NIRF-2024
Engineering Rank Band (151-200)
Pharmacy Rank - 77
Innovation Rank Band (11-50)

www.kiet.edu
Delhi-NCR, Ghaziabad

KIET GROUP OF INSTITUTIONS
Connecting Life with Learning

Accredited by
NAAC with Grade "A+"

Introduction to AI Project (AI1O1B)

Even Semester

Session 2024-25

Speech to Text Converter

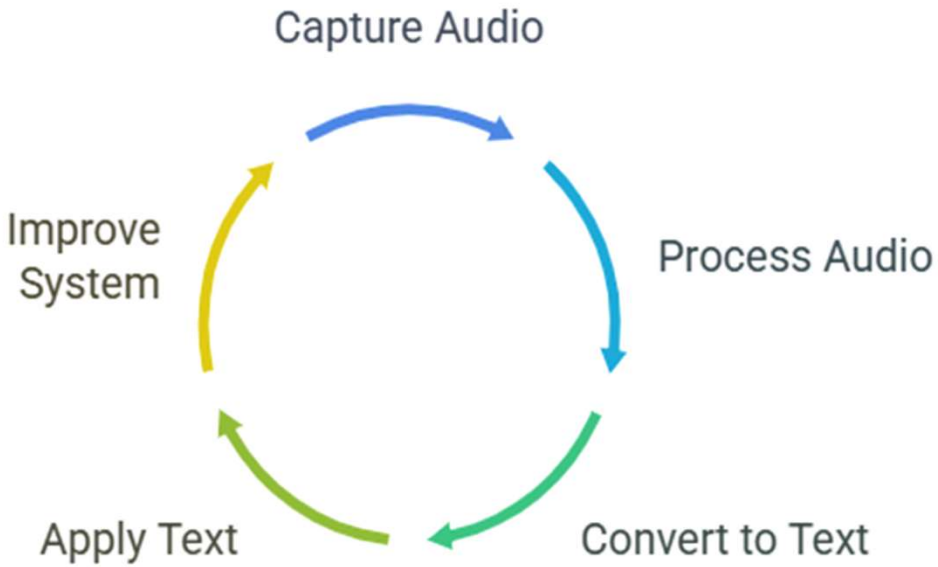
Abhishek Mishra 202410116100008
Ankit Kumar Shahi 202410116100028
Bhaskar Divedi 202410116100048
Avnish Kaushik 202410116100045

Project Supervisor:

Apoorv Jain
Assistant Professor

Introduction

This project aims to convert spoken audio into text in real-time using Python. It helps in applications like voice assistants, dictation tools, and accessibility support for hearing-impaired users.



Technologies Used

Python serves as the core scripting language, integrating PyAudio for capturing live audio and SpeechRecognition for transcription via Google API. NumPy analyzes audio signal levels to detect silence and segment meaningful speech portions.

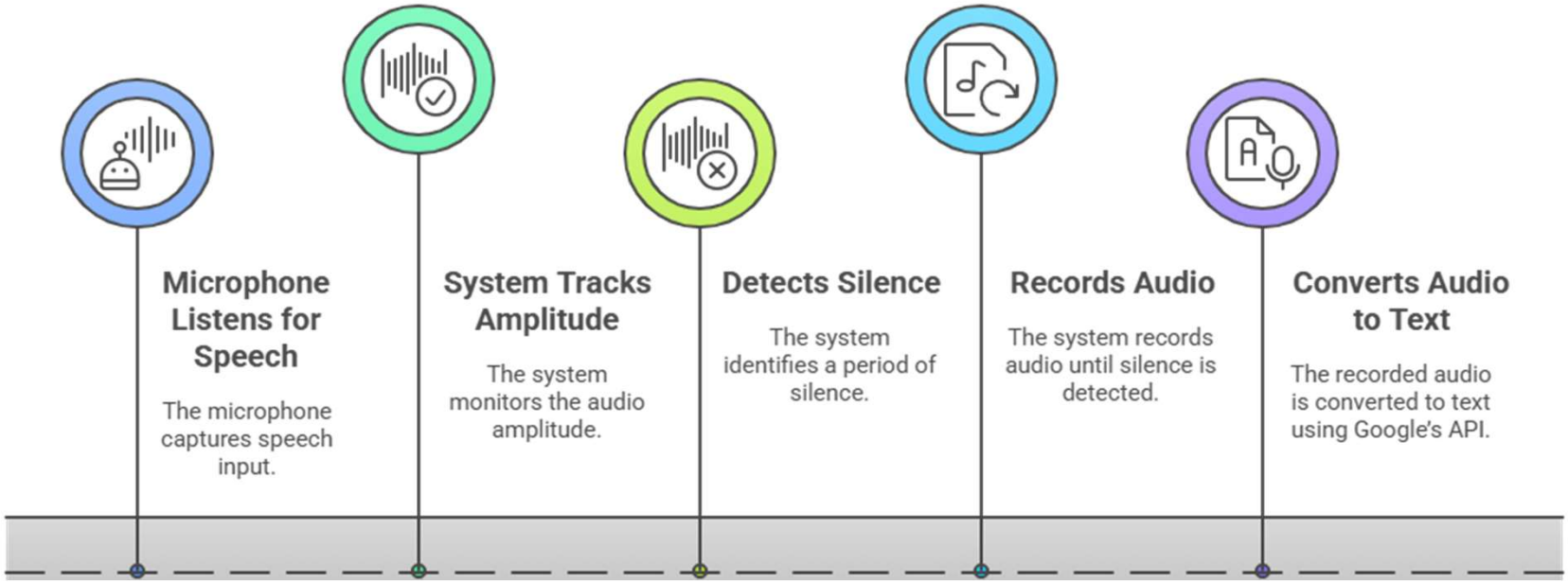
How It Works

•Microphone listens for speech input.

•System tracks amplitude to detect silence.

•Records audio in real-time until silence is detected.

•Converts audio to text using Google’s API.



Sample Output

- **Spoken Input:** *"hello what are you doing"*
- **Transcribed Output:** >>> hello what are you doing
- If unclear, returns: (Could not understand audio)

```
In [6]: # Start listening
        listen_and_transcribe()

        Listening... (Press Ctrl+C to stop)

        Speak now...
        >>> hello what are you doing

        Speak now...
        (Could not understand audio)

        Speak now...

        Stopped listening
```

Challenges Faced

Noise Filtering

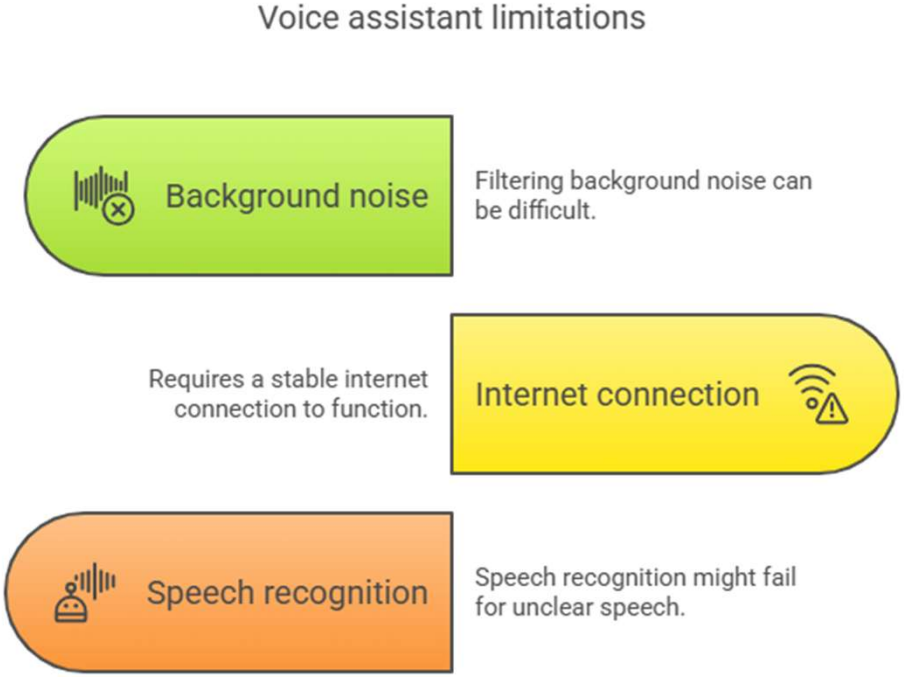
Improve algorithms to reduce background noise interference

Connectivity Dependence

Implement offline capabilities or enhance network resilience

Accent Recognition

Train models on diverse accents and speech patterns



Conclusion & Future Scope

The project efficiently converts real-time speech into text using Python and Google's API. Future improvements include adding offline transcription through Vosk or Whisper and integrating a GUI. It can also be enhanced to support multiple languages and regional dialects.