Speaker Verification System for Noisy English Speech

GROUP MEMBERS

Harshil Bharagava
Viraj Busa
Esha Jain

MENTOR

Dr. Nitin Khanna

Department of Electrical Engineering , IIT Bhilai , Chhattisgarh, India

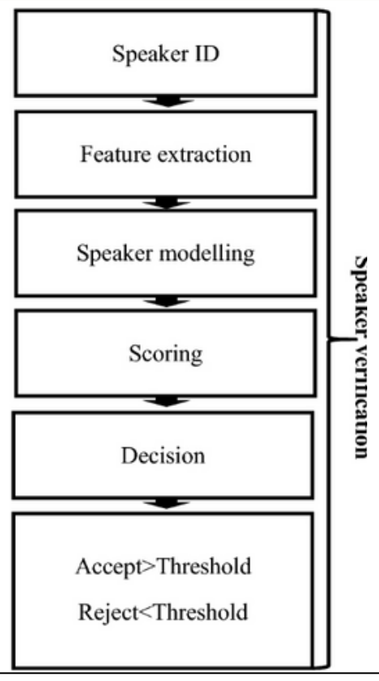
Group-9

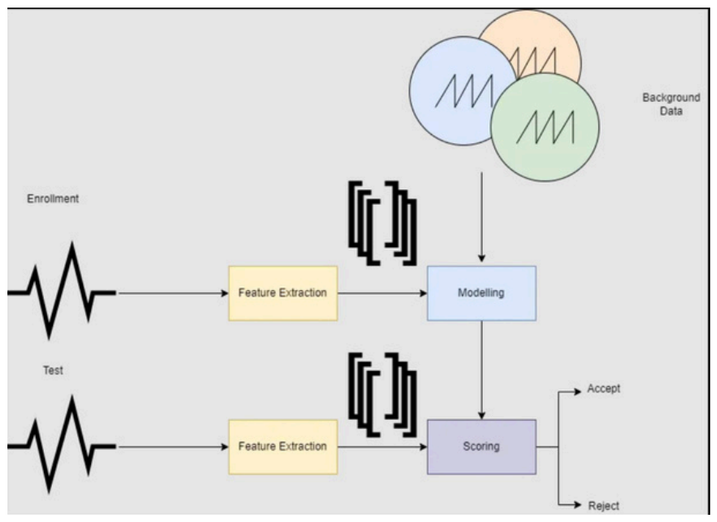
Introduction

Noisy environments challenge traditional speaker verification systems. From enhancing user authentication in mobile devices to bolstering security in financial transactions. By developing a robust system capable of effectively identifying speakers amidst noise, we can mitigate risks associated with impersonation and unauthorized access, ultimately fortifying the integrity of voice-based authentication systems.

Methodology

- We have used Resnet-34 pre-trained linear model For getting the embeddings of the audio file which is cleaned in the first step.
- After getting the embeddings we use it to Compare with the embeddings of the ground truth.
- The comparison is based on similarity score. Threshold set for that is 0.5.





Model Used

- **Pre-trained for Speaker Identification:** wespeaker-voxceleb-resnet34-LM is a pre-trained model designed to identify speakers from their voices. It leverages the VoxCeleb dataset, to learn speaker-specific characteristics.
- **ResNet-34 Architecture:** The core of the model is a convolutional neural network (CNN) called ResNet-34. It is trained on 21.8 Million Parameters
- **Learned Speaker Embeddings:** The model extracts speaker-specific features from audio data and projects them into a lower-dimensional embedding space. Each speaker is represented by a unique vector in this space, capturing their vocal characteristics.
- **Potential for Fine-Tuning:** The model can benefit from further fine-tuning on specific noisy data to achieve optimal performance in speaker verification systems.

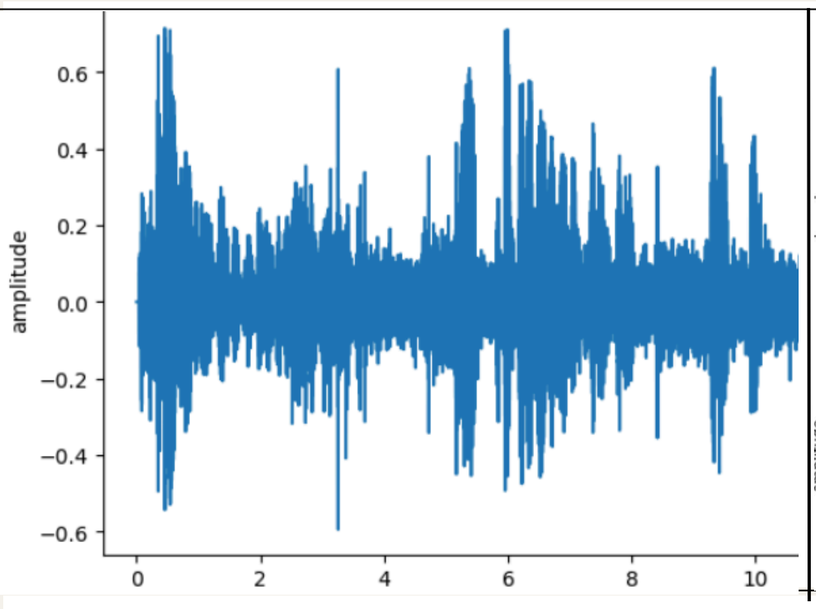
Audio Processing techniques tried

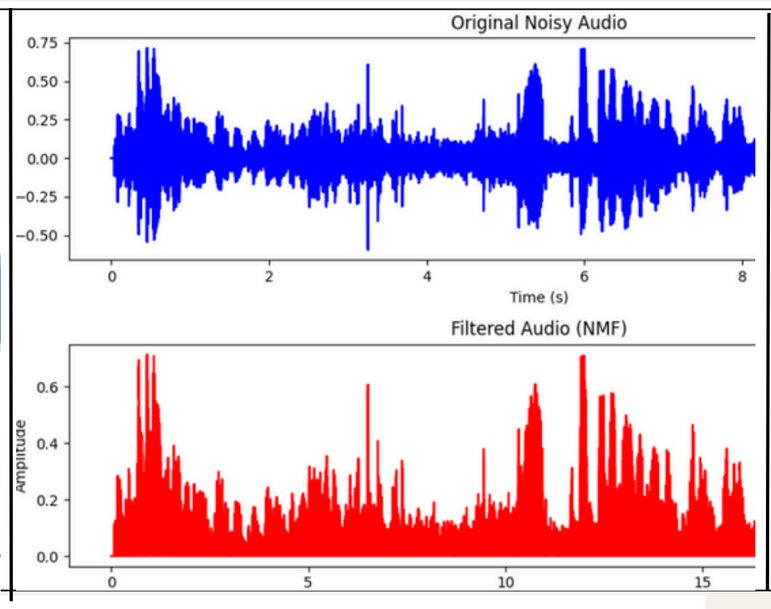
- **Spectral Subtraction:** It is a method used in audio signal processing for noise reduction, where the noise spectrum estimated from a noise-only portion of the signal is subtracted from the original spectrum.
- **NMF** is a technique that decomposes a given spectrogram into a set of basis vectors and their corresponding activations, often employed in noise reduction tasks to separate sources from a mixture.
- The **Wiener filter**, commonly utilized in audio processing, operates by minimizing the mean square error between the desired signal and the filtered signal, effectively reducing noise by exploiting the signal-to-noise ratio.

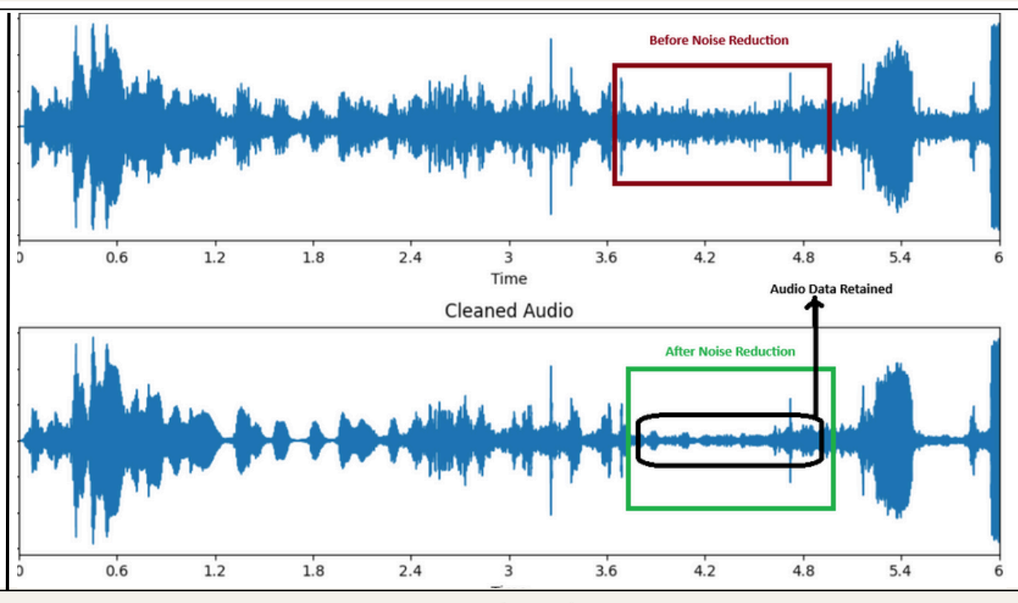
RESULTS

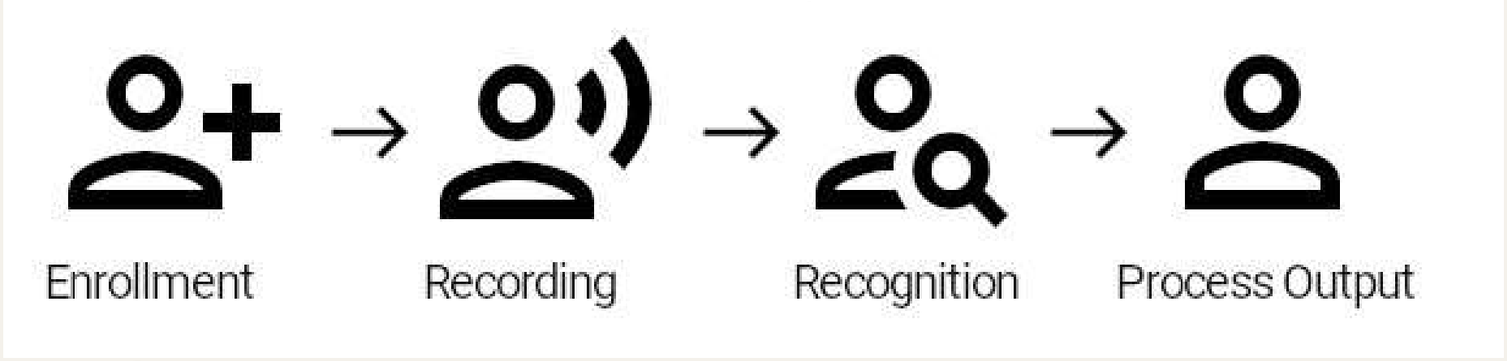
- Better results were shown when we opted for noise reduction instead of audio enhancement and other filtering techniques
- We observed that there was loss of original information on using too many enhancement and denoising features which resulted in poor results.
- After careful trials on real data and observations we came to a conclusion that if similarity score is above 0.5 means the speaker is correctly identified.

Analysis









Conclusion

This speaker verification system effectively identifies speakers in noisy English environments. By leveraging noise-resistant techniques, it ensures accurate verification even in challenging conditions. This paves the way for improved security, hands-free interaction, and user experience in real-world applications.

References

1. https://www.ibm.com/docs/en/SSMQSV_6.1.1/com.ibm.websphere.wvs.doc/wvs/wvs_sv.pdf (Denoising)
2. https://medium.com/@proof1234/mel-frequency-cepstral-coefficients-filter-banks-terminated-d02ae99cab3c (MfCC)
3. https://www.researchgate.net/figure/Wiener-Filter-implementation-using-Python_fig3_332574579 Denoising and Audio Enhancement (weiner Filter)