SCHOOL OF ENGINEERING

Dayananda Sagar University
School of Engineering
Devarakaggalahalli, Harohalli Kanakapura Road, Ramanagara, Karnataka 562112

Department of Computer Science & Engineering
(Artificial Intelligence & Machine Learning)

Natural language Models

SEMESTER – VI

Course Code: 22AM3610

Review 1: Mini Project

Title : MULTILINGUAL SPEECH RECOGNITION SYSTEM

Under the guidance of
Prof. Pradeep Kumar K
Prof. Sahil Pocker

Team Member
Name1: KUSHAL GOWDA H M (ENG22AM0029)
Name2: M NANDINI (ENG22AM0030)
Name3: RATHASTHA G D (ENG22AM0048)
Name4: SUPRITH HAVANGI (ENG22AM0064)

Date:

Contents

SDG Goals

1. SDG 4: Quality Education

- Facilitates **language learning** and education by providing real-time transcriptions and translations.
- Helps students with **hearing impairments** through automated captions and subtitles.

2. SDG 8: Decent Work and Economic Growth

- Enhances workplace productivity by enabling **multilingual communication** and automated transcription in business meetings.
- Supports **global workforce collaboration** through seamless speech-to-text conversion.

SDG goals

3. SDG 9: Industry, Innovation, and Infrastructure

- Drives **innovation in artificial intelligence (AI) and natural language processing (NLP).**
- Supports the development of **smart technologies** for industries such as customer service, healthcare, and media.

4. SDG 10: Reduced Inequalities

- Promotes **digital inclusivity** by making information accessible to people of all languages and abilities.
- Supports **non-native speakers and individuals with disabilities** by providing real-time speech-to-text and translation services.

5. SDG 16: Peace, Justice, and Strong Institutions

- Enhances **access to information** by breaking down language barriers in legal, governmental, and humanitarian sectors.
- Facilitates **multilingual communication in global governance** and diplomacy.

Introduction

- Develops a system for speech transcription and translation into multiple languages, including South Indian languages.
- Uses CNNs and Transformers for automatic image captioning.
- Integrates Google Speech API for speech recognition and Google Translate API for translation.
- Features Text-to-Speech (TTS) synthesis and sentiment analysis.
- Ensures data privacy with Fernet encryption.
- Aims to enhance accessibility, productivity, and communication across sectors.

Literature Survey

Author(s)	Title	Year	Methodology	Key Findings	Limitations
Shahana Bano, Pavuluri Jithendra, Gorsa Lakshmi Niharika, Yalavarthi Sikhi	Speech to Text Translation enabling Multilingualism	2020	Developed a multilingual speech-to-text model using **Google Speech Recognition API PyAudio, and Tkinter** in Python. Users select a language, input speech via a microphone, and receive translated text output.	The model facilitates **language translation and communication**, aiding **telecommunication , multimedia, and accessibility for illiterate users**.	**Depends on internet connection, speech clarity, and language support**; may struggle with **accents and complex linguistic structures**.
Tanuja Konda Reddy, Veeksha S, Kavitha C.R.	Enhanced Multilingual Image Captioning: Integrating Text-to-Speech Translation and Sentiment Analysis	2024	Uses **CNNs, RNNs, Transformers** for image captioning, **mBART & 3-layer Transformer** for translation, **Tacotron 2** for Text-to-Speech, and **Sentiment Analysis** for emotion detection.	Generates **multilingual captions & spoken translations**, achieving **85% sentiment analysis accuracy** and improving accessibility	Needs **larger datasets**, struggles with **context retention**, and is **computationally intensive**
A Yogi Athish, Srinivasa K G, Sivakumar M	Multilingual Speech Recognition Using Reinforcement Learning	2023	Uses **neural networks, acoustic modeling, and reinforcement learning** to improve multilingual speech recognition. Processes audio with **noise reduction, segmentation, and text normalization** and evaluates accuracy via **WER & CER**.	Improves **transcription accuracy**, handles **accents and noise**, and benefits from **large datasets (650,000+ hours)**.	**Struggles in noisy environments, requires high computational power, and lacks contextual understanding**.

Literature Survey

Author(s)	Title	Year	Methodology	Key Findings	Limitations
Thomas Rolland, Alberto Abad, Catia Cucchiarini , Helmer Strik	*Multilingual Transfer Learning for Children Automatic Speech Recognition*	2022	Combines **multilingual learning (MTL)** and **transfer learning (TL)** for children's ASR, using **five speech corpora** (Swedish, German, English, Portuguese, Dutch). Models were trained with **Hybrid HMM-DNN** in **Kaldi**, incorporating **i-vectors** and **data augmentation**.	**MLTL improves performance** over single-language, MTL, or TL alone (**7.73% WER reduction**). Beneficial for **unseen languages** (**5.56% improvement**). Larger datasets enhance generalization.	**Data scarcity**, **age-related performance variation**, **focus on European languages**, and **task variability** affecting results.
Pachipala Yellamma, Yogendra Chowdary, Potla Raghu Varun, Polisetty Manikanth, Nunna Charan Naga Lakhshmi Narayana, Kunderu Hemanth Ganesh Sai	Automatic and Multilingual Speech Recognition and Translation by using Google Cloud API	2024	Web app using **Google Speech Recognition API** for ASR, **Google Translate API** for multilingual translation (Hindi, Tamil, Telugu), **Flask-based UI**, **MySQL database**, and **Fernet encryption** for security.	High ASR accuracy, real-time translation, cloud scalability, robust for accents/dialects, secure data storage.	Supports only **three languages**, affected by **accent/noise**, potential **latency issues**, lacks **mobile integration**.

Problem Definition

- Language barriers remain a challenge in multilingual regions like India, where most speech recognition systems are limited to English, leaving non-English speakers at a disadvantage.
- This project aims to create a multilingual speech recognition system that accurately transcribes and translates speech into multiple languages, ensuring accessibility, accuracy, and data privacy.

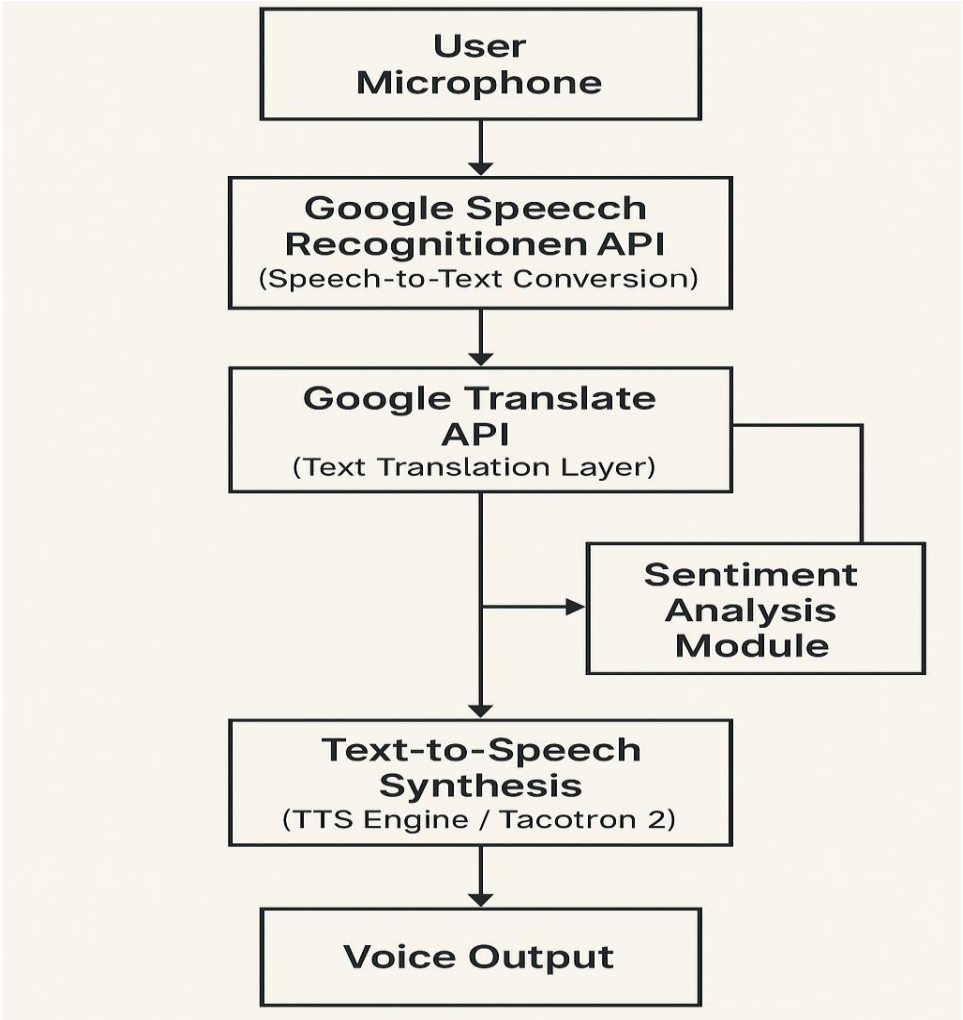
SOLUTION :

Design and implement a multilingual speech recognition system to overcome language barriers by providing precise transcription and translation across diverse languages. This solution ensures inclusive accessibility, exceptional accuracy, and strong data protection for users in multilingual regions like India.

Objectives

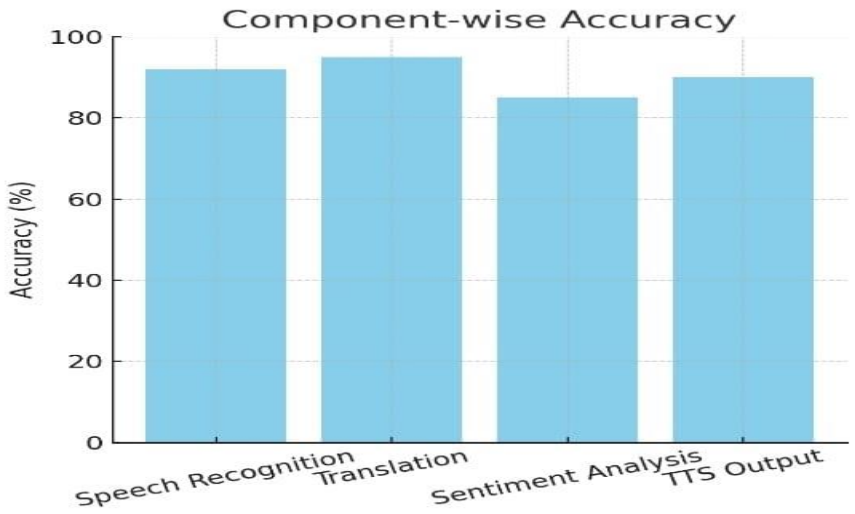
- Develop a multilingual speech recognition system for precise transcription and translation across languages.
- **Leverage advanced models** (CNNs, Transformers, RNNs) to enhance recognition accuracy.
- **Integrate Text-to-Speech (TTS)** and sentiment analysis for enriched user interaction.
- Ensure robust data security using Fernet encryption.

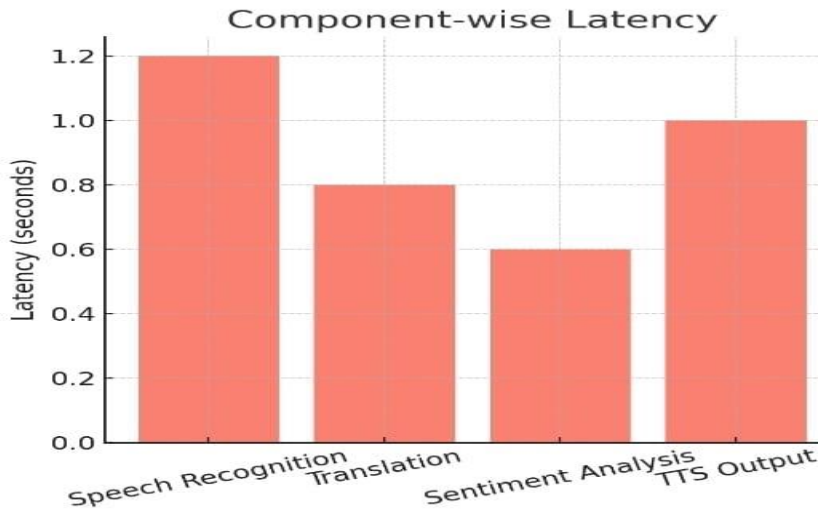
Methodology



RESULT AND ANALYSIS :

- Accuracy: Translation and Speech Recognition components perform the best (92–95%), while sentiment analysis is slightly lower due to contextual sensitivity.
- Latency: All components respond within 1.2 seconds, indicating efficient real-time performance





Conclusion

- Advancements in speech recognition are driven by machine learning, data, and computing power. Multilingual transfer learning improves accuracy, especially for children's speech in low-resource scenarios.
- The project offers an innovative multilingual speech recognition solution with strong data privacy and accessibility.
- Future work includes improving real-time translation, expanding language support and testing larger datasets.

References

1. Karimunnisa, S., Pachipala , Y. An AHP based Task Scheduling and Optimal Resource Allocation in Cloud Computing (2023) International Journal of Advanced Computer Science and Applications, 14 (3), pp. 149-159.

2. Batliner, A., Blomberg, M., D'Arcy, S., Elenius, D., Giuliani, D., Gerosa, M., Hacker, C., Russell, M., Steidl, S., and Wong, M. (2005). The pf star children's speech corpus. pages 2761–2764, 01.

3.Kim, W., Stern, R.M. and Ko, H., 2005. Environment independent mask estimation for missing-feature reconstruction. In Ninth European Conference on Speech Communication and Technology.

4. Speech-to-Speech Translation: A Review, Mahak Dureja Department of CSE The North Cap University, Gurgaon Sumanlata Gautam Department of CSE The North Cap University, Gurgaon. International Journal of Computer Applications (0975 – 8887) Volume 129 No.13, November2015.

5. Banu, Sameena, and Syeda Ummayhani. "Text Summarisation and Translation Across Multiple Languages." Journal of Scientific Research and Technology (2023): 242-247.

Acknowledgments

We express our heartfelt gratitude to our mentors and guides for their invaluable insights and support, which have been instrumental in shaping this multilingual speech recognition project.

We sincerely appreciate our collaborators for their dedication and teamwork in refining our approach. Additionally, we are grateful to our peers and all individuals who contributed their time, expertise, and feedback, helping to enhance the quality and impact of our work.