Analysis of different voice notes

1. Introduction

This report details the analysis of 10 different audio recordings using various speech signal processing techniques. The goal is to examine different speech characteristics such as amplitude, pitch, RMS energy, zero-crossing rate (ZCR), mel-frequency cepstral coefficients (MFCCs), fundamental frequency (f0), and spectrograms. The dataset includes audio samples with varied pitch, volume, and speaking styles to analyze their effects on the extracted features.

---

2. Methodology

2.1 Dataset Description

The dataset consists of 10 different speech recordings:

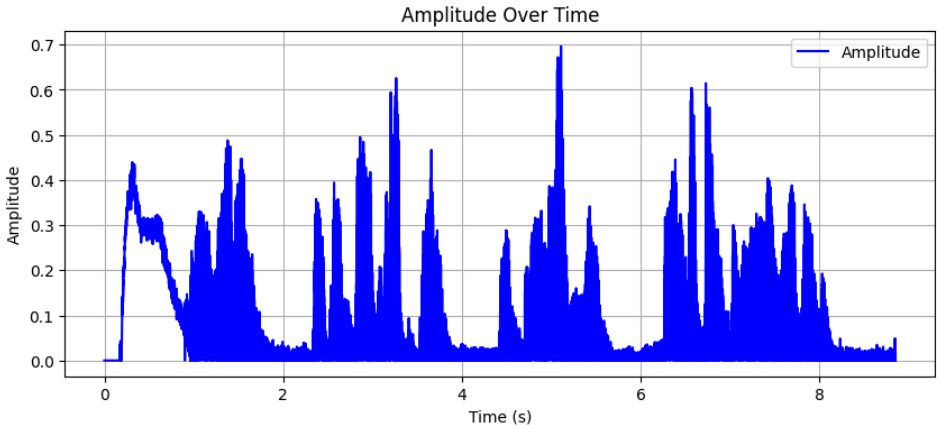
1. **News Report** (Neutral, Moderate Volume, Medium Pitch)
2. **Excited Announcement** (High Pitch, Loud Volume)
3. **Whispered Secret** (Low Pitch, Soft Volume)
4. **Formal Speech** (Medium Pitch, Moderate Volume, Slow Pace)
5. **Angry Complaint** (Low Pitch, Loud Volume, Fast Pace)
6. **Childlike Excitement** (High Pitch, Soft Volume, Fast Pace)
7. **Robotic/Monotone** (Flat Pitch, Moderate Volume, Even Pace)
8. **Dramatic Storytelling** (Varied Pitch and Volume, Expressive Tone)
9. **Relaxing Meditation Guide** (Low Pitch, Soft Volume, Slow Pace)
10. **Sarcastic Remark** (Medium Pitch, Moderate Volume, Slow Drawl)

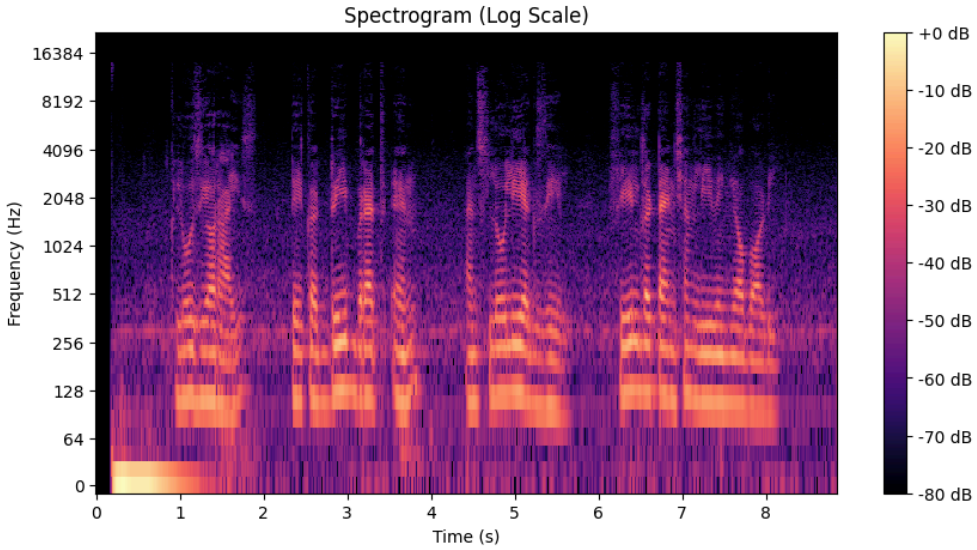
2.2 Processing Steps

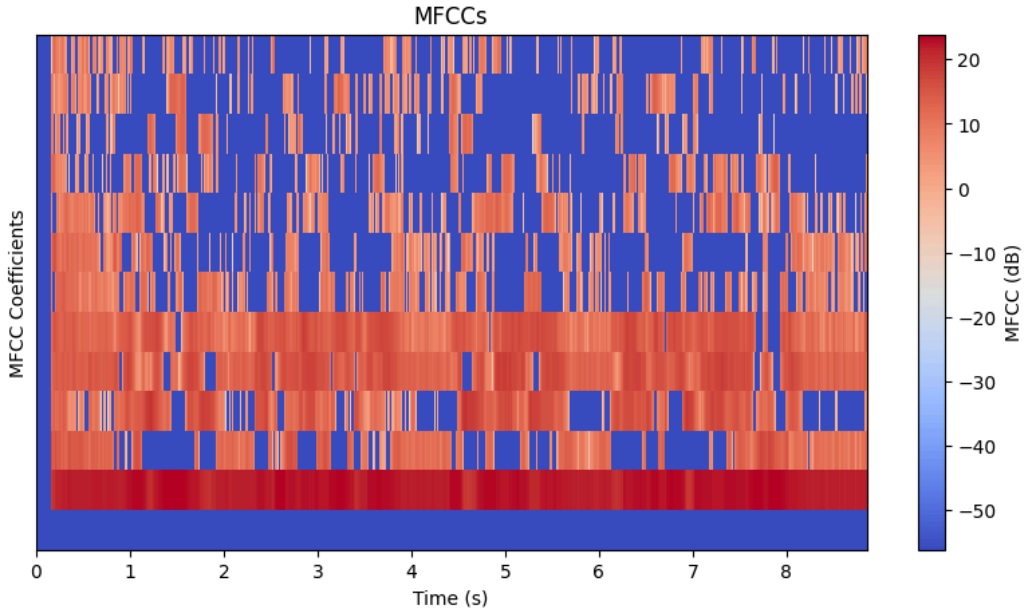
1. Load the audio file and extract the sampling rate.
2. Compute various speech features:
    - **Amplitude**: Extracted from the absolute values of the waveform.
    - **Pitch**: Extracted using `librosa.piptrack()`.
    - **RMS Energy**: Computed using `librosa.feature.rms()`.
    - **Zero-Crossing Rate (ZCR)**: Derived from `librosa.feature.zero_crossing_rate()`.
    - **MFCCs**: Extracted using `librosa.feature.mfcc()`.
    - **Fundamental Frequency (f0)**: Computed with `librosa.pyin()`.
    - **Spectrogram**: Generated using `librosa.stft()` and converted to a decibel scale.

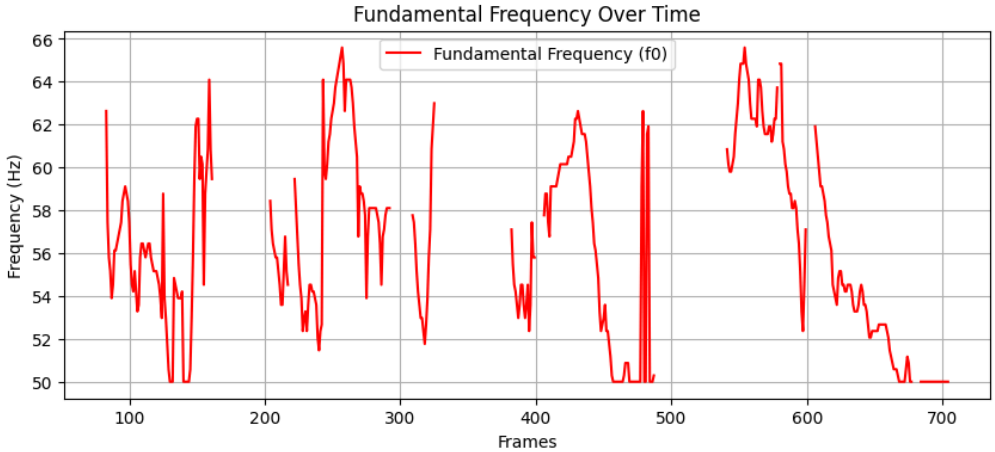
3. Results

Processing: Relaxing Meditation Guide (Low Pitch, Soft Volume, Slow Pace).wav









Pitch Mean: 297.19573974609375 Hz

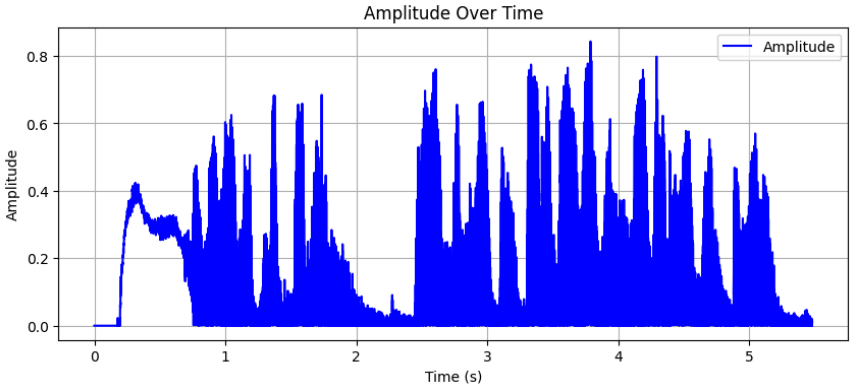
RMS Energy Mean: 0.0891302078962326

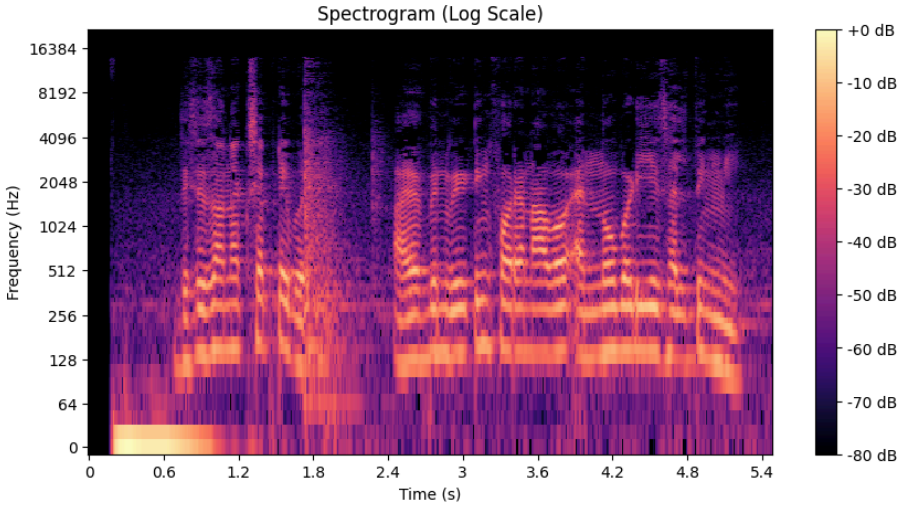
Zero-Crossing Rate Mean: 0.0237069186189384

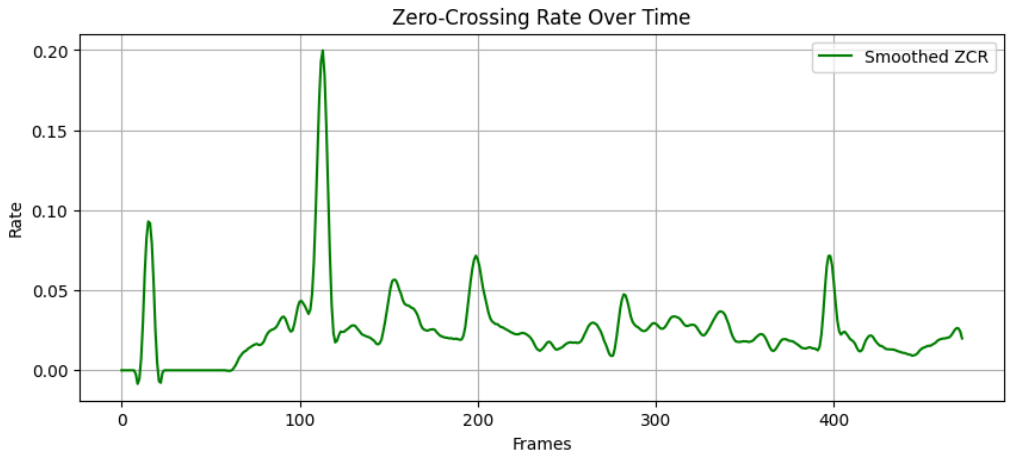
MFCCs Shape: (13, 763)

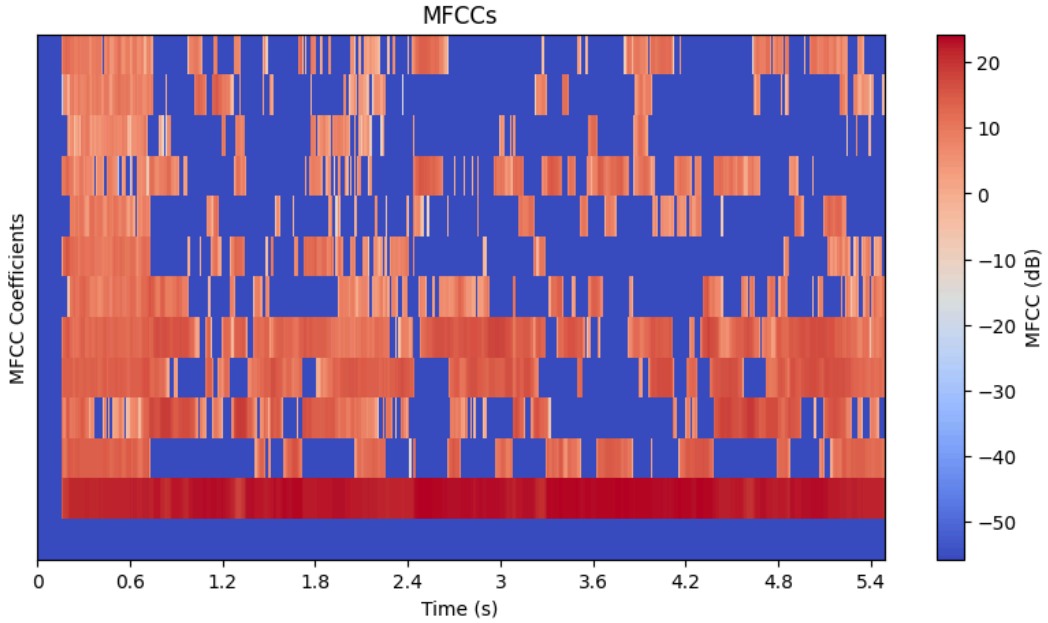
Mean Fundamental Frequency (f0): 56.23263115073206 Hz

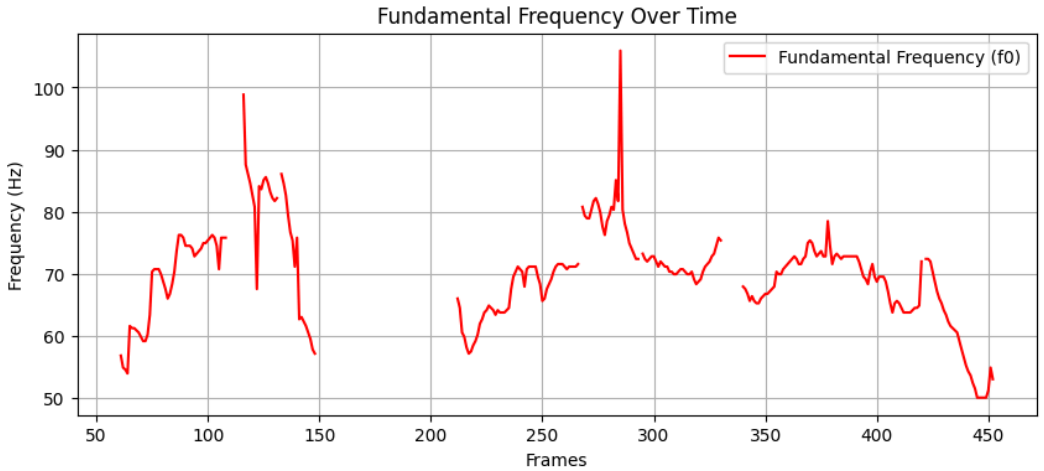
Processing: Angry Complaint (Low Pitch, Loud Volume, Fast Pace).wav






Zero-Crossing Rate Over Time
Smoothed ZCR
Rate
Frames
0.20
0.15
0.10
0.05
0.00
0
100
200
300
400


MFCCs
MFCC Coefficients
Time (s)
MFCC (dB)
0
0.6
1.2
1.8
2.4
3
3.6
4.2
4.8
5.4
20
10
0
−10
−20
−30
−40
−50


Fundamental Frequency Over Time
Fundamental Frequency (f0)
Frequency (Hz)
Frames
100
90
80
70
60
50
50
100
150
200
250
300
350
400
450

Pitch Mean: 389.0041198730469 Hz

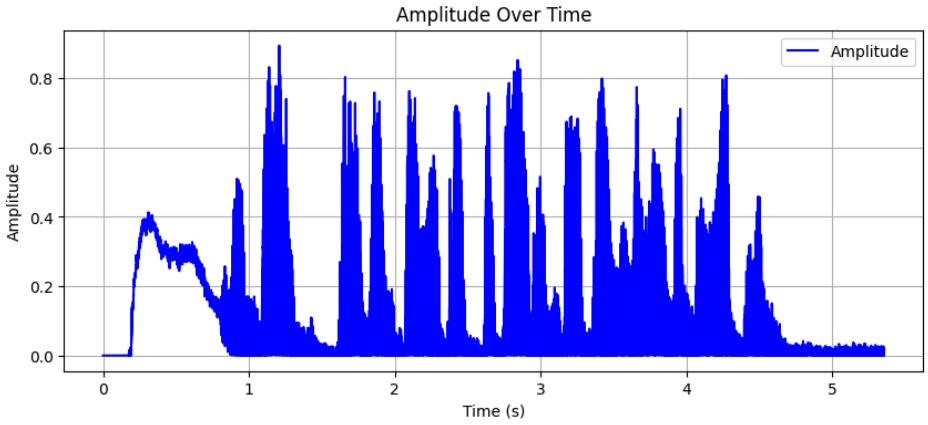
RMS Energy Mean: 0.14974799752235413

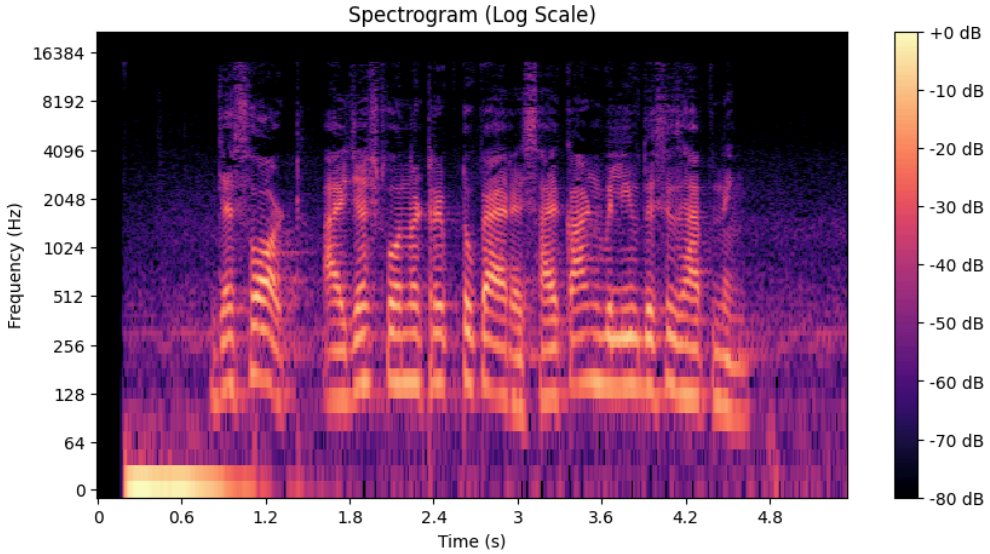
Zero-Crossing Rate Mean: 0.024840405325052856

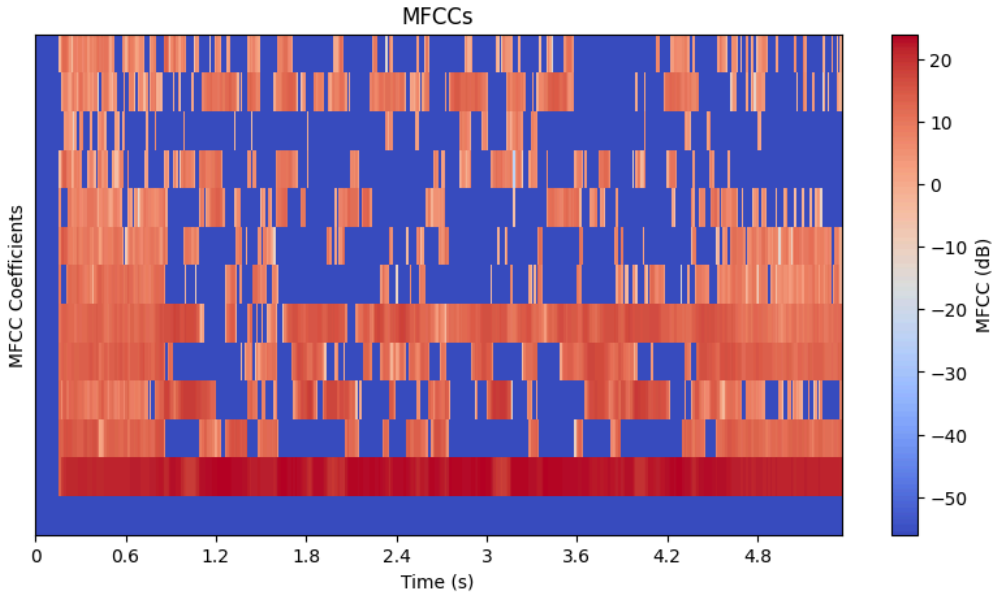
MFCCs Shape: (13, 473)

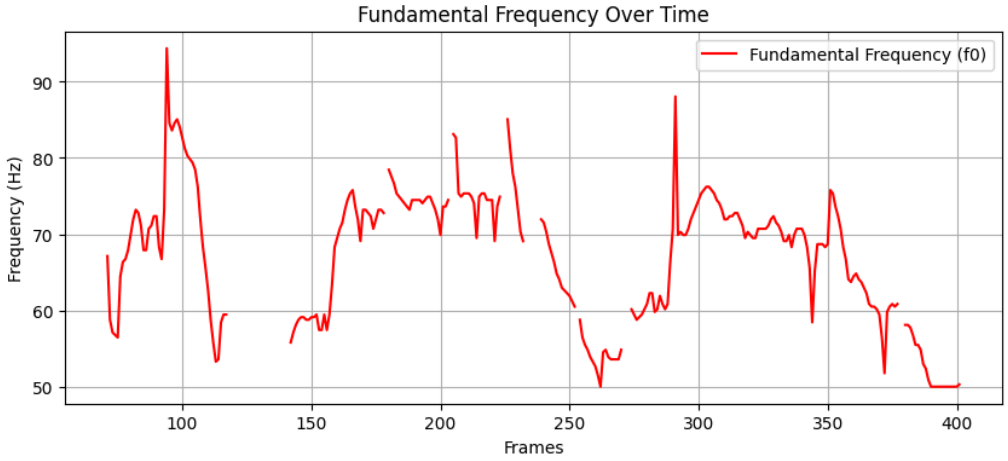
Mean Fundamental Frequency (f0): 69.7042731550207 Hz

Processing: Excited Announcement (High Pitch, Loud Volume).wav









Pitch Mean: 364.42547607421875 Hz

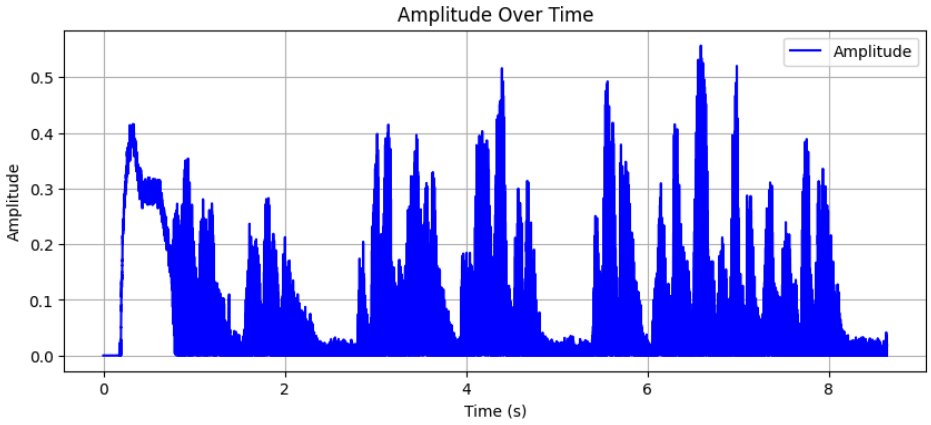
RMS Energy Mean: 0.1334761530160904

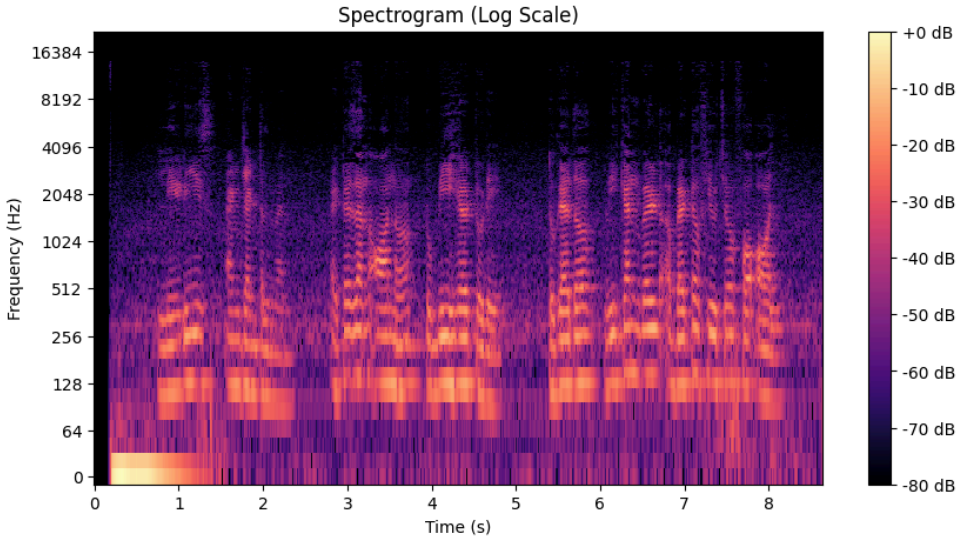
Zero-Crossing Rate Mean: 0.031114718614718616

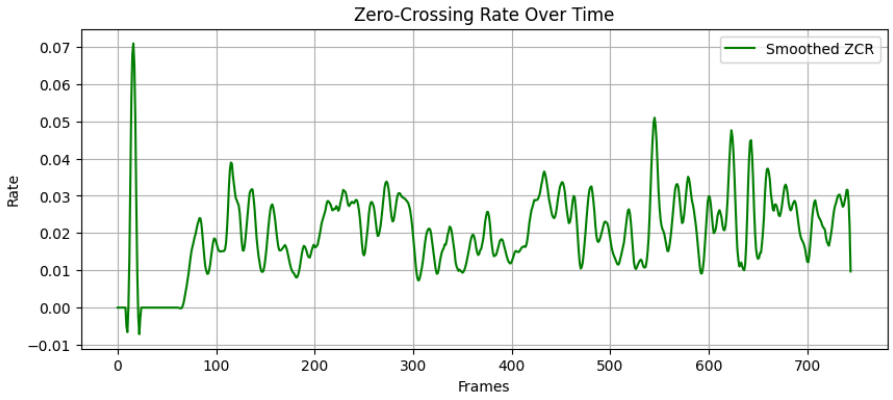
MFCCs Shape: (13, 462)

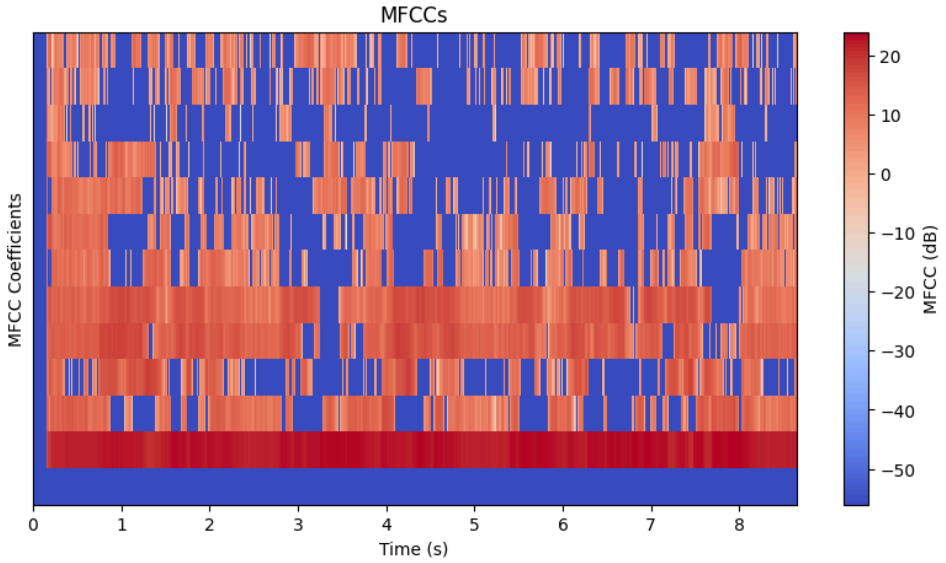
Mean Fundamental Frequency (f0): 67.06374495619055 Hz

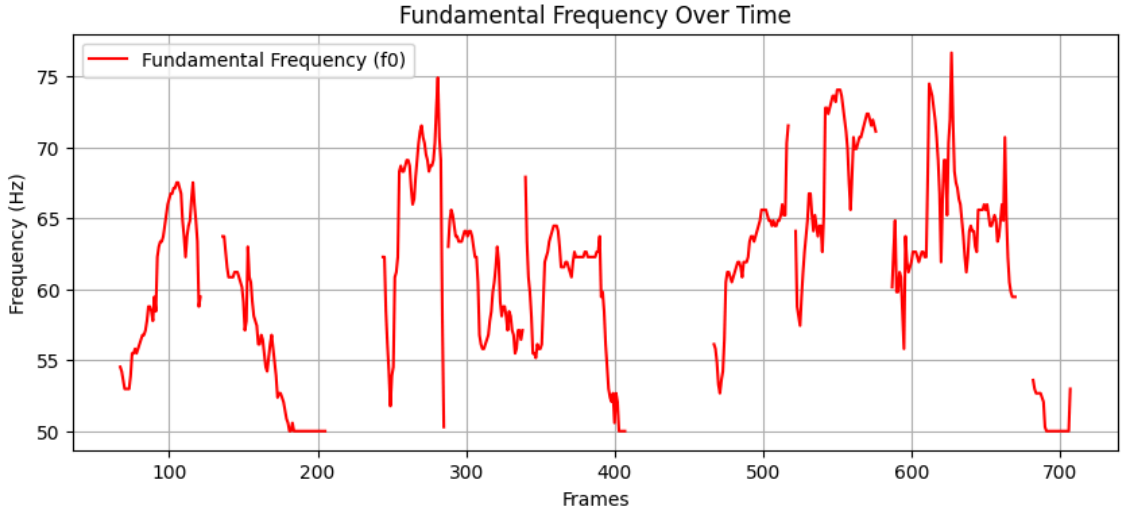
Processing: Formal Speech (Medium Pitch, Moderate Volume, Slow Pace).wav






Zero-Crossing Rate Over Time
Smoothed ZCR
Rate
Frames


MFCCs
MFCC Coefficients
Time (s)
MFCC (dB)


Fundamental Frequency Over Time
Fundamental Frequency (f0)
Frequency (Hz)
Frames

Pitch Mean: 308.978759765625 Hz

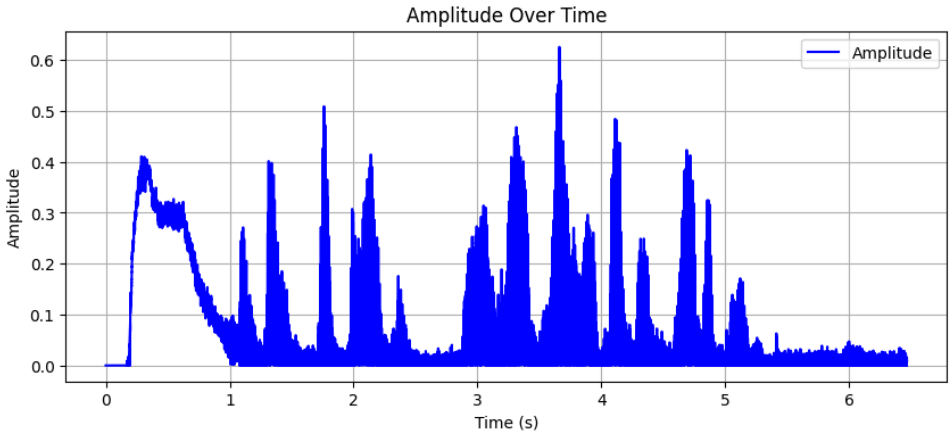
RMS Energy Mean: 0.08348973095417023

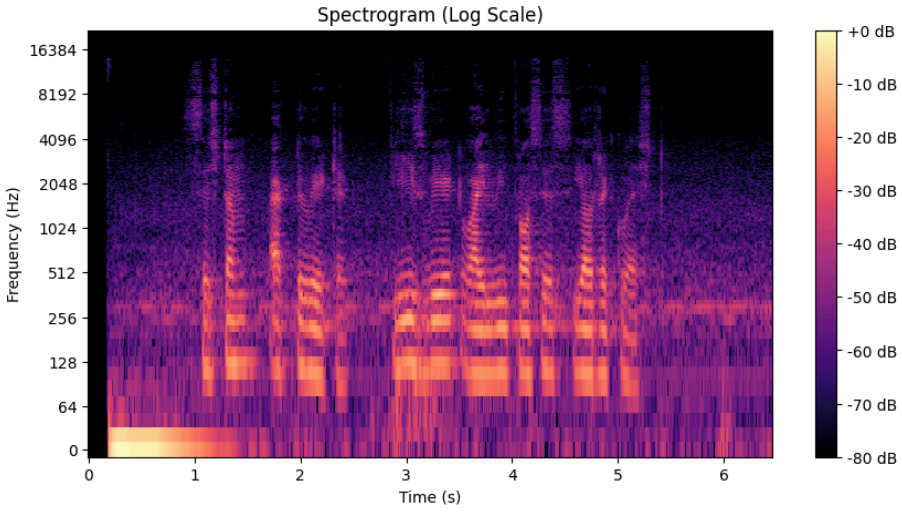
Zero-Crossing Rate Mean: 0.020229918204697987

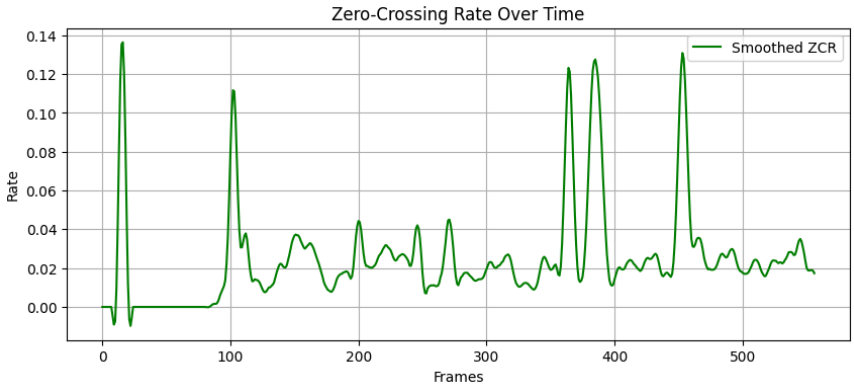
MFCCs Shape: (13, 745)

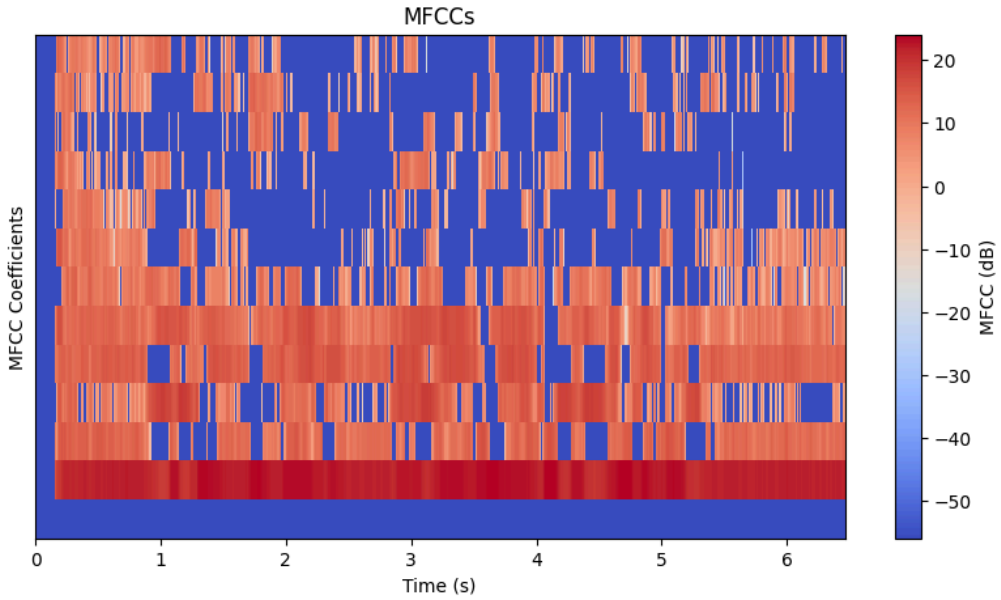
Mean Fundamental Frequency (f0): 61.290164354424114 Hz

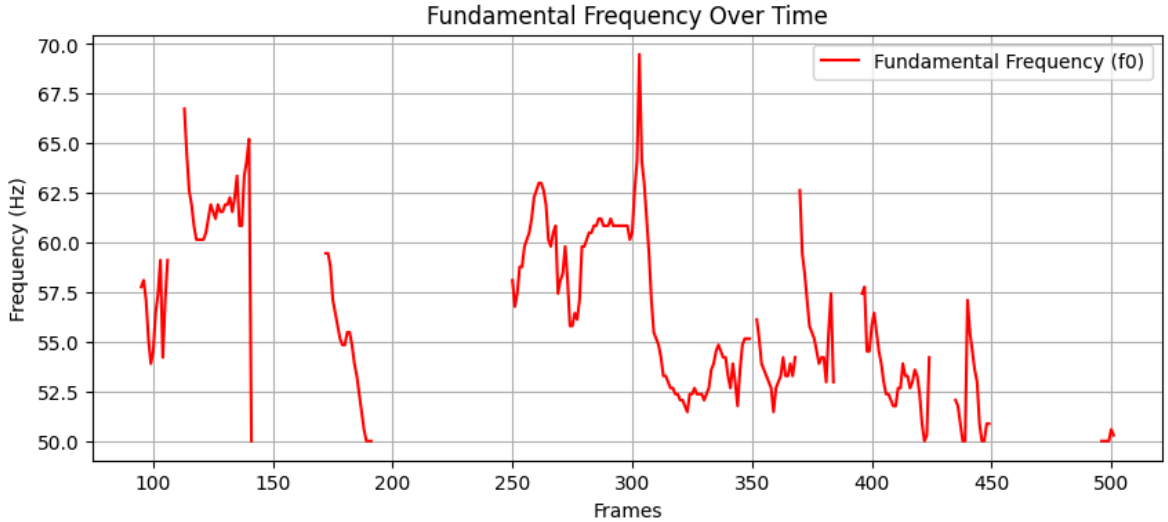
Processing: RoboticMonotone (Flat Pitch, Moderate Volume, Even Pace).wav






Zero-Crossing Rate Over Time
Smoothed ZCR
Rate
Frames
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
100
200
300
400
500


MFCCs
MFCC Coefficients
Time (s)
MFCC (dB)
0
1
2
3
4
5
6
20
10
0
−10
−20
−30
−40
−50


Fundamental Frequency Over Time
Fundamental Frequency (f0)
Frequency (Hz)
Frames
70.0
67.5
65.0
62.5
60.0
57.5
55.0
52.5
50.0
100
150
200
250
300
350
400
450
500

Pitch Mean: 316.6019592285156 Hz

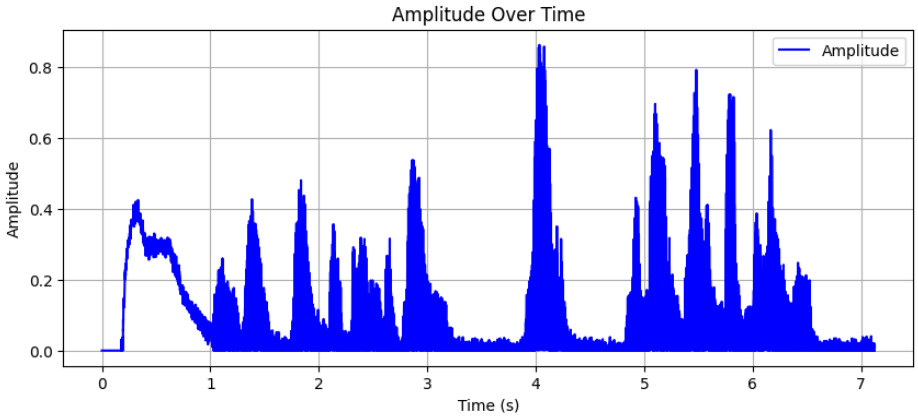
RMS Energy Mean: 0.07597647607326508

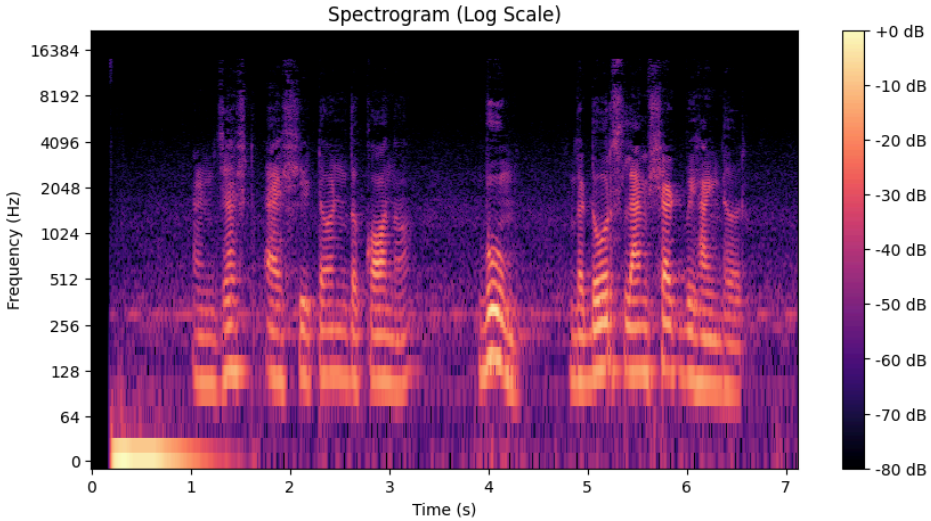
Zero-Crossing Rate Mean: 0.02512413038599641

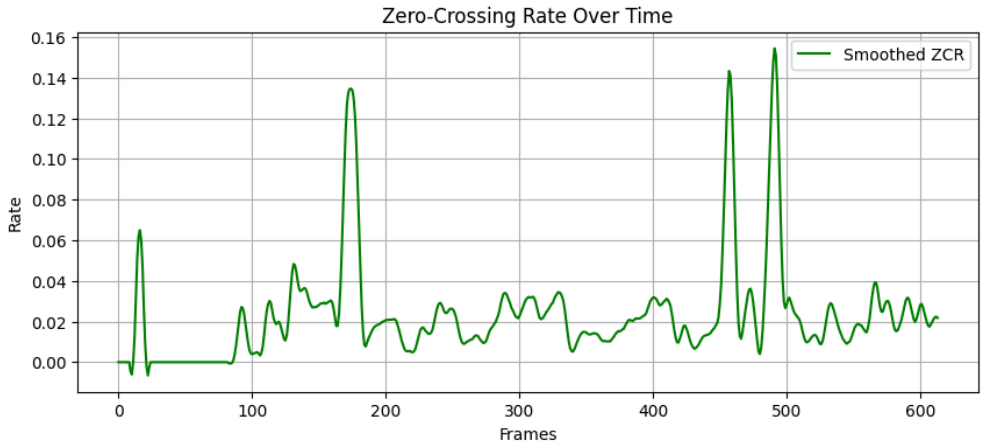
MFCCs Shape: (13, 557)

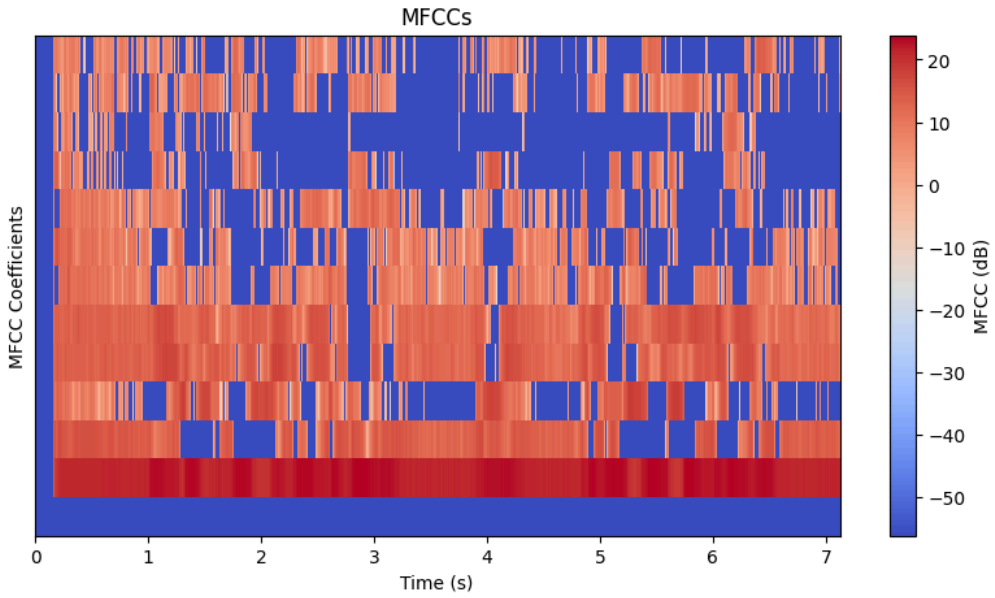
Mean Fundamental Frequency (f0): 56.30224794315477 Hz

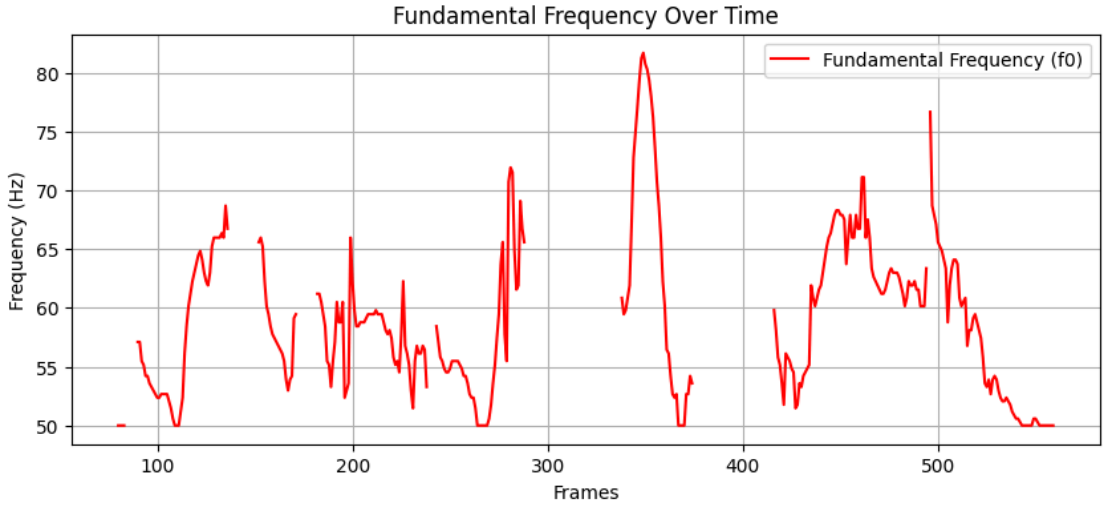
Processing: Dramatic Storytelling (Varied Pitch and Volume, Expressive Tone).wav






Zero-Crossing Rate Over Time
Smoothed ZCR
Rate
Frames
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
100
200
300
400
500
600


MFCCs
MFCC Coefficients
Time (s)
MFCC (dB)
0
1
2
3
4
5
6
7
20
10
0
−10
−20
−30
−40
−50


Fundamental Frequency Over Time
Fundamental Frequency (f0)
Frequency (Hz)
Frames
80
75
70
65
60
55
50
100
200
300
400
500

Pitch Mean: 333.0775451660156 Hz

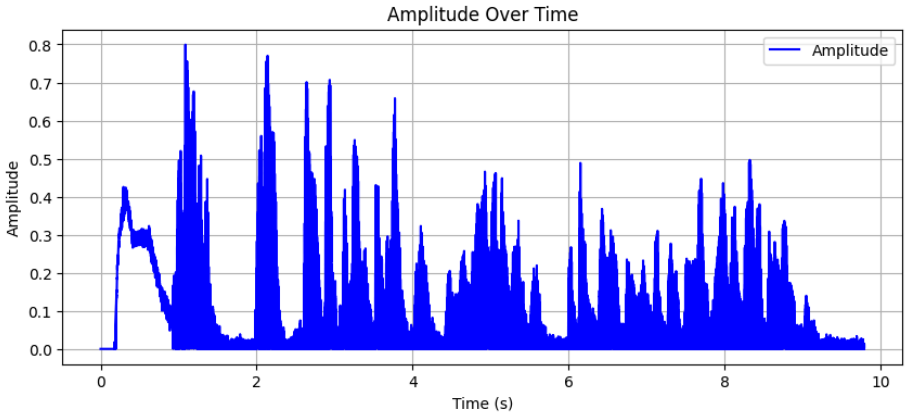
RMS Energy Mean: 0.0969623401761055

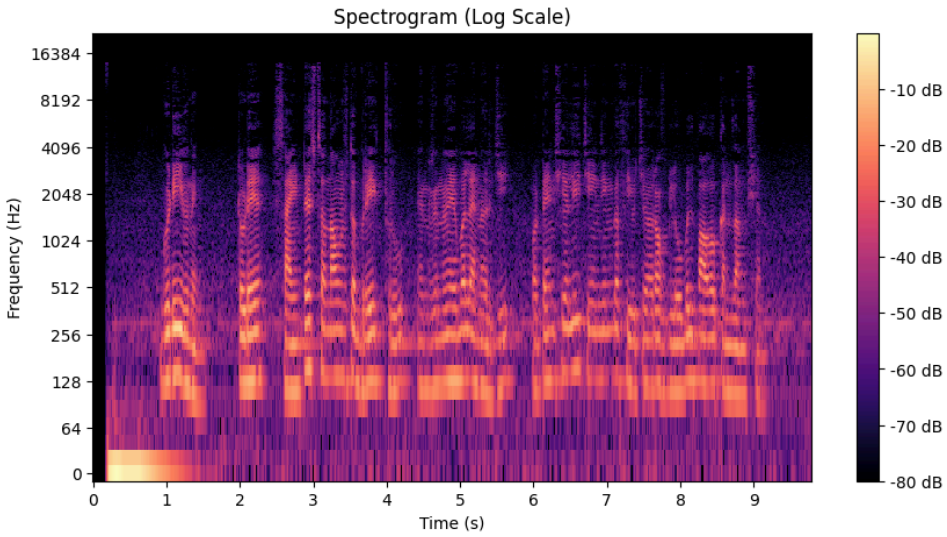
Zero-Crossing Rate Mean: 0.023141668363192182

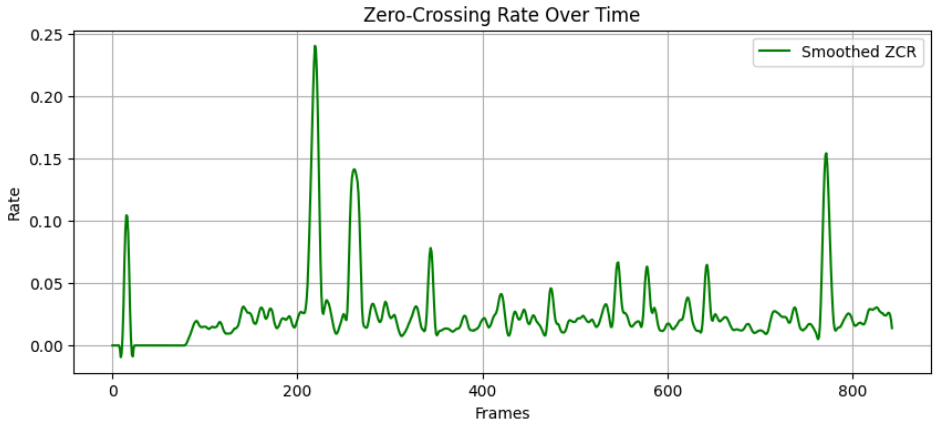
MFCCs Shape: (13, 614)

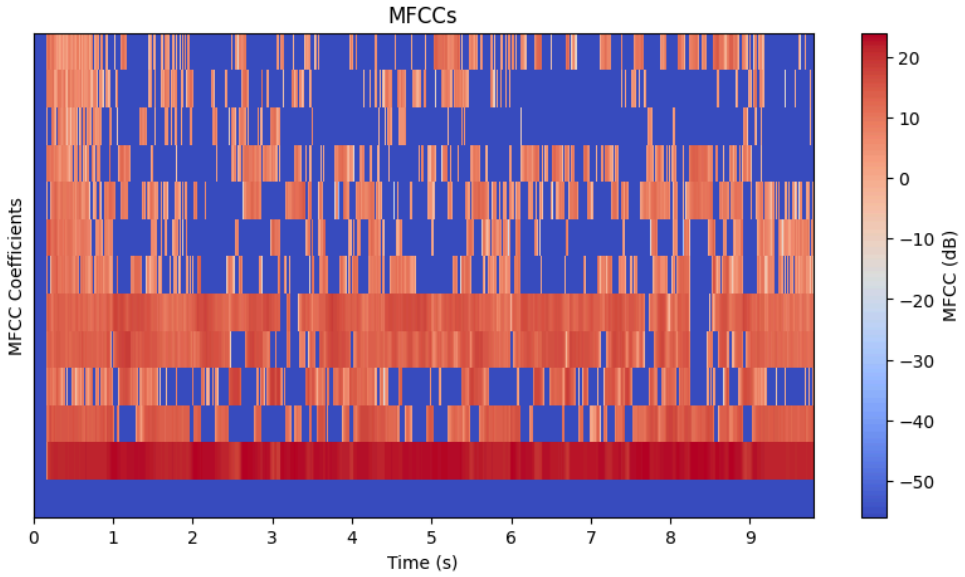
Mean Fundamental Frequency (f0): 58.886485128796636 Hz

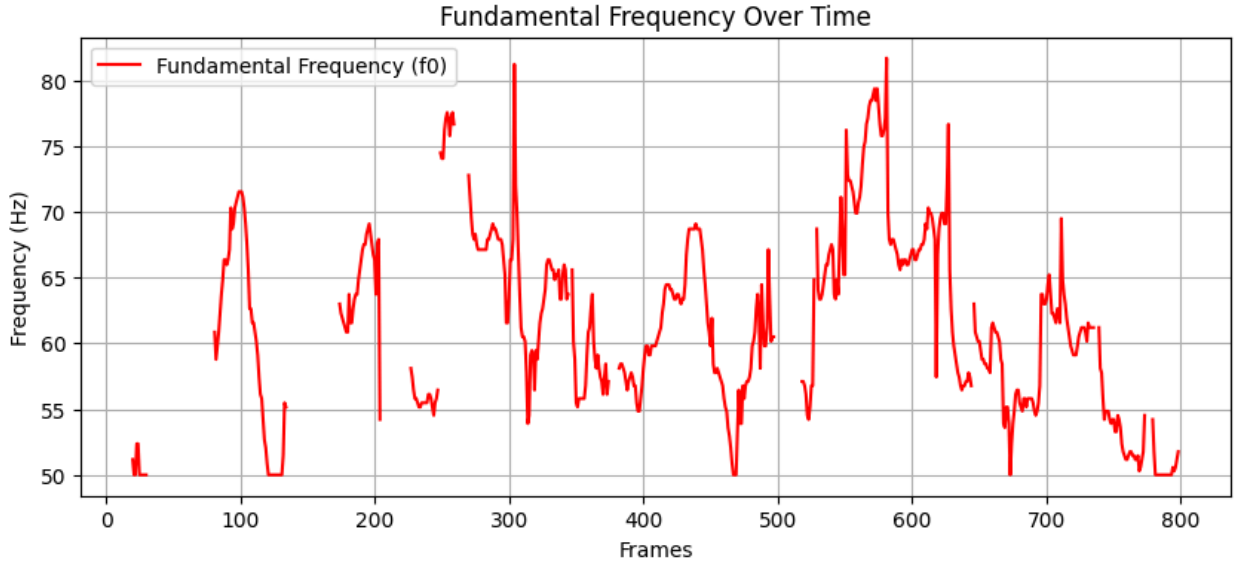
Processing: News Report (Neutral, Moderate Volume, Medium Pitch).wav






Zero-Crossing Rate Over Time
Smoothed ZCR
Rate
Frames
0.25
0.20
0.15
0.10
0.05
0.00
0
200
400
600
800


MFCCs
MFCC Coefficients
Time (s)
MFCC (dB)
0
1
2
3
4
5
6
7
8
9
20
10
0
−10
−20
−30
−40
−50


Fundamental Frequency Over Time
Fundamental Frequency (f0)
Frequency (Hz)
Frames
80
75
70
65
60
55
50
0
100
200
300
400
500
600
700
800

Pitch Mean: 312.0110168457031 Hz

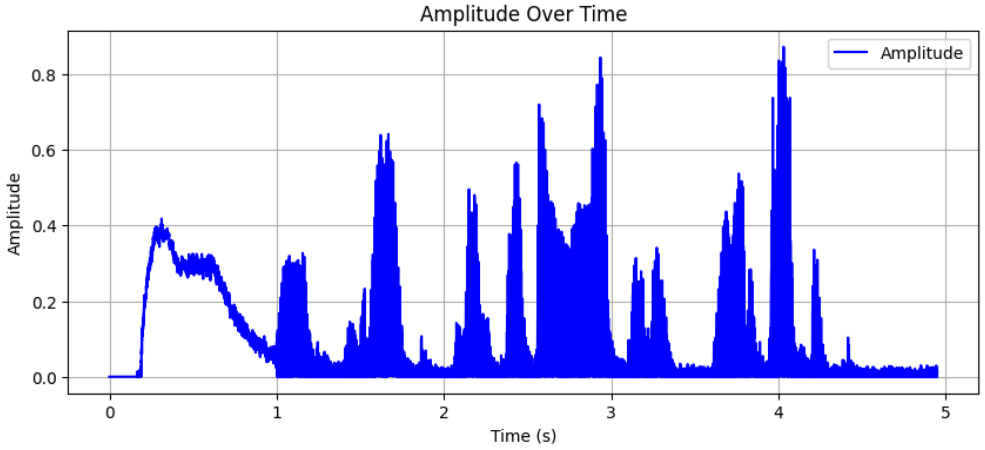
RMS Energy Mean: 0.09271269291639328

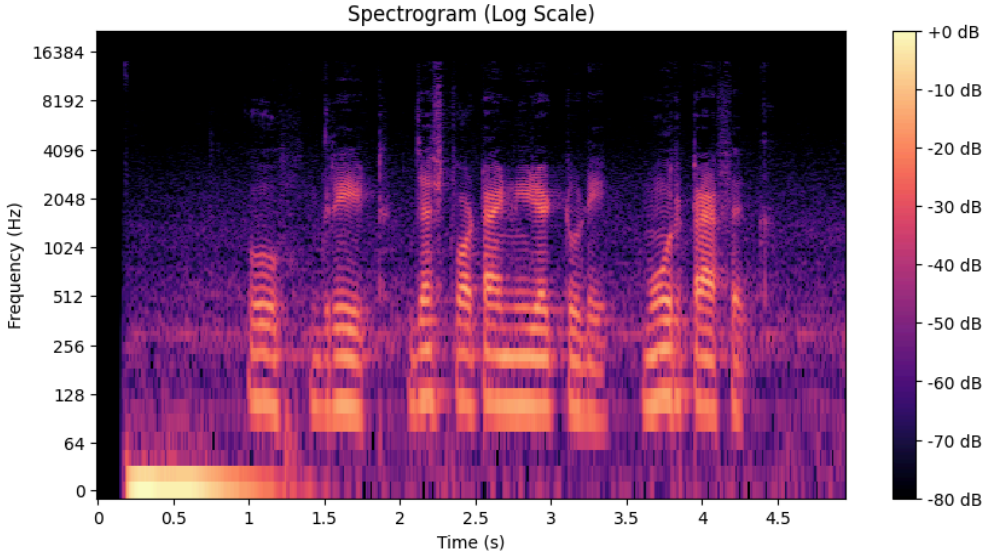
Zero-Crossing Rate Mean: 0.02495151899437204

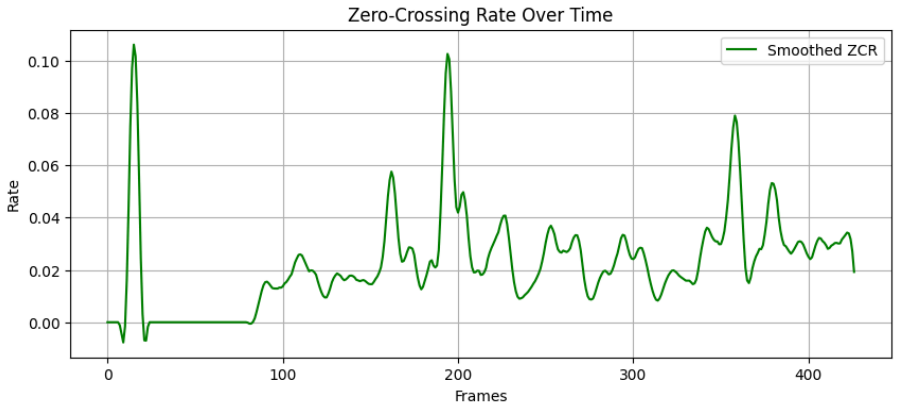
MFCCs Shape: (13, 844)

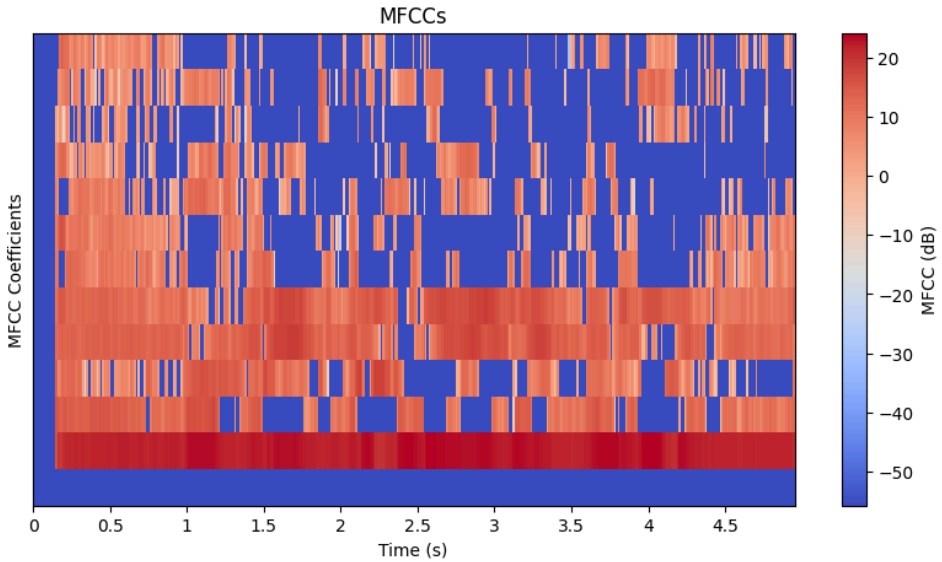
Mean Fundamental Frequency (f0): 61.49005287838933 Hz

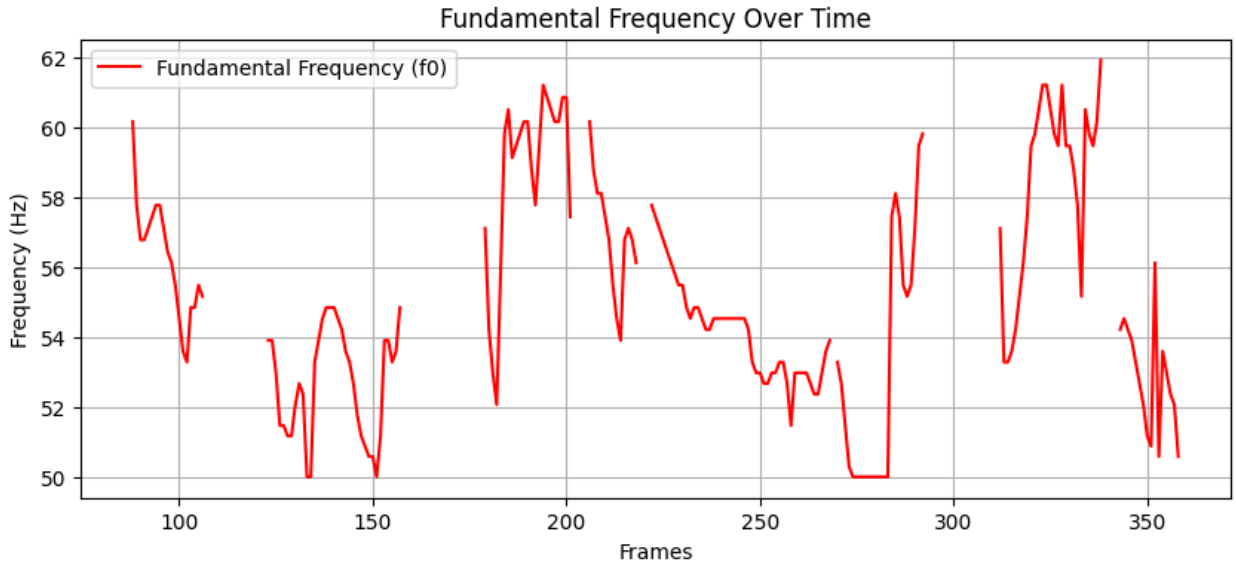
Processing: Sarcastic Remark (Medium Pitch, Moderate Volume, Slow Drawl).wav






Zero-Crossing Rate Over Time
Smoothed ZCR
Rate
Frames
0.10
0.08
0.06
0.04
0.02
0.00
0
100
200
300
400


MFCCs
MFCC Coefficients
Time (s)
MFCC (dB)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
20
10
0
−10
−20
−30
−40
−50


Fundamental Frequency Over Time
Fundamental Frequency (f0)
Frequency (Hz)
Frames
62
60
58
56
54
52
50
100
150
200
250
300
350

Pitch Mean: 338.747314453125 Hz

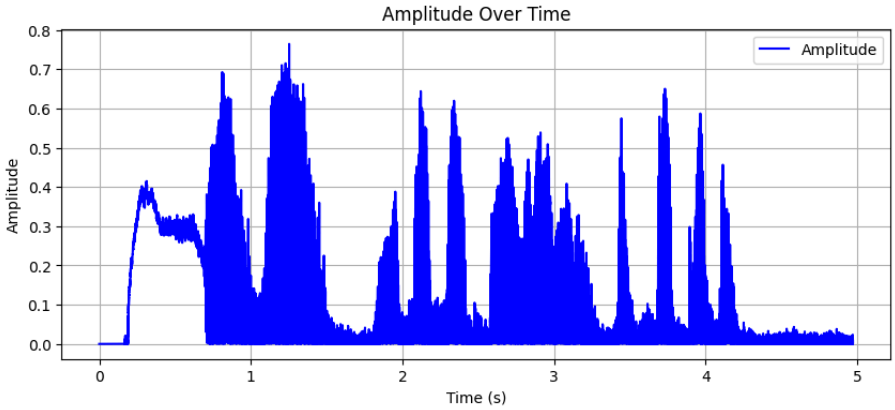
RMS Energy Mean: 0.09930200129747391

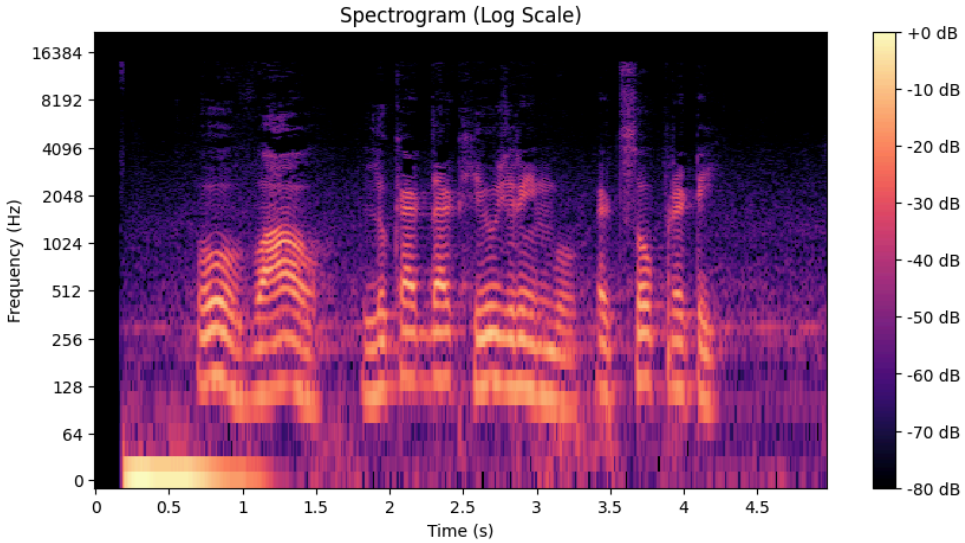
Zero-Crossing Rate Mean: 0.022513539227166278

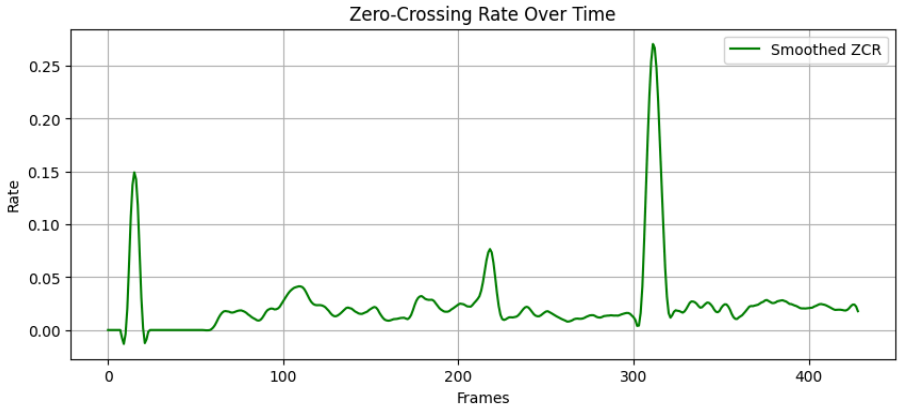
MFCCs Shape: (13, 427)

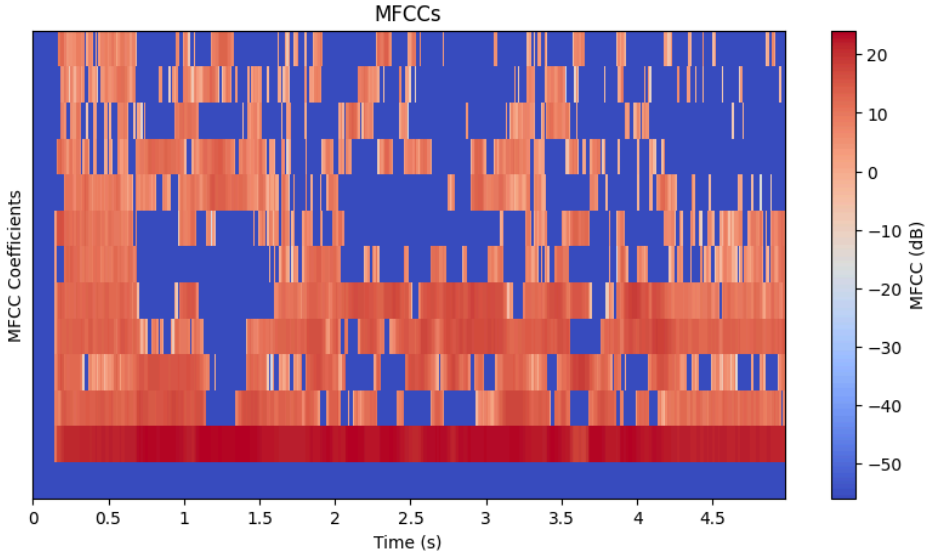
Mean Fundamental Frequency (f0): 55.12937744652043 Hz

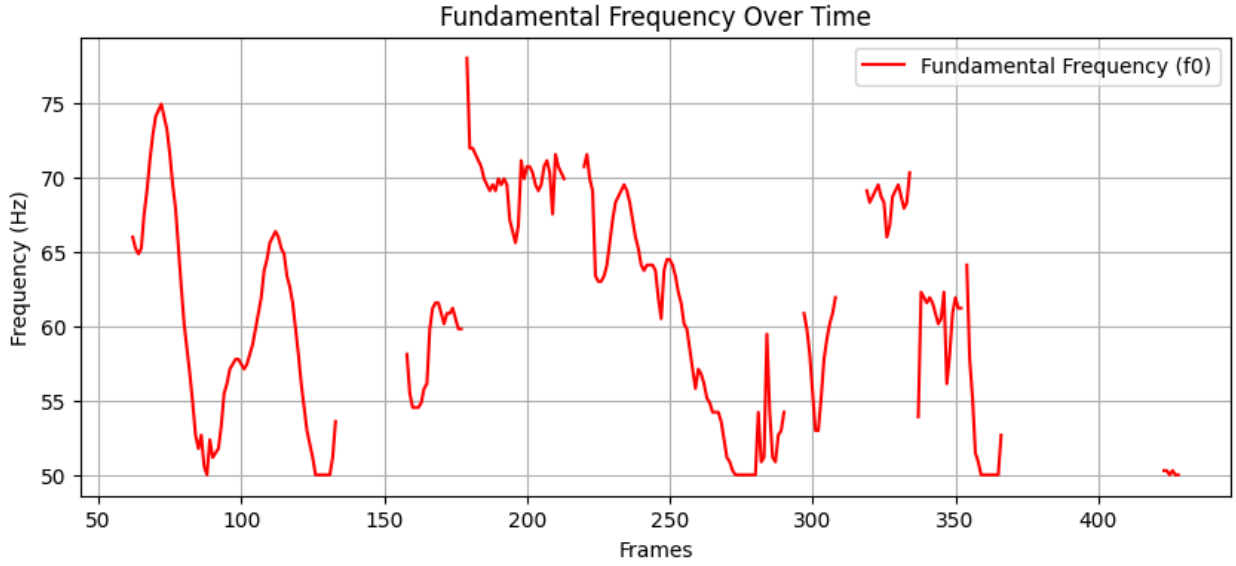
Processing: Childlike Excitement (High Pitch, Soft Volume, Fast Pace).wav






Zero-Crossing Rate Over Time
Smoothed ZCR
Rate
Frames
0.25
0.20
0.15
0.10
0.05
0.00
0
100
200
300
400


MFCCs
MFCC Coefficients
Time (s)
MFCC (dB)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
20
10
0
−10
−20
−30
−40
−50


Fundamental Frequency Over Time
Fundamental Frequency (f0)
Frequency (Hz)
Frames
75
70
65
60
55
50
50
100
150
200
250
300
350
400

Pitch Mean: 343.05120849609375 Hz

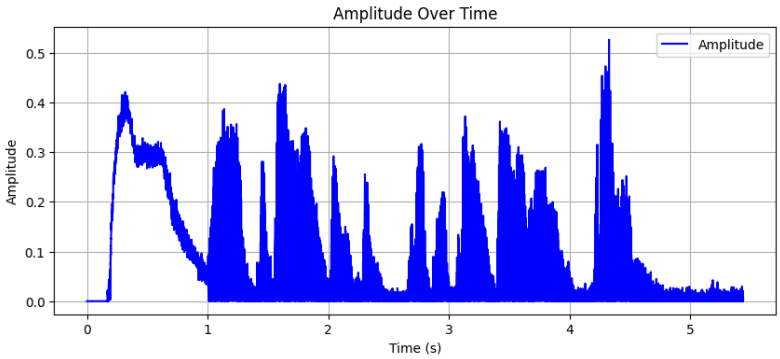
RMS Energy Mean: 0.11976262181997299

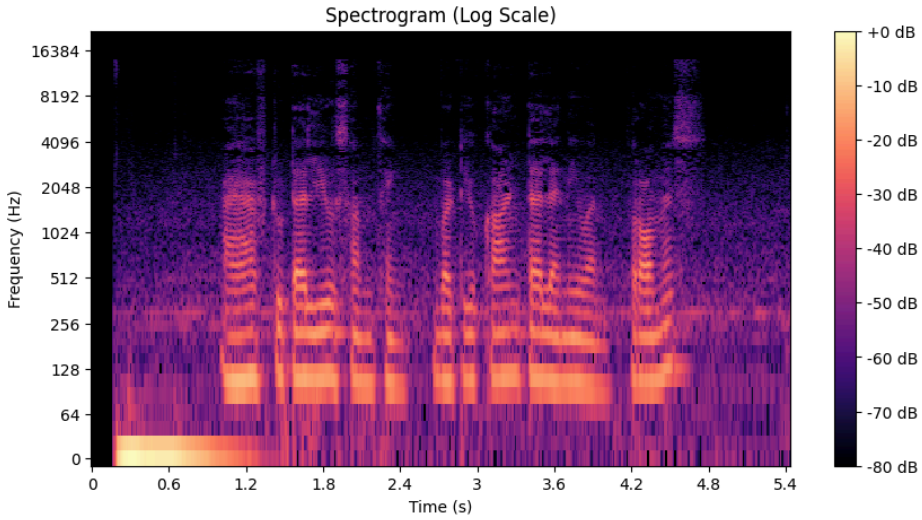
Zero-Crossing Rate Mean: 0.024112443546037296

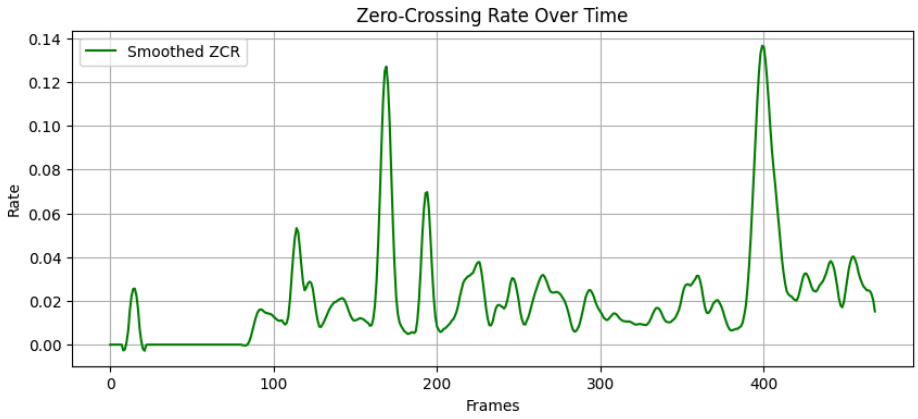
MFCCs Shape: (13, 429)

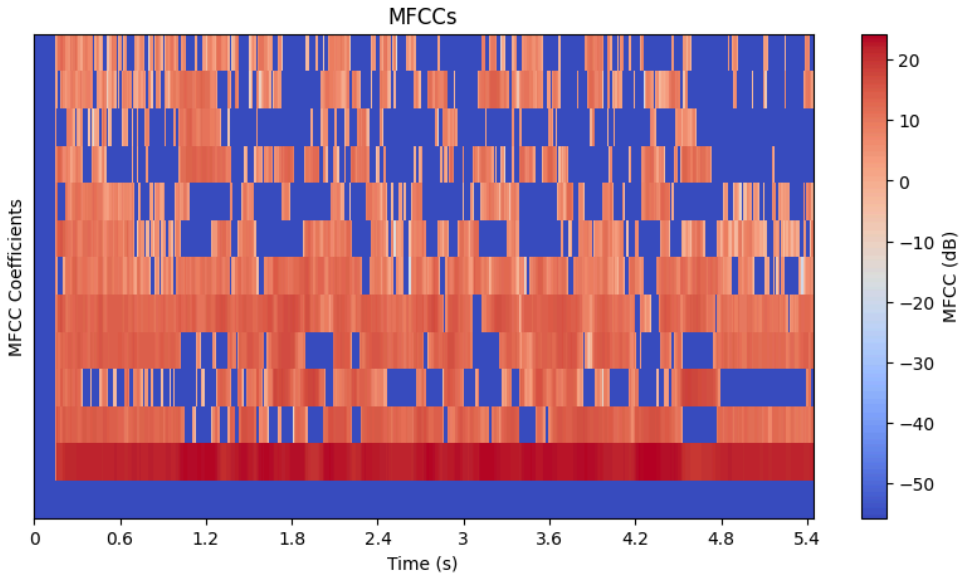
Mean Fundamental Frequency (f0): 60.87752016537706 Hz

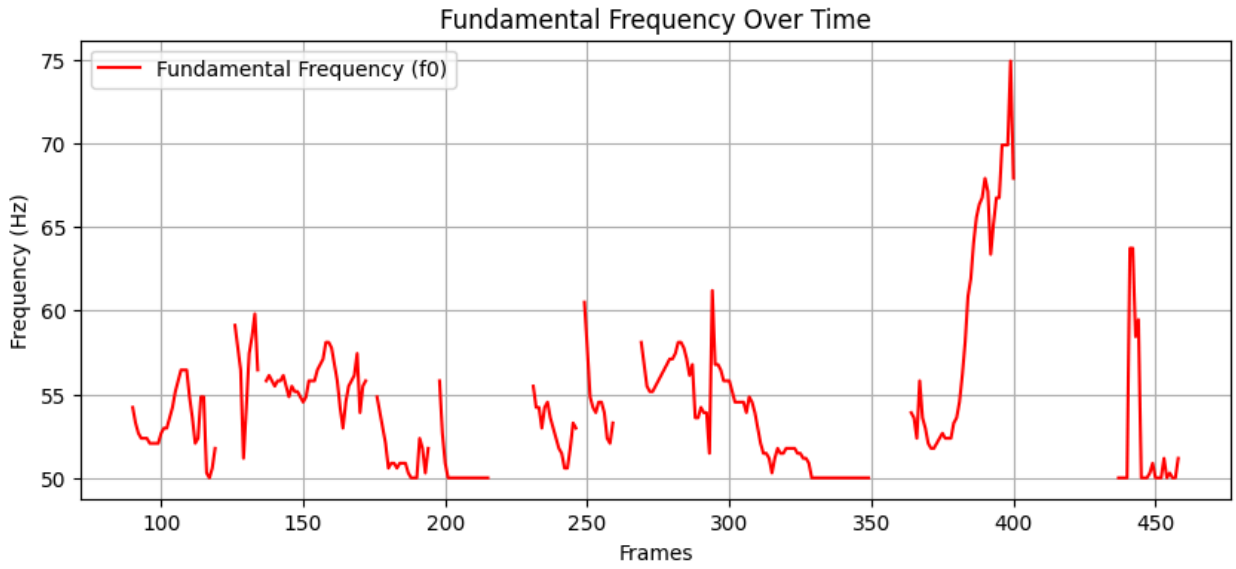
Processing: Whispered Secret (Low Pitch, Soft Volume).wav






Zero-Crossing Rate Over Time
Smoothed ZCR
Rate
Frames
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
100
200
300
400


MFCCs
MFCC Coefficients
Time (s)
MFCC (dB)
0
0.6
1.2
1.8
2.4
3
3.6
4.2
4.8
5.4
20
10
0
−10
−20
−30
−40
−50


Fundamental Frequency Over Time
Fundamental Frequency (f0)
Frequency (Hz)
Frames
75
70
65
60
55
50
100
150
200
250
300
350
400
450

Pitch Mean: 309.77197265625 Hz

RMS Energy Mean: 0.09022488445043564

Zero-Crossing Rate Mean: 0.021020039312366737

MFCCs Shape: (13, 469)

Mean Fundamental Frequency (f0): 54.18077990526883 Hz

---

4. Analysis

4.1 Visual Representations

Each audio file was analyzed and plotted using the following graphs:

- **Amplitude Over Time**: Displays variations in amplitude over time.
- **Spectrogram (Log Scale)**: Visualizes frequency components over time.
- **Zero-Crossing Rate (ZCR)**: Smoothed ZCR plot to observe signal changes.
- **MFCCs Plot**: Showcases extracted MFCC coefficients.
- **Fundamental Frequency (f0) Plot**: Illustrates variations in pitch.

4.2 Observations

- High-pitched speech (e.g., Excited Announcement) exhibits greater variations in pitch and higher ZCR.
- Whispered speech (e.g., Whispered Secret) has lower energy and pitch values.
- Formal Speech and Robotic Speech display stable pitch and energy patterns.
- Angry speech shows rapid fluctuations in amplitude and pitch due to emotional intensity.

---

5. Conclusion

This analysis highlights how different speech characteristics manifest in audio signals. Variations in pitch, volume, and speaking style influence amplitude, frequency content, and energy distribution. These features can be useful for applications such as speaker recognition, emotion detection, and speech synthesis.

---

6. References

1. **Librosa**
   M. McFee, B. McVicar, C. Raffel, et al., "librosa: Audio and Music Signal Processing in Python," Version 0.10.0, 2023. [Online]. Available: https://librosa.org/

2. **NumPy**
   C. R. Harris, K. J. Millman, S. J. van der Walt, et al., "Array programming with NumPy," *Nature*, vol. 585, pp. 357–362, 2020. [Online]. Available: https://numpy.org/

3. **Matplotlib**
   J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90-95, 2007. [Online]. Available: https://matplotlib.org/

4. **SciPy**
   P. Virtanen, R. Gommers, T. E. Oliphant, et al., "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020. [Online]. Available: https://scipy.org/

5. **os (Python Standard Library)**
   Python Software Foundation, "os — Miscellaneous operating system interfaces," Python Documentation, 2023. [Online]. Available: https://docs.python.org/3/library/os.html