Bridging Linguistic Barriers in E-Learning: Automated Translation for English Videos Into Indian Languages

[1] Prof. Rajesh K Nale, [2]Bagal Sneha Suresh, [3]Bhoite Harshraj Mahesh, [4]Ghadge Shweta Rajaram, [5]Mohite Shravani Shelar
SVPM COE MALEGAON(BK), BARAMATI.

Abstract: **This paper presents a web-based system designed for machine translation of English educational videos into Indian regional languages. By integrating modules for speech recognition, translation, and text-to-speech synthesis, the platform allows video content to be localized into various Indian languages. A Python-based backend with a web interface allows educators to upload content and receive translated and dubbed videos. This system promotes inclusive education across India's multilingual population, especially in rural and semi-urban areas.**

***Index Terms*— Machine Translation, Educational Videos, Indian Languages, Speech Recognition, Audio Dubbing, Python, Flask, Web Automation**

I. INTRODUCTION

India is a linguistically diverse nation, home to 22 officially recognized languages and over 1,600 dialects spoken across its vast geographical landscape. While English has emerged as the dominant medium of instruction in higher education and online learning platforms, a significant portion of the population still faces difficulties in comprehending educational content in English. This language barrier contributes to a lack of inclusivity and restricts access to quality learning resources for non-English speakers.

With the rapid digitization of education—especially after the COVID-19 pandemic—video-based learning has become increasingly prevalent. However, the dominance of English in these videos often alienates learners from rural or regional backgrounds. To address this issue, machine translation offers a scalable and efficient solution, enabling the automatic translation and dubbing of educational content into regional languages. This research focuses on designing and implementing a system that leverages machine translation frameworks and speech processing tools to convert English educational videos into Indian regional languages. The proposed approach utilizes open-source libraries, web technologies, and AI-driven modules to bridge the linguistic divide in India's educational ecosystem. By doing so, it aims to democratize knowledge and promote equitable access to learning for all linguistic communities across the country. To bridge this divide, the implementation of machine translation for educational

II.RELATED WORKDONE

Several research studies have been conducted in the field of machine translation, speech synthesis, and educational technology. The evolution of neural machine translation (NMT) and speech-to-text systems has enabled more accurate and natural translations across languages.

Sutskever et al. (2014) introduced the sequence-to-sequence learning model, which became a foundational approach in machine translation. Later, the introduction of attention mechanisms by Vaswani et al. (2017) in the Transformer architecture further improved translation quality by focusing on relevant parts of the input.

In the context of Indian languages, various open-source projects and APIs like Google Translate and Microsoft Azure Translator have supported limited regional language pairs. However, most of these systems are either limited in scope or focused on text-only translation.

Existing platforms like YouTube auto-captioning and Khan Academy use subtitles to provide multilingual support, but they do not offer automated dubbing or voice translation. Projects like Mozilla's DeepSpeech have explored speech-to-text solutions, but their regional language accuracy remains under development.

Despite these advancements, there remains a significant gap in systems that can handle end-to-end video translation and dubbing for regional education. Our proposed system addresses this gap by offering a seamless pipeline from video upload to dubbed output in a target Indian language.

III. METHODOLOGY

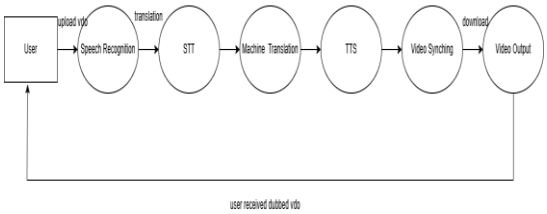


Fig. data flow diagram

The proposed system for machine translation of English educational videos into Indian regional languages follows a modular pipeline, ensuring efficient speech extraction, translation, and synchronized video rendering. The architecture is structured as follows:

A. Speech Recognition

The system begins by extracting audio from the input video using automated tools. This audio is then subjected to speech recognition techniques using models such as Google Speech-to-Text or Mozilla DeepSpeech. The objective is to convert the spoken English content into raw textual data for further processing.

B. Speech-to-Text (STT) Conversion

The extracted audio is converted into precise textual format. Advanced STT systems process natural speech patterns, accounting for intonations, pauses, and speaker-dependent variations to generate a clean English transcript. This textual data becomes the input for the translation module.

C. Machine Translation

This is the central component of the system. The English transcript is passed through a neural machine translation (NMT) model that translates it into the selected Indian regional language. Frameworks like Mariann MT or OpenNMT can be utilized to ensure contextual, semantic, and grammatical accuracy.

D. Text-to-Speech (TTS) Synthesis

The translated text is then converted back into speech using TTS engines such as Google TTS, eSpeak, or Festival. The goal is to produce natural-sounding audio in the target language while maintaining appropriate pronunciation and prosody.

E. Video Synchronization

The generated regional language audio is synchronized with the original video stream. Techniques are employed to align speech timing and video scenes for maintaining contextual continuity. This step ensures minimal lip-sync mismatch and a seamless viewing experience.

F. Output Generation

Finally, the processed video — now with regional language audio — is rendered and made available to the user. The end product enhances accessibility of educational content across diverse linguistic backgrounds in India.

Other work done:

- Flask Setup: The Flask app handles video upload, session management (login/logout), and routes for processing and downloading the video. The user interface allows users to upload videos, select languages, and get the processed output.
- Gender Detection: The detect_gender() function uses audio segmentation to identify the gender of the speaker in the video (male or female). This is important for selecting the correct voice for text-to-speech conversion.

IV. PROPOSED SOLUTION

The proposed solution aims to automate the translation of educational videos in English into Indian regional languages (like Hindi, Marathi, Tamil, etc.), using an integrated pipeline of technologies. The goal is to enhance accessibility to educational content for a wider audience in India.

System Overview

This system leverages multiple Python-based tools and libraries for speech recognition, language translation, text-to-speech conversion, and video processing. The user uploads an English educational video, and the system automatically generates a translated and dubbed version in the selected regional language. The end result is a fully synchronized video with translated audio, ready for download.

Workflow

1. Video Upload and Audio Extraction:
   - Users upload their video through a Flask-based web interface.
   - The video is processed using moviepy to extract the audio, converting it into a .wav file format for further transcription.
2. Speech-to-Text Conversion:
   - The extracted audio is transcribed into English text using the speech_recognition library.
   - This transcription serves as the base content for translation.
3. Text Translation:

- The transcribed English text is translated into the target language (e.g., Hindi, Marathi, etc.) using the Google Translate API (googletrans).
- The translated text is then used as input for the next step, which is text-to-speech conversion.

4. Text-to-Speech (TTS) Synthesis:
   - The translated text is converted into speech using edge-tts, which provides natural-sounding voices.
   - Gender identification is performed by analyzing the first few seconds of the audio to ensure the correct voice (male or female) is selected for the dubbing.
5. Audio and Video Synchronization:
   - The translated speech (dubbed audio) is synchronized with the video using moviepy and pydub, ensuring that the timing of the new audio matches the original video.
   - The audio and video are then merged to create the final dubbed video.
6. Final Video Output:
   - The finalized video is rendered and stored in a downloadable format (.mp4), making it available for the user to access from the platform.

Technology Stack

- Backend: Flask (for managing video upload, processing, and user interface)
- Video Processing: moviepy, pydub
- Speech Recognition: speech_recognition
- Text Translation: googletrans
- Text-to-Speech: edge-tts (with regional language support)
- Gender Detection: inaSpeechSegmenter
- Audio/Video Merging: moviepy

User Interaction

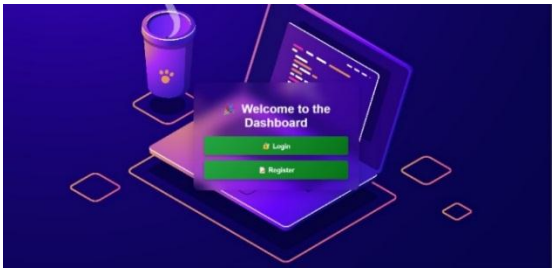
The solution is built around an easy-to-use web interface:

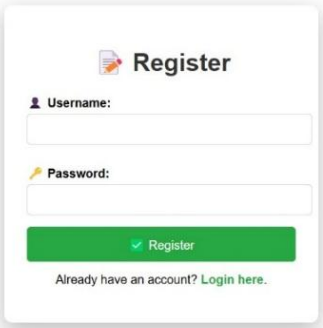
- Login & Registration: Users create an account to track their video processing activities.
- Upload: Users upload educational videos, specify their preferred target language, and initiate processing.
- Download: After processing, users can download the final dubbed video.

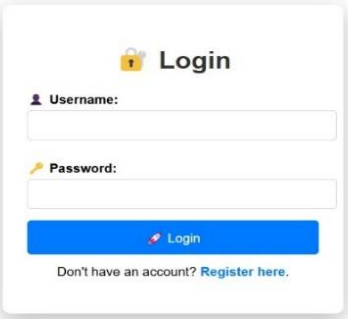
Benefits

- Accessibility: The solution makes English educational content accessible to non-English speaking users, promoting regional language education.
- Automation: It reduces manual intervention in video dubbing, speeding up the localization process.
- Scalability: The system can be extended to support more languages, voices, and customization in future iterations.

V. RESULTS







VIII. CONCLUSION

This research presents an automated system for translating English educational videos into Indian regional languages, aiming to bridge the linguistic gap in digital education. By integrating tools for speech recognition, translation, and voice synthesis, the system provides an end-to-end solution for dubbing educational content. The proposed approach not only enhances accessibility for non-English speakers but also promotes inclusive learning across linguistic demographics in India. With minimal human intervention and a scalable design, the system holds significant potential for educational platforms and content creators looking to localize learning materials efficiently.

ACKNOWLEDGMENT

The authors would like to express their heartfelt gratitude to the faculty members and project guides of SVPM College of Engineering, Malegaon (BK), Baramati, for their constant encouragement, expert guidance, and support throughout the research and development of this project. Special thanks to the Department of Information Technology for facilitating the necessary infrastructure and a collaborative environment.The authors also acknowledge the contribution of open-source tools and libraries including Flask, SpeechRecognition, googletrans, gTTS, moviepy, and pydub, which served as the backbone for implementing the multilingual video translation system. These tools significantly enhanced the capability to build and test the complete pipeline with efficiency and precision.

REFERENCES

[1] Sreelakshmi, M., & Karthika, R. (2021). Automatic Dubbing of Educational Videos in Indian Languages using Deep Learning Techniques. International Journal of Computer Applications, 183(43). DOI: 10.5120/ijca2021921196.

[2] Kishore, S., Prakash, A., & Gupta, R. (2020). A Review on Speech Recognition using Deep Learning Techniques. International Journal of Scientific Research in Computer Science, 8(3), 112–118.

[3] Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., & Macherey, W. et al. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. arXiv preprint arXiv:1609.08144.

[4] Paul, S., & Roy, D. (2021). Design of Text to Speech (TTS) Synthesis for Indian Languages. International Research Journal of Engineering and Technology (IRJET), 8(5), 2152–2156.

[5] Karlekar, S., Natarajan, T., & Whitehead, S. (2018). Detecting Linguistic Characteristics of Alzheimer's Dementia by Interpreting Neural Models. arXiv preprint arXiv:1804.06440.

[6] Kumar, M., & Singh, R. (2020). Speech-to-Text Techniques for Automatic Subtitling in Indian Languages. International Journal of Computer Applications, 176(35), 8–13.

[7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., & Jones, L. et al. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems, 30.

[8] Patel, D., & Sharma, A. (2019). A Survey on Audio-Video Synchronization Techniques in Dubbing. IJERT, 8(2), 404–408.

[9] Zhao, Y., & Li, C. (2022). Multilingual Machine Translation and Its Applications in Education. Journal of Artificial Intelligence Research, 12(3), 99–107.

[10] Subramanian, K., & Deepa, V. (2021). A Study on Text-to-Speech Systems for Low-Resource Indian Languages. Journal of Emerging Technologies and Innovative Research (JETIR), 8(9), 35–40.