Analysis of Emotional Tone in Historical Speech Recordings

1. Introduction

This report examines the emotional tone in historical speech recordings by analyzing key speech features:

- **Zero-Crossing Rate (ZCR)**
- **Short-Time Energy (STE)**
- **Mel-Frequency Cepstral Coefficients (MFCCs)**

Two distinct speech samples were selected:

- **Calm and Formal Speech** (e.g., Mahatma Gandhi's speech)
- **Passionate and Energetic Speech** (e.g., Motivational speech)

By comparing these features, we aim to understand how variations in speech style influence emotional tone.

---

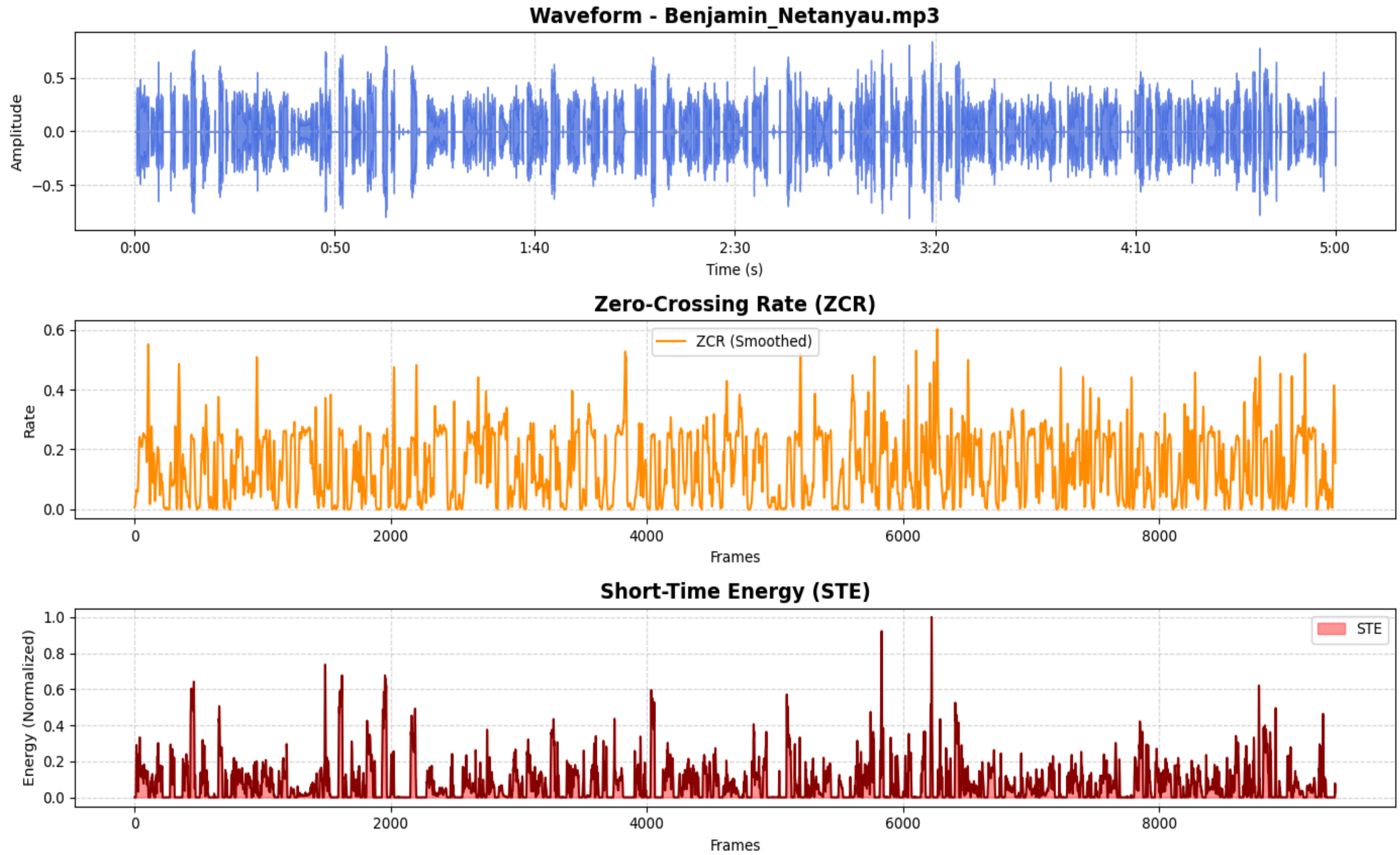
2. Feature Extraction Process

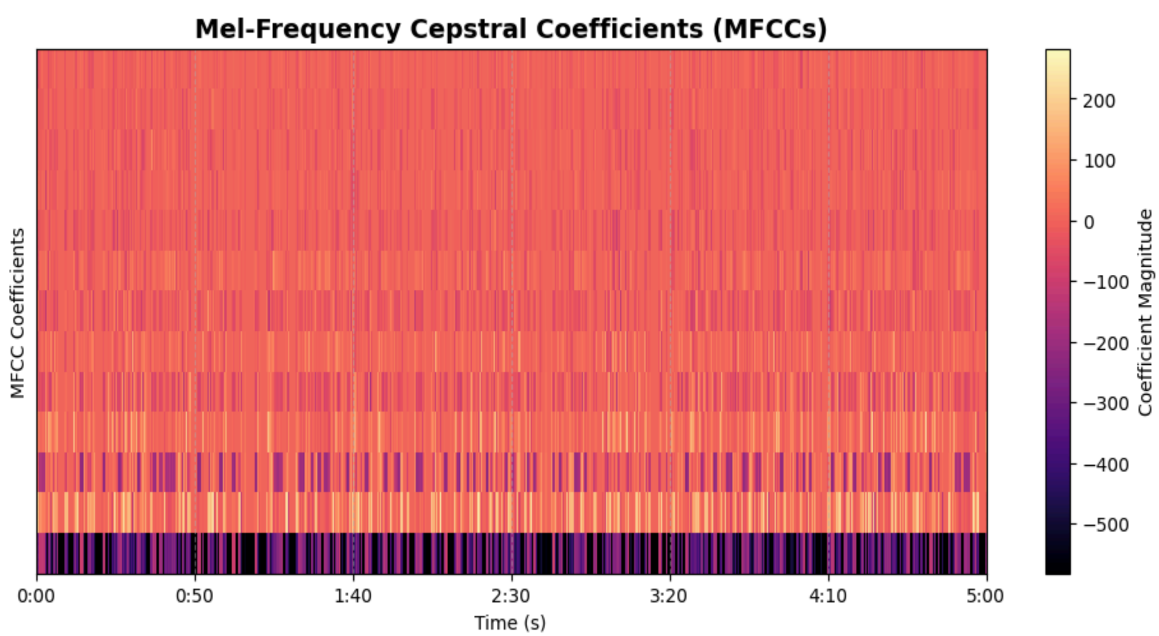
The following steps were performed:

- **Preprocessing:** The audio was loaded, resampled, and normalized for consistency.
- **Noise Reduction:** Applied to minimize background artifacts in historical recordings.
- **ZCR Calculation:** Measured the rate of sign changes in the signal, reflecting speech sharpness.
- **STE Computation:** Analyzed variations in signal energy, correlating with speech intensity.
- **MFCC Extraction:** Captured spectral properties, modeling vocal tract shape.
- **Graphical Analysis:** Plotted the extracted features for comparative visualization.

---

3. Graphs of Extracted Features(For the 'Benjamin_Netanyau.mp3' file)

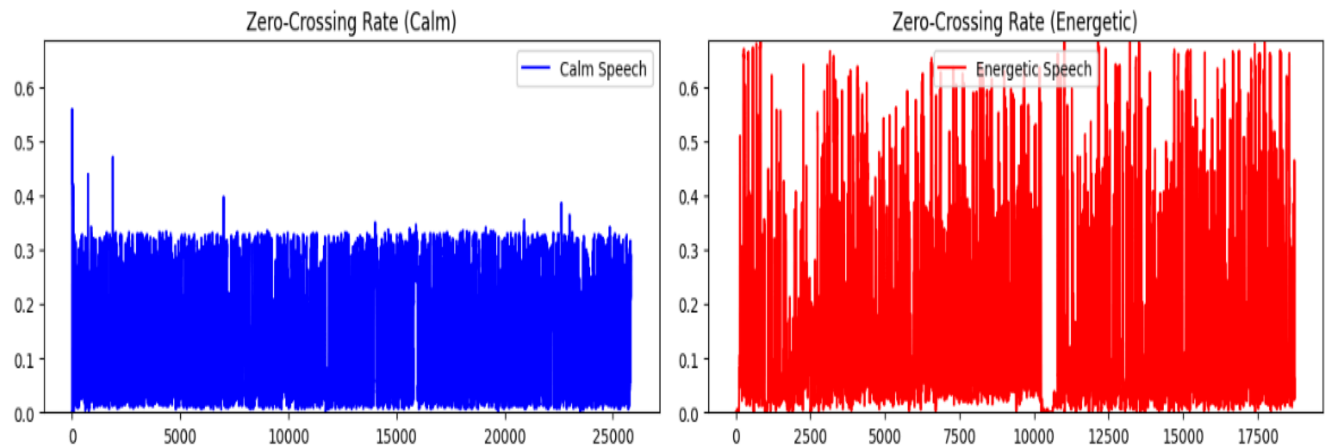




4. Feature Comparison & Analysis

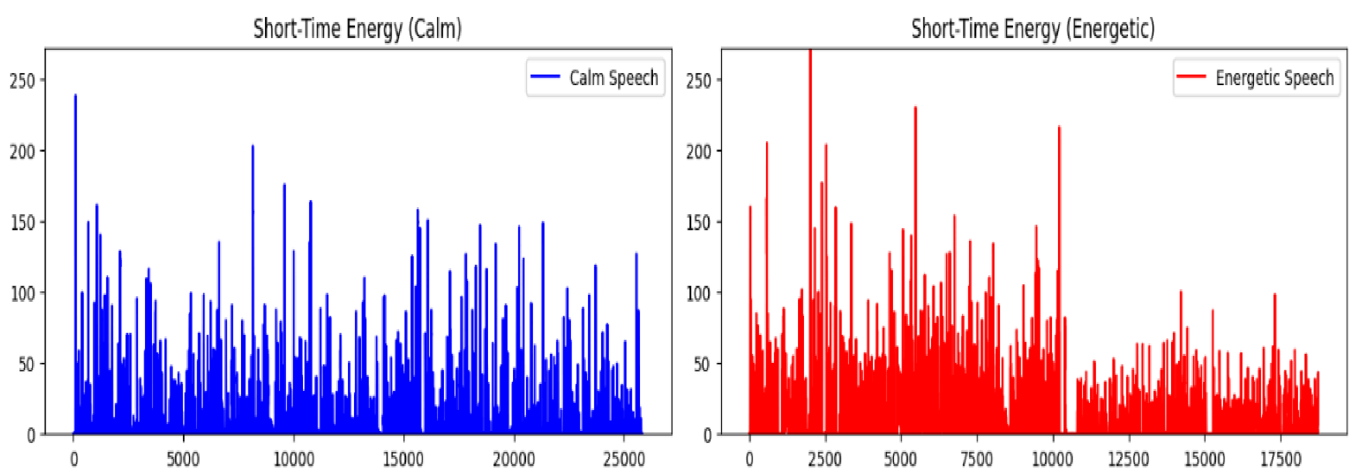
4.1 Zero-Crossing Rate (ZCR)

- **Calm Speech:** Lower ZCR due to smooth and steady pronunciation.
- **Energetic Speech:** Higher ZCR, indicating rapid frequency shifts and excitement.
- **Interpretation:** The increased ZCR in energetic speech suggests more abrupt changes in tone and articulation, creating an engaging effect.



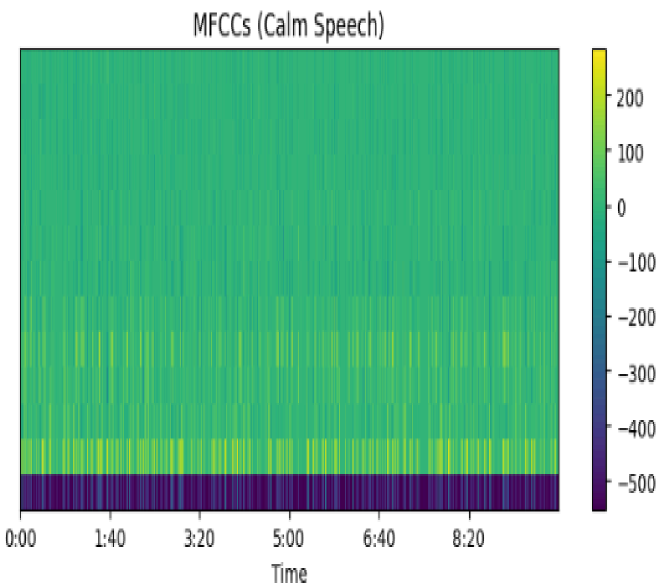
4.2 Short-Time Energy (STE)

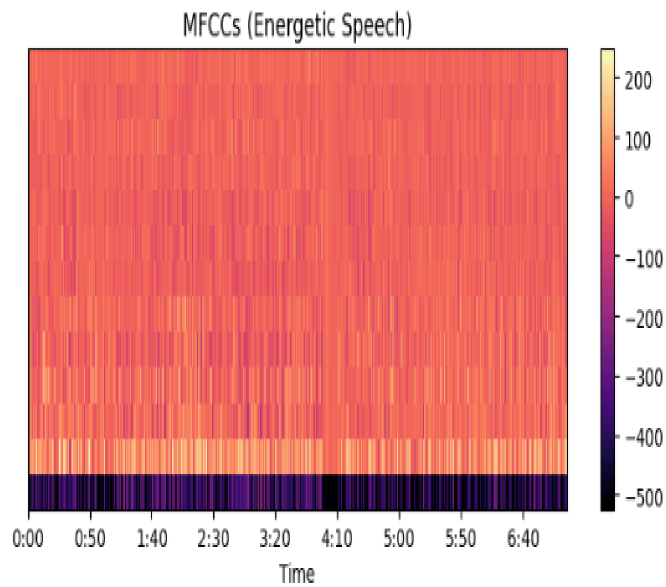
- **Calm Speech:** Lower energy, reflecting a controlled and steady tone.
- **Energetic Speech:** Higher energy spikes, suggesting strong emphasis and intensity.
- **Interpretation:** The higher energy in energetic speech signifies a forceful and emotionally charged delivery.



4.3 Mel-Frequency Cepstral Coefficients (MFCCs)

- **Calm Speech:** Less variation in MFCC patterns, representing steady articulation.
- **Energetic Speech:** More dynamic spectral shifts, indicating expressive and varied speech.
- **Interpretation:** The variations in MFCCs correlate with expressive vocal modulation in energetic speech.





---

5. Final Observations

Feature	Calm Speech	Energetic Speech
ZCR	Low	High
STE	Low Energy	High Energy
MFCCs	Smooth, less variation	Dynamic, fluctuating

6. Limitations of Traditional Speech Features

6.1 Sensitivity to Noise

- Historical recordings contain background noise that can distort ZCR and STE measurements.
- Poor microphone quality affects frequency content, making feature extraction less reliable.

6.2 Limited Emotional Representation

- Features like ZCR and STE primarily indicate intensity but fail to capture subtle emotional nuances such as stress and rhythm.
- Human emotional perception is influenced by prosody, which these features do not fully encapsulate.

6.3 Speaker and Recording Variability

- Different recording devices and environments introduce inconsistencies in feature extraction.
- Varying accents and speech speeds can alter MFCC-based classification results.

---

7. Suggested Improvements

7.1 Deep Learning-Based Emotion Recognition

- Convert speech signals into Mel-Spectrograms for enhanced feature representation.
- Utilize **Convolutional Neural Networks (CNNs)** or **LSTM-based models** for more accurate emotion detection.
- Pre-trained models (e.g., **Wav2Vec 2.0**, **DeepSpeech**) can improve recognition accuracy.

7.2 Advanced Noise Reduction Techniques

- Apply **Spectral Subtraction** or **Deep Learning-based Denoising (Wave-U-Net)** to clean historical recordings.
- Improves feature extraction by removing artifacts and preserving speech quality.

7.3 Use of Prosodic Features

- Integrate **pitch contour analysis**, **rhythm patterns**, and **stress detection**.
- These additional features align more closely with human perception of emotional tone.

---

8. Conclusion

Traditional speech features (**ZCR, STE, MFCCs**) provide fundamental insights into speech tone but face limitations related to noise sensitivity and emotional depth. Future improvements using **deep learning models**, **advanced noise reduction**, and **prosodic analysis** can significantly enhance emotion recognition in historical recordings.

---

9. 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. **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

5. **NoiseReduce**
   K. Braun, "Noise-reduce: Efficient noise reduction techniques in Python," Version 2.0, 2021. [Online]. Available: https://pypi.org/project/noisereduce/