Comprehensive Voice Analysis using Signal Processing Techniques

Neelima Pulipati
210108032

November 1, 2024

1 Introduction

This report outlines a comprehensive voice analysis system designed to process and analyze audio signals. The system employs various signal processing techniques, including audio enhancement, pitch detection, formant extraction, and cepstral analysis, using libraries such as NumPy, SciPy, and librosa.

2 Methodology

The analysis workflow involves the following key steps:

2.1 Audio Signal Enhancement

The audio signal is first enhanced by converting stereo signals to mono, normalizing the audio, removing DC bias, and applying an emphasis filter. A speech-optimized bandpass filter (50Hz-10000Hz) is also utilized to improve the quality of the signal.

2.2 Audio Format Transformation

The audio files, originally in M4A format, are converted to WAV format for further processing. This involves standardizing audio parameters such as channels, frame rate, and sample width.

2.3 Spectrogram Visualization

The time-frequency characteristics of the processed audio are visualized using spectrograms. Two types of spectrograms are generated: a detailed narrow-band spectrogram and a wide-band overview spectrogram.

2.4 Pitch Analysis

Pitch analysis is performed using Average Magnitude Difference (AMD). The results are visualized to provide insights into the fundamental frequency of the voice signal.

2.5 Formant Extraction

The first three formants of the audio signal are extracted using Linear Predictive Coding (LPC) analysis. The mean formant frequencies are computed for further analysis.

2.6 Cepstral Analysis

Cepstral analysis is conducted to estimate the pitch of the audio signal. The cepstrum is calculated, and its visualization aids in understanding the underlying periodicities of the voice.

3 Results

The analysis was conducted on three audio files. The results for each file are summarized in the tables below.

3.1 Audio File 1: `aa.m4a`

Table 1: Analysis Results for `aa.m4a`

Parameter	Value
Mean Pitch	117.9 Hz
Median Pitch	117.3 Hz
F1	759.0 Hz
F2	1211.2 Hz
F3	1963.3 Hz
Estimated Pitch from Cepstral Analysis	234.6 Hz

3.2 Audio File 2: `ii.m4a`

Table 2: Analysis Results for `ii.m4a`

Parameter	Value
Mean Pitch	127.7 Hz
Median Pitch	127.9 Hz
F1	293.7 Hz
F2	544.4 Hz
F3	2805.9 Hz
Estimated Pitch from Cepstral Analysis	254.9 Hz

3.3 Audio File 3: `uu.m4a`

Table 3: Analysis Results for `uu.m4a`

Parameter	Value
Mean Pitch	153.8 Hz
Median Pitch	155.6 Hz
F1	314.2 Hz
F2	641.7 Hz
F3	1886.9 Hz
Estimated Pitch from Cepstral Analysis	234.6 Hz

4 Conclusion

This voice analysis system successfully processes audio signals and extracts relevant acoustic features. The methodologies implemented provide a robust framework for future research and applications in speech processing and analysis.

Repo - https://github.com/Neelima-Pulipati/Speech-Signal-Processing.git