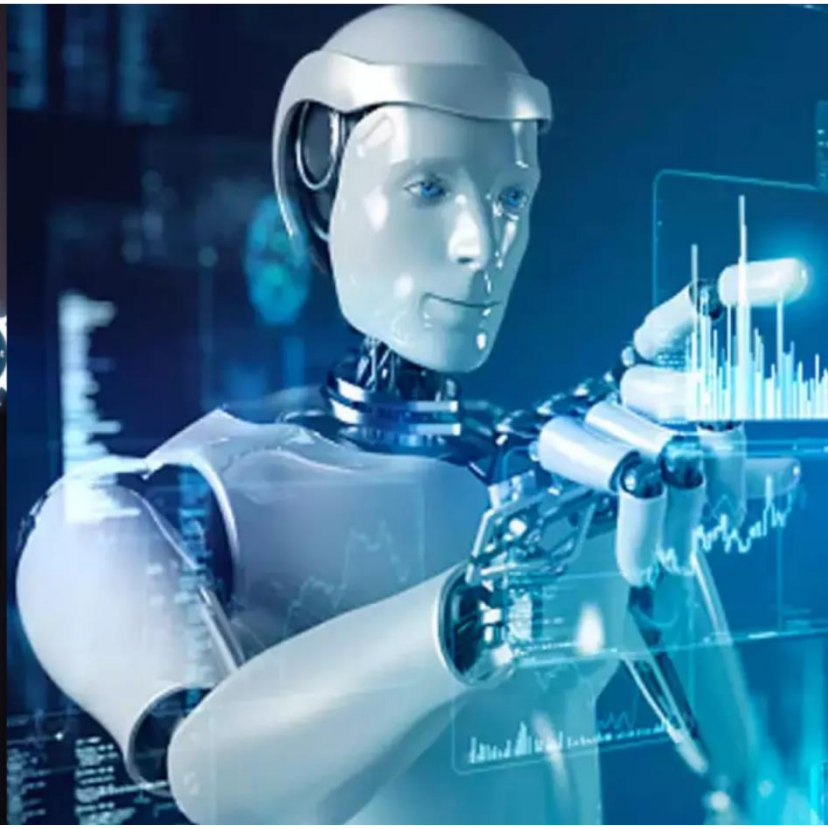
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404588|

Speech-To-Text Conversion using Machine Learning

Pranita Bhopale ,Sayali Chougale , Shweta Kamble Shraddha Mandave

C.O.Student, Department of Computer Engineering, K.P.Patil Institute of Technology, Mudal Maharashtra, India

Prof.V.V.Kumbhar

B.E.Information Technology, K.P.Patil Institute of Technology, Mudal Maharashtra, India

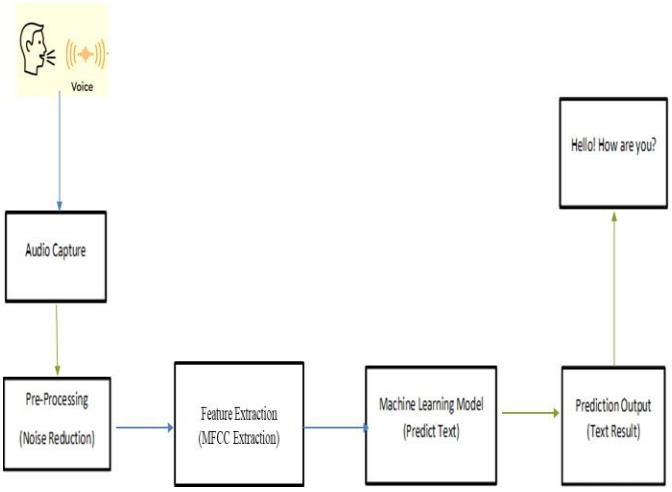
ABSTRACT: This project presents the development of a Speech-to-Text Converter that integrates Speech Recognition with Librosa and Matplotlib to analyze and visualize the audio signal. The system captures speech using a microphone, converts it to text using Google's Speech Recognition API, and simultaneously processes the audio to extract its Mel-Frequency Cepstral Coefficients (MFCCs), a key feature in speech analysis. The MFCCs are scaled and visualized as a heatmap using Matplotlib, providing insights into the spectral characteristics of the spoken words.

I. INTRODUCTION

Speech-to-text technology has become a critical component in modern applications, ranging from virtual assistants to transcription services. The ability to convert spoken language into written text is vital for creating more accessible and efficient human-computer interactions. Traditional speech recognition systems rely on various algorithms to process and transcribe audio signals, and advanced techniques have integrated feature extraction methods such as Mel-Frequency Cepstral Coefficients (MFCCs) to improve accuracy and performance.

The system utilizes Tkinter to provide an intuitive Graphical User Interface (GUI), where users can interact with the application to record speech, view the transcribed text, and visualize the MFCC features of the audio. By incorporating real- time speech-to-text conversion and audio feature visualization, this project demonstrates an effective tool for exploring the intersection of speech recognition, signal processing, and machine learning in practical applications.

II. BLOCKDIAGRAM



User Interaction:

The user opens the application and clicks on a button to start recording their speech.The system activates the microphone and begins recording the user's speech.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404588|

Speech Recognition:
The SpeechRecognition library captures the audio and sends it to Google's Speech API for transcription.The API returns the text, which is displayed in a text box on the GUI

Algorithm
1. Audio Input Acquisition**:**
Capture real time or pre-recorded audio
2. Preprocessing:
Noice reduction,normalization,resampling
3. Feature Extraction:
Extract MFCCS, Spectrogram,Chroma Features
4. Model Traninig and Prediction**:**
Train deep learning models(LSTM,CNN,Transformer)
Convert extracted features into text
5. Post Processing:
Aplly language model,error correction,punctuation
6. Output Display:
Show Transcribed text,store for further use

III. METHODOLOGY

The Speech-to-Text Converter project utilizes a multi-step process to convert speech into text, extract audio features, and visualize the audio characteristics.

The methodology can be broken down into the following steps:

1. Data Collection (Speech Input): Microphone Input:
The system uses a microphone as the input device to record the user's speech. This is captured using the SpeechRecognition library, which provides an interface to work with various speech recognition APIs (such as Google's API). The recorded speech is captured in real-time and saved as a temporary audio file in WAV format for further processing.

2. Speech Recognition: Speech to Text:
The Speech Recognition library is used to convert the recorded audio into text. The system listens to the audio, processes it through the chosen recognition engine (Google's Speech-to-Text API), and outputs the transcribed text.

Ambient Noise Adjustment:
To improve the accuracy of speech recognition, the system adjusts for background noise by using the recognizer.adjust_for_ambient_noise (source) function. This ensures that the recognition system adapts to varying noise levels in the environment.

3. Audio Feature Extraction with Librosa: MFCC Extraction:
In addition to transcribing the speech, the system extracts Mel-Frequency Cepstral Coefficients (MFCCs) from the speech signal. MFCCs are widely used features in speech processing

IV. FUTURE SCOPE

1] Multilingual Support:
- Adding support for multiple languages will make the system more accessible globally, allowing it to convert speech from different languages.

2] Machine Learning:
- Integrating machine learning modelscanimprovetran scription accuracy, especiallyin noisy environments or with different accents.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404588|

3] Noise Filtering:

- Advanced noise reduction techniques could make the system more accurate in real- world settings, where background noise is common.

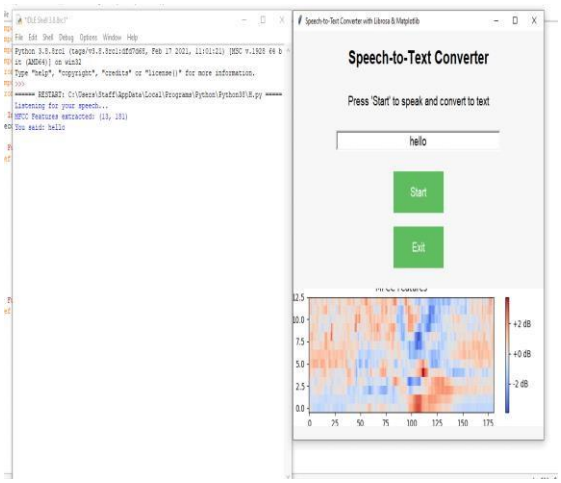
4] Real-Time Translation:

- The system could be expanded to not only convert speech to text but also translate it into other languages in real time.

5] Emotion Detection:

- By analyzing speech features, the system could detect emotions like happiness or sadness, useful for customer service or mental health applications.

V. RESULT



VI. CONCLUSION

Speech recognition and text summarization are two vast areas to be explored. The proposed research work aims to reduce the time and effort of manual documentation of lengthy speeches in an event. Speech recognition and text summarization can ease the work of documentation. Even for the verification of the summarized content, the system can be automated to read out the summarized content with the help of text to speech conversion. As of now, speech summarization for sentences terminating with a full stop or containing a small pause shown by comma is experimented. The future work is to include all punctuation marks in the recognized speech which helps in improving the text summarization performance

REFERENCES

- Python SpeechRecognition Library:

1. SpeechRecognition Documentation. (n.d.).Retrieved fromhttps://pypi.org/project/Speech Recognition/R. A. (2018). Speech recognition using Python. Medium. Retrieved from https://medium.com

- Librosa: Audio and Music Signal Processing:

2. McFee, B., & Raffel, C. (2015). librosa: Audio and music signal processing in Python. Proceedings of the 14th Python in Science Conference, 18-24. Retrieved fromhttps://librosa.org/Matplotlib: Visualization with Python:
3. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90–95. DOI:10.1109/MCSE.2007.55Matplotlib Documentation. (n.d.). Retrieved from https://matplotlib.org/

- Tkinter: Python GUI Library:

4. Grayson, D. (2000). Tkinter: A comprehensive guide to Tkinter. Retrieved from https://tkdocs.com/

- Speech Recognition with Deep Learning:

5. Hinton, G., et al. (2012). Deep neural networks for acoustic modeling in speech recognition. IEEE Transactions on Audio, Speech, and Language Processing, 12(1), 2-3. DOI: 10.1109/TASL.2012.2182145

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404588|

- Noise Reduction Techniques in Speech Processing:

6. Boll, S. F. (1979). Suppression of acoustic noise in speech using spectral subtraction. IEEE Transactions on Acoustics, Speech, and Signal Processing,27(2),113120.DOI:10.1109/TASSP.1979.1163209

- Real-Time Speech Recognition for Applications:

7. Zhou, X., & Peng, Q. (2019). Real-time speech recognition and its applications in intelligent systems. nternational Journal of Speech Technology, 22(4), 689-703. DOI: 10.1007/s10772-019- 09539-5

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details