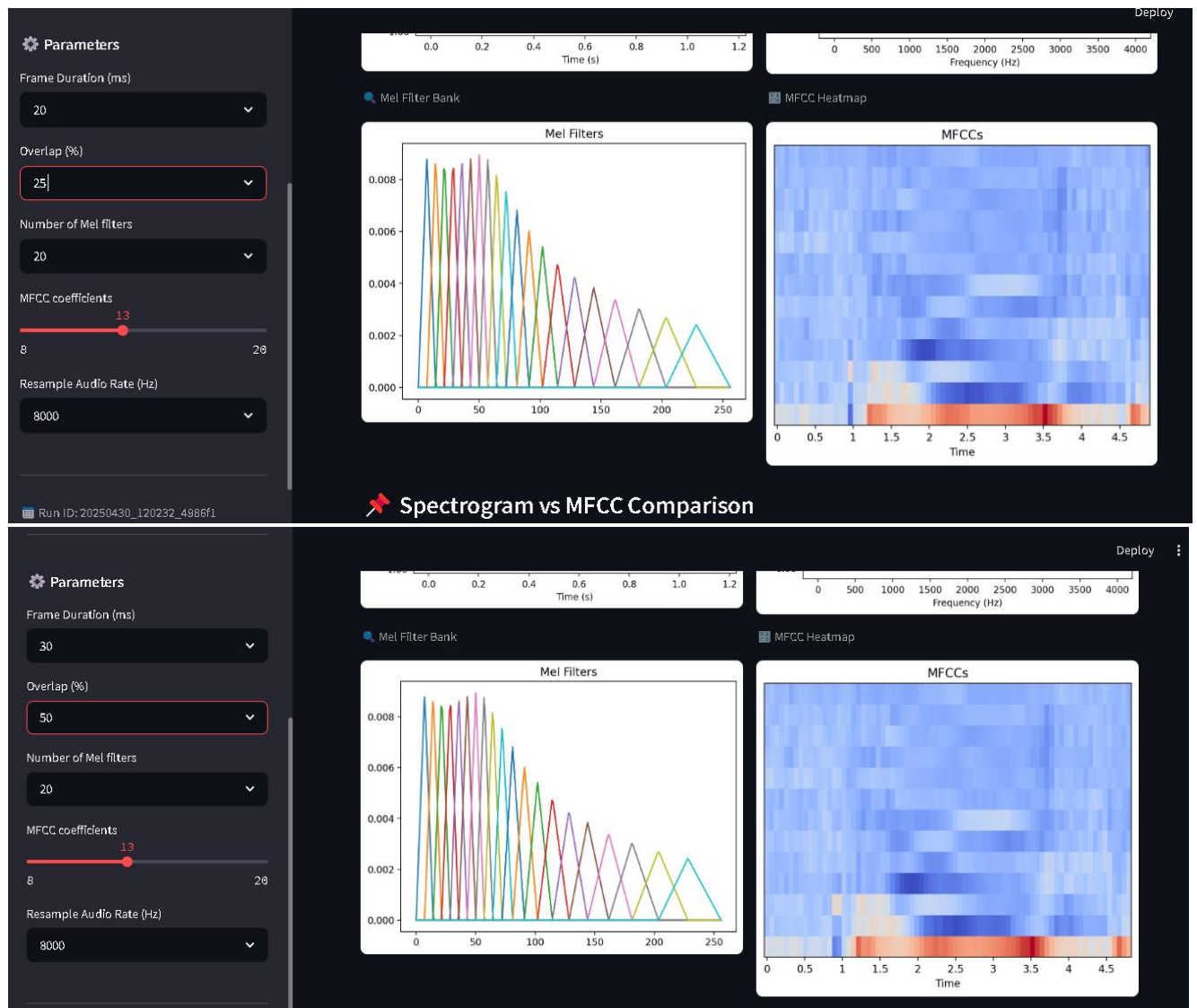
### 4. Experimenting with different frame sizes and sampling rates

#### Q1. Spectrogram vs. MFCC Interpretation



The spectrogram provides a dense time-frequency representation of the audio signal, highlighting all frequency components across time. In contrast, the MFCC heatmap presents a more compact and perceptually meaningful representation of the same audio by emphasizing frequencies important to human hearing. Upon visual comparison, it is evident that MFCCs smooth out finer spectral details while preserving speaker-specific features, making them more suitable for recognition tasks. The MFCC heatmap displays band-like structures aligned with phonetic variations, while the spectrogram captures transient and harmonic structures more directly.

## Q2. Interactive Frame Size and Overlap Analysis:



Changing the frame size affects the time-frequency resolution of the analysis. A smaller frame size (e.g., 20 ms) provides better temporal resolution but poorer frequency resolution, resulting in more variable MFCC patterns that closely track rapid speech

changes. Conversely, larger frame sizes (e.g., 50 ms) offer better frequency resolution at the cost of temporal precision, producing smoother and more stable MFCC contours. Similarly, increasing the overlap (e.g., from 25% to 75%) yields more densely sampled MFCC frames, which leads to smoother transitions between frames and improves the continuity of the feature map, especially useful in fast or fluent speech.

### Q3. Mel Filter Bank Customization and Visualization



Adjusting the number of mel filters directly impacts the granularity of frequency analysis. A lower filter count (e.g., 20) results in a broader and more generalized spectral view, leading to MFCCs with less detail. As the number increases to 30 or 40, the filter bank becomes denser, enabling finer resolution in the frequency domain and producing more detailed MFCCs. This enhances the system's ability to capture subtle variations in vocal characteristics, but also increases the dimensionality of the feature vectors, which could affect computation time and generalization in classification.

## Q4. Dynamic Sampling Rate Adjustment and MFCC Degradation Study



Altering the sampling rate affects both the fidelity of the audio and the quality of extracted features. Downsampling (e.g., to 8000 Hz) reduces the frequency range captured, leading to MFCCs that lose high-frequency details and may exhibit coarser patterns. Upsampling or retaining the original rate (e.g., 22050 Hz or above) preserves the full spectral range, resulting in clearer MFCCs with better resolution. Visually, degraded sampling rates produce flatter spectrograms and simplified MFCC maps, potentially reducing the distinctiveness of speaker-specific features in recognition tasks.

## 7. Conclusion

This project successfully implemented an MFCC-based feature extraction pipeline and a Streamlit-based visualization app. It demonstrates the impact of audio processing

Run ID: 20250430\_121826\_35e226

✓ Processing completed at 2025-04-30 12:18:29

parameters on MFCC quality and lays the foundation for future integration into a full speaker recognition system.