ISSN: 2395-7852

# International Journal of Advanced Research in Arts, Science, Engineering & Management

Volume 11, Issue 6, November - December 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.583

+91 9940572462 +91 9940572462 ijarasem@gmail.com www.ijarasem.com

International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 7.583 | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 11, Issue 6, November-December 2024 |

# Software for Dubbing English Videos into Indian Languages

**V.V.S.S. Balaram[1], V. Deepak[2], Bharathchandra[3], Tahura Afsheen[4]**

Assistant Professor, Department of Computer Science and Engineering, Anurag University, Hyderabad, Telangana, India[1]

UG Student (Department of Computer Science and Engineering, Anurag University), Hyderabad, Telangana, India[3,4,5]

**ABSTRACT:** In India, where there are as many languages spoken as there are leaves on a tree, the demand for video content in Indian languages is growing at a phenomenal pace. Dubbing into all Indian languages like Hindi, Tamil, Telugu, and Bengali is critically important to make content more accessible to the masses. It is in this context that the paper presents an automation solution of dubbing through technology, that will integrate Automatic Speech Recognition, Machine Translation, and Text-to-Speech. The entire process begins when the ASR converts the audio of the English input into text, which is then translated by using advanced models of NLPs into the target Indian language. Further, the resultant text is synthesized as natural speech through TTS engines such as Google Text-to-Speech (gTTS). Finally, getting the dubbed audio assimilated with the original video with proper lip-sync and timing is the last step. This auto-method introduces dubbing as an automated process that saves time and cost while still maintaining the translation intact on the audio and voice quality respectively, along with synchronization, thus providing a cost-effective solution for content creators, businesses, and even educators who face the problem of communication in multiple languages across India.

**KEYWORDS:** Natural Language Processing, Gemini API, GTTS, Automatic Speech Recognition.

## I. INTRODUCTION

In today's digital age, video content is a powerful medium for communication, education, entertainment, and marketing. However, language barriers can limit the reach and impact of this content, especially in linguistically diverse countries like India. With over 1.4 billion people and more than 20 official languages, there is a growing need to make English-language videos accessible to regional audiences.

And that is where the software for dubbing English videos in Indian languages comes in. This software keeps the original message but changes it to regional languages and, therefore, lets the creators of content, teachers, and organizations reach more people. It functions through complex technologies like artificial intelligence, machine learning, and natural language processing to automatically or manually translate one language into another.

Thus, integrating the Gemini API with gTTS gives an end-to-end solution that is not only able to translate text but provides proper professional-grade voiceover so that video content is accessible and engaging for all the viewers across India's many different linguistic regions.

## II. RESEARCH METHODOLOGY

The research into this project focuses on developing and optimizing technologies that automate the process of language translation**,** voice synthesis**,** and audio synchronization. Referred to the afore mentioned research papers for the process and flow.

## III. PROBLEM IDENTIFICATION

The first and foremost and most important step of the methodology is problem identification. This step focuses on understanding the challenges faced by people in any part of India who do not understand English even though there is vast amount of content available. Individuals in rural areas prefer using their native language such as Hindi, Telugu, Tamil etc, as their means of communication and for receiving knowledge through videos whether for education or for entertainment purposes.

| ISSN: 2395-7852 | www.ijarasem.com | Impact Factor: 7.583 | Bimonthly, Peer Reviewed & Referred Journal|

| Volume 11, Issue 6, November-December 2024 |

## IV. SYSTEM ARCHITECTURE

**Client-side (front-end):** The user interface of the application was developed by using Streamlit Web based framework to focus on functionality of web design.
**Server-side (Backend):** Backend consists of Python web framework to handle API requests from Streamlit client and perform backend processing.

## V. INPUT AND PROCESSING

Users can provide an English video as an input to the application. The system uses its APIs to process the file that is uploaded. When the user selects the target language and clicks on "process video" button the process starts by extracting the audio, transcribing it into text, translating and converting into audio where the original English audio is replaced with translated audio. Supported languages include Hindi, Telugu, Tamil, Kannada, Urdu, Malayalam, Punjabi, Marathi, Gujarati and Nepali.

## VI. RESULTS AND DISCUSSION

During its run, the process of subtitling English videos into Indian languages will start typically involving several steps including speech recognition, translation, text-to-speech synthesis, and then synchronization. Automatic Speech Recognition process commences by inputting the English video into the software. The software uses Automatic Speech Recognition (ASR) technology to convert the audio in the video into text. In this process Google-Speech-to-Text tools are utilized for transcribing the speech. The English transcript is broken into sentences or phrases; that will facilitate the translation and aligning to the video. The processed text of English is then translated into the target Indian language-for example, Hindi, Tamil, Bengali-using the MT engine. In the process of translation, Gemini API tool has been used along with gTTs tools. The translated audio then replaced by the original English audio provides the user with the expected translation of the video.

**Libraries used**
MoviePy is used for audio extraction, synchronisation of dubbed audio with the video while Gemini API is a machine translation tool that converts English text into Indian languages, and handling context.
All the tools used in this project form a cohesive pipeline for automated, high-quality dubbing of English videos into Indian languages.

**Table 1:** *Summary of libraries used.*

| Library | General use | Contextual use |
|---|---|---|
| Streamlit | Open-source Python framework that allows developers to create interactive web applications quickly. It's an ideal tool for building a front-end interface | Streamlit allows users to upload English videos via a simple drag-and-drop interface. This provides a user-friendly way for content creators or businesses to interact with the dubbing system. |
| Moviepy | Python library used for video editing, which provides essential functionalities for processing video and audio in the dubbing pipeline. | Extracts and synchronises the audio to the input video. |
| gTTS | That converts text into speech, enabling the translated text (from Gemini API) to be spoken aloud in the chosen Indian language. | After translation, gTTS converts the Indian language text into speech gTTS supports various Indian languages, such as Hindi, Tamil, Telugu, and Bengali, making it a flexible tool for multi-language dubbing software. |
| Gemini | Machine translation of the extracted English transcript into the target Indian languages. | To automate translation for low-resource languages, which is crucial when working with Indian languages. The transcript is passed through the Gemini API to translate it into the desired language. |

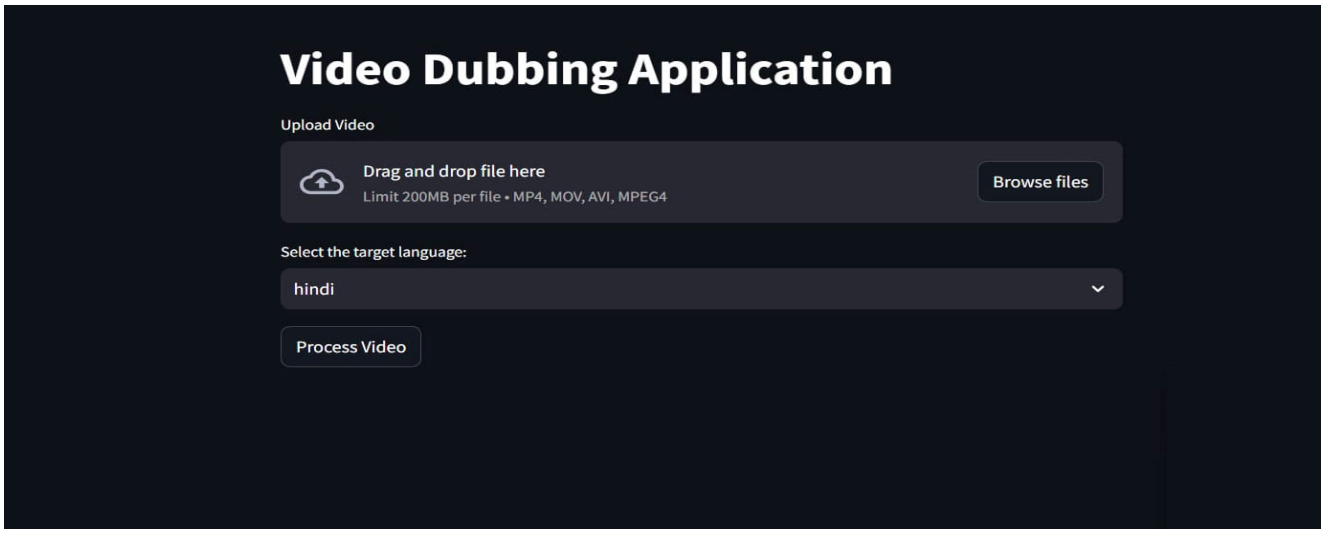



**Figure 1**: User interface

## VII. CONCLUSIONS

In conclusion, the development of the video translation software marked a significant stride towards bridging linguistic gaps and fostering cultural inclusivity. The project successfully showcased the feasibility of translating videos from English to various Indian religious languages, contributing to enhanced accessibility and understanding across diverse communities.

## VIII. DECLARATIONS

**Study Limitations**

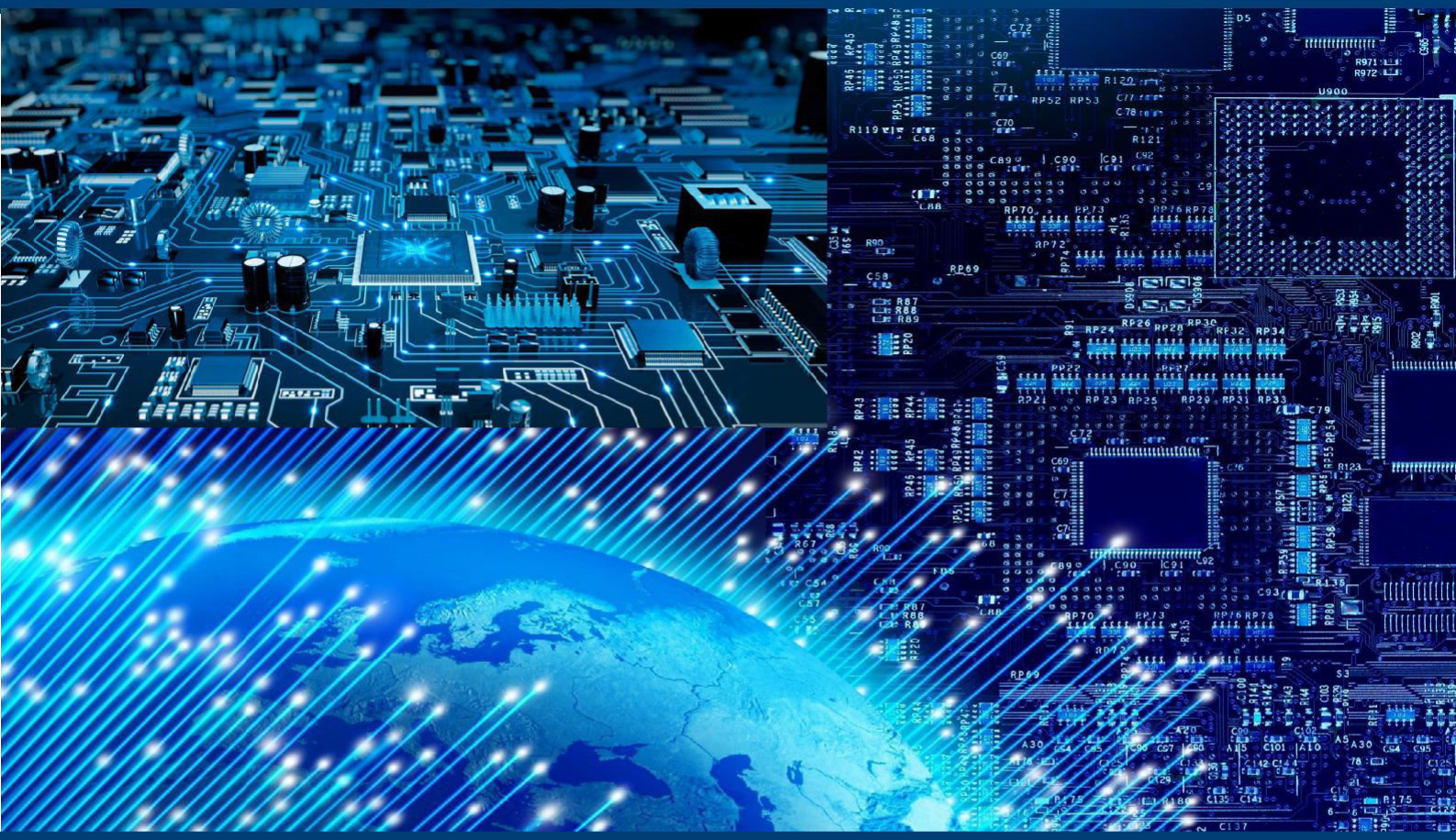
The current project limits to only one speaker per video, though the input video might contain multiple speakers i.e. a video having lengthy conversations, the audio that is synchronised for the entire video is of only one voice.
Indian languages are complex and diverse, and preserving meaning, tone, and cultural nuances is critical. Some words in English do not have a direct equivalent in Indian languages, so it requires some context-aware translation. People like to talk in their native accent or slang in rural areas with some speed; while the current project is not equipped with speed and accent-changing, it could be improved in the future with evolving technology.

**Acknowledgements**

We would like thank our mentors and colleagues for their invaluable insights and support throughout the development of this project. Sincere thanks to Dr. V.V.S.S.S. Balaram sir for his guidance in completing the project.

## REFERENCES

[1]. Saad A. Bazaz et. al. (2022) “Automated Dubbing and Facial Synchronization using Deep Learning”. https://ieeexplore.ieee.org/abstract/document/9773697
[2]. Chenxu Hu et. al. (2021) “Neural Dubber: Dubbing for Videos According to Scripts” in 35th Conference on Neural Information Processing Systems (NeurIPS 2021). https://arxiv.org/abs/2110.08243
[3]. Mattia Di Gangi et. al. (2022) “Automatic Video Dubbing at AppTek” in Proceedings of the 23rd Annual Conference of the European Association for Machine Translation. https://aclanthology.org/2022.eamt-1.65/
[4]. Marcello Federico et. al. (2020) “From Speech-to-Speech Translation to Automatic Dubbing” in Proceedings of the 17th International Conference on Spoken Language Translation https://aclanthology.org/2020.iwslt-1.31/
[5]. Andrew Gibiansky (2017) "Deep Speech: Scaling up end-to-end speech recognition"



IJARASEM

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

# International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| Mobile No: +91-9940572462 | Whatsapp: +91-9940572462 | ijarasem@gmail.com |

www.ijarasem.com