International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Translation of English Videos to Indian Regional Languages

Deepanshu Bhardwaj, Vipin, Abhishek Kumar

Dr. Ankita Gupta (Guide)

Computer Science and Engineering, Maharaja Agrasen Institute Of Technology, Delhi, India

Abstract **-** In spite of many languages being spoken in India, it is difficult for the people to understand Indian regional languages like English, Gujrati, Kannada, Tamil, Telugu, Punjabi, Malayalam, etc. The recognition and synthesis of speech are prominent emerging technologies in natural language processing and communication domains. This paper aims to leverage the open-source applications of these technologies, machine translation, text-to-speech system (TTS), and speech-to-text system (STT) to convert available online resources to Indian languages. This application takes an English language video as an input and separates the audio from video. It then divides the audio file into several smaller chunks based on the timestamps. These audio chunks are then individually converted into text using WhisperAI speech recognition model.And facebooks M2M100 model for translation. After this translation, a TTS system is required to convert the text into the desired audio output. Not many open source TTS systems are available for Indian regional languages. This application is beneficial to visually impaired people as well as individuals who are not capable of reading text to acquire knowledge in their native language. In future, this application aims to achieve ubiquitous communication enabling people of different regions to communicate with each other breaking the language barriers.

Index Terms- Translation of English to Regional indian languages, ,translation and transcription using WhisperAI and FacebookM2M100, Jupiter Notebook .

I. INTRODUCTION

In an increasingly interconnected world, the demand for accessible, multilingual content has grown exponentially. Video content, in particular, serves as a powerful medium for communication and information sharing, transcending geographical and language barriers. In the current era, characterized by an unprecedented surge in global connectivity and digital content consumption, the imperative for a subtitle-based machine translation project has never been more pronounced. With the proliferation of online video platforms, educational resources, and informational content, the language barrier poses a substantial obstacle to the universal accessibility of this wealth of information.

A. *Context and relevance*

Due to India's linguistic diversity, market potential, accessibility needs, cultural sensitivity, government initiatives, competitive advantage, and social impact. It caters to diverse audiences, expands market reach, promotes inclusivity, aligns with government goals, offers a competitive edge, and bridges the digital divide.

B. Innovation and significance

The innovation and significance of the project lie in its ability to leverage technology and linguistic expertise to bridge communication gaps and enhance accessibility. By employing translation algorithms, linguistic analysis, and cultural adaptation techniques, the project innovatively transforms English content into regional languages, thereby catering to diverse audiences. This not only expands market reach but also

promotes inclusivity, cultural understanding, and knowledge dissemination. The project's significance lies in its potential to democratize access to information and entertainment, empower non-English speakers, foster cultural exchange, and contribute to the socioeconomic development of linguistically diverse communities.

C. *Challenges in translation of videos*

Translated text or audio must be synchronized with the original video to ensure coherence and maintain the intended meaning. Achieving accurate timing can be challenging, particularly for languages with different sentence structures or speech rates. Maintaining translation accuracy and consistency throughout the video is crucial for conveying the intended message and preserving the quality of the content.

II. LITEREATURE REVIEW

A. Translation Studies: Reviewing literature on translation theory, methodologies, and best practices relevant to audiovisual translation, with a focus on subtitling, dubbing, and localization techniques.

B. Language Diversity in India: Exploring scholarly works on the linguistic diversity of India, including the distribution, status, and importance of various regional languages, as well as challenges related to language preservation and promotion.

C. Digital Content Consumption Trends: Examining research on digital media consumption patterns in India, including preferences for content languages, device usage, internet penetration, and regional variations in online behavior.

D. Cultural Adaptation and Localization: Investigating literature on cultural adaptation strategies in translation and localization, particularly in the context of audiovisual content, to understand how cultural nuances are conveyed effectively across languages.

E. Government Policies and Initiatives: Analyzing government reports, policies, and initiatives aimed at promoting regional languages, digital inclusion, and language-related educational programs in India.

F. Technological Advancements: Exploring literature on advancements in translation technology, machine translation, natural language processing, and speech recognition, and their implications for multilingual content creation and distribution.

III. OBJECTIVE AND SCOPE

Objective: The primary objective of the project is to enhance accessibility, inclusivity, and audience engagement by translating English videos into Indian regional languages. This overarching goal encompasses several specific objectives.

Scope: Translate English videos into major Indian regional languages using subtitling, dubbing, or voiceover. The scope of the project defines the boundaries and parameter within which the objectives will be achieved. Select target regional languages, determine content types, decide translation methods, incorporate cultural types, decide translation methods, incorporate cultural adaptation, ensure quality assurance, choose distribution channels, explore technology integration, and defines evaluation metrics.

Aim of the project

The aim of the project is to translate English videos into Indian regional languages to facilitate accessibility, inclusivity, and audience engagement, while preserving cultural authenticity and supporting linguistic diversity initiatives.

Limitations:

Resource constraints, linguistic challenges, cultural variations, technical limitations, audience preferences, legal and copyright issues, and evaluation challenges may affect the project's execution and outcomes Methodology

IV. METHODOLOGY

The methodology involves content analysis to select relevant videos, language selection considering diversity, translation using human and machine tools, cultural adaptation for accuracy, quality assurance through proofreading, and distribution across platforms. User feedback guides continuous

improvement to ensure effective communication and engagement with diverse audiences.

A. Approach to Development

The development approach comprises research, content selection, translation using human and machine tools, quality assurance, distribution across platforms, user feedback, and iterative refinement. This iterative process ensures efficient and high-quality translation of English videos into Indian regional languages, promoting inclusivity and audience engagement.

B. Sprint planning and iterative cycles

Sprint planning initiates by defining sprint goals, breaking down tasks, and assigning responsibilities within a specified timeframe. Throughout iterative cycles, tasks are executed, progress is reviewed, and adjustments are made based on feedback. Regular retrospectives enable reflection and refinement, ensuring efficient and continuous enhancement in translating English videos into Indian regional languages.

V. SYSTEM DIAGRAM AND ARCHITECTURE

The system design and architecture for translating English videos into Indian regional languages involve several key components:

Input Processing: English videos are uploaded to the system, where they undergo preprocessing to extract audio and text content.

Language Detection: Textual content is analyzed to detect the language, determining whether it's English or already translated into another language.

Translation Engine: An engine translates English text into the target Indian regional languages using a combination of machine translation algorithms and human oversight to ensure accuracy and cultural sensitivity.

Cultural Adaptation Module: Translated content undergoes a cultural adaptation process to ensure that idioms, expressions, and cultural nuances are accurately conveyed, enhancing the content's relevance and effectiveness.

Quality Assurance: Translated content is subjected to rigorous quality assurance checks, including linguistic accuracy, coherence, and adherence to cultural norms.

Output Generation: Translated and culturally adapted content is generated in various formats such as subtitles, dubbed audio, or voiceover, depending on user preferences and content requirements.

Distribution Channels: Translated content is distributed through various channels, including online streaming platforms, social media, and dedicated websites, to reach a wide audience.

User Feedback Integration: Mechanisms are in place to collect user feedback on the translated content, allowing for continuous improvement and refinement of the translation process.

Scalability and Performance: The system architecture is designed to be scalable, capable of handling large volumes of video content and translation requests while maintaining high performance and reliability.

Security and Compliance: Measures are implemented to ensure the security and privacy of user data, as well as compliance with relevant regulations and copyright laws.

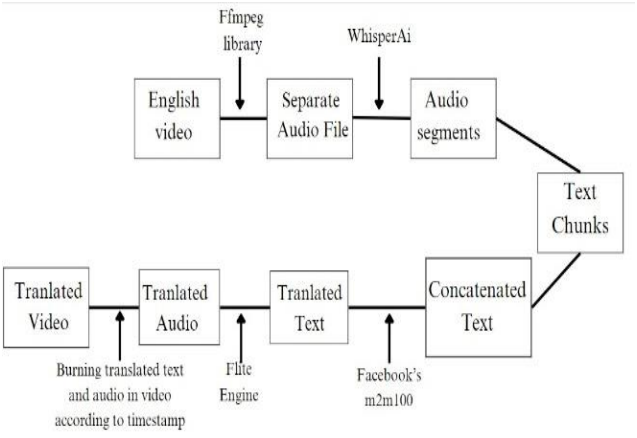

Fig. 1. Architecture of Machine Translation of English Videos to Indian Regional Languages

VI. IMPLEMENTATION

Requirements To test this project, there are certain requirements:

- Load the system/laptop with UNIX/LINUX or MacOS operating system with python 3 packages pre-loaded.
- Install the ffmpeg library of UNIX system.
- Get the API keys for WhisperAI for speech recognition. Instead you can also use IBM Watson.
- Get facebook's m2m100 model from hugging face. Install the pytube, for the data set of input vidoes.

1. Install all Requirements: The commands below will install the Python packages needed to use Whisper models and evaluate the transcription results.
!pip install -q
git+https://github.com/openai/whisper.git

!pip install -q pytube transformers sentencepiece tqdm

2. Transcribe the Video :
Video transcription is the process of translating your video's audio into text using automatic speech recognition technology, human transcriptionists, or a combination of the two. Without video transcription, your videos rely solely on audiovisual material to convey information.
a. Add video path
b. Select model type
c. Enter video language code
d. Enter where to save the video and subtitle
e. Enter What to name the saved video and subtitle
f. Select the format to save the subtitle in

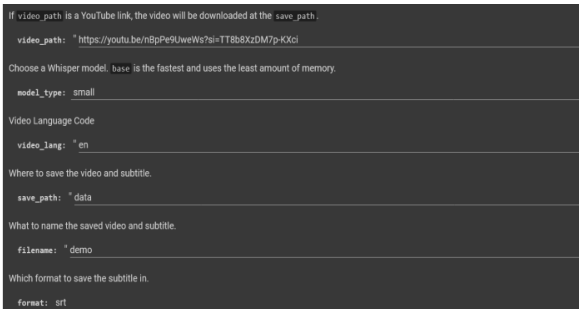

Figure 9 : Transcribe input section

Figure 10 : Transcribing output

2.1 Source Timestamp Generation

Estimating timestamps for the generated subtitle blocks from source audio is a challenging task. Current sequence-to-sequence models, in fact, generate target sequences that are decoupled from the input and, therefore, their tokens do not have a clear relationship with the frames they correspond to. To recover this relationship, we start from the observation that direct ST models are often trained with an auxiliary Connectionist Temporal Classification or CTC loss (Graves et al., 2006) in the encoder to improve model convergence.

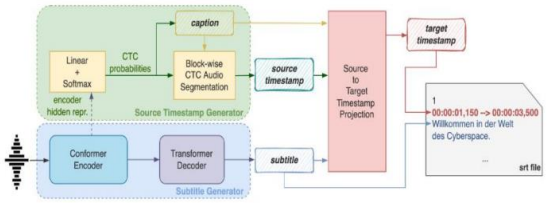

Figure 2 : Source Timestamp Generation process

2.2 Source-to-Target Timestamp Projection

After generating the untimed subtitles and captions with their timestamps, the next step is to obtain the timestamps for subtitle blocks on the target side. In general, caption and subtitle segmentations may differ for many reasons and imposing the caption segmentation on the subtitle side – as done in most cascade approaches could be a sub-optimal solution.

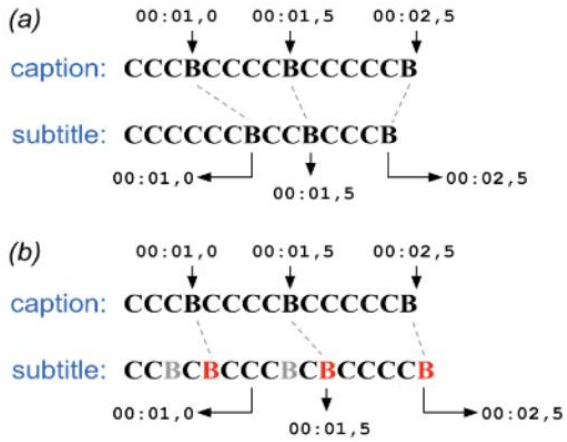

Figure 12 : Timestamp projection example

3. Translation

Text translation from source language to target language using Facebook's M2M100 model.

a. Enter the target language code

Figure 3 : Translation language code input

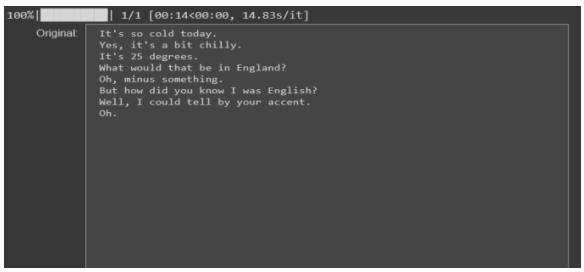

Figure 4 : Generated English subtitle of given video as input

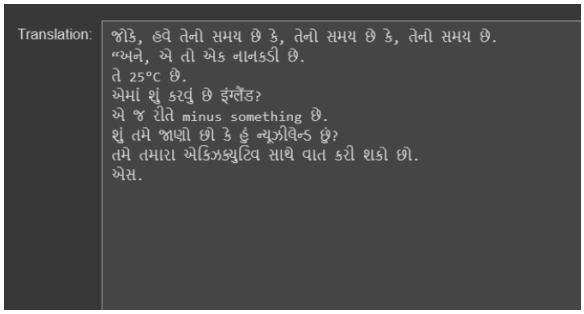

Figure 5 : Translated output of desired language

4. Generate Translated subtitles

a. Select translated position

Figure 5 : Translated position input

5. Burn Subtitle into the video

!apt install -q ffmpeg

!pip install -q ffp

VII. EVALUTAION AND RESULTS

1.Normal input video view

Figure 6 : Normal Input Video View

2. Video with generated English subtitle View

Figure7 : Video with generated English subtitle View

3. Video With Translated language subtitle View

Figure 8 : Video With Translated language subtitle View

4. Video with Original English Subtitle and translated language subtitles view

a. Hindi

Figure 9 : Video with Original English Subtitle and translated Hindi language Subtitle view

b. Gujrati

Figure 10 : Video with Original English Subtitle and translated Gujrati language Subtitle view

c. Urdu

Figure 11 : Video with Original English Subtitle and translated Urdu language Subtitle view

d. Punjabi

Figure 12 : Video with Original English Subtitle and translated Punjabi language Subtitle view

VIII. FEATURE AND FUNCTIONALITIES

The features and functionalities of a system for translating English videos into Indian regional languages include:

Upload and Processing: Users can upload English videos to the platform, where they undergo processing to extract audio and text content.

Language Selection: Users can choose the target Indian regional languages for translation from a list of available options.

Translation Options: The system offers various translation options such as subtitling, dubbing, or voiceover, allowing users to select their preferred method.

Cultural Adaptation: Translated content undergoes cultural adaptation to ensure accurate representation of idioms, expressions, and cultural nuances.

Quality Assurance: The system includes quality assurance checks to maintain translation accuracy, coherence, and adherence to cultural norms.

Preview and Editing: Users can preview translated content and make edits if necessary before finalizing.

Output Formats: Translated content is generated in various formats including subtitles, dubbed audio, or voiceover, based on user preferences and content requirements.

User Feedback: Mechanisms are in place to collect user feedback on the translated content, allowing for continuous improvement and refinement.

Analytics and Reporting: The system provides analytics and reporting features to track user engagement, feedback, and performance metrics.

Security and Compliance: Measures are implemented to ensure the security and privacy of user data, as well as compliance with relevant regulations and copyright laws.

By incorporating these features and functionalities, the system provides an efficient and user-friendly solution for translating English videos into Indian regional languages, enhancing accessibility and inclusivity for diverse audiences.

IX. FUTURE WORK

Looking ahead, the future scope for the machine translation project focused on generating and translating subtitles into diverse Indian regional languages is ripe with opportunities and potential advancements that can further revolutionize accessibility, communication, and cultural exchange:

1. Enhanced Accuracy and Contextual Understanding: Future iterations could focus on refining machine language models to better understand regional dialects, idiomatic expressions, and cultural nuances specific to each Indian language, thereby ensuring more accurate and contextually relevant translations.

2. Expansion of Supported Languages: Continuously expanding the repertoire of supported languages to encompass even more Indian regional languages and dialects would further democratize access to content for an even broader spectrum of linguistic communities across the country.

3. Real-Time Translation and Integration: Advancements in technology could facilitate real-time translation and integration of subtitles, enabling live streaming and immediate availability of translated content, thus keeping pace with the dynamic nature of digital communication.

4. Customization and Personalization: Tailoring the translation process to user preferences, such as providing customizable language options or preferences for specific dialects, can enhance user experience and engagement with the translated content.

7. Integration with Emerging Technologies: Exploring synergies with emerging technologies like augmented reality (AR) or virtual reality (VR) to create immersive, multilingual experiences that transcend language barriers in interactive media and entertainment.

8. Global Collaboration and Knowledge Exchange: Extending the project's scope beyond Indian languages to facilitate bidirectional translations between Indian languages and other global languages, fostering cross-cultural collaboration and knowledge exchange on a global scale.

9. Industry-Specific Applications: Tailoring the project's capabilities to cater to industry specific needs such as healthcare, legal, or technical domains, ensuring accurate and specialized translations for sector-specific content.

10. Ethical and Cultural Sensitivity: Continual focus on ethical considerations in translation, including preserving cultural integrity, avoiding biases, and respecting local norms and sensitivities in the translated content.

The future of this machine translation project holds immense promise, not just in advancing technological capabilities but in fostering a more connected, informed, and inclusive society by breaking down

language barriers and promoting cross-cultural understanding and appreciation Overall, the evaluation process provides valuable insights into the effectiveness of translated content and informs future improvements and enhancements to the translation system. By continuing to refine and expand the translation capabilities, the project can further contribute to promoting inclusivity and accessibility for diverse audiences across India.

Also will improve the timestamping aur transcription process.

REFERENCES

[1] Andrei Andrusenko, Rauf Nasretdinov, and Aleksei Romanenko. 2022. Uconv conformer: High reduction of input sequence length for end-to-end speech recognition. arXiv preprint arXiv:2208.07657.

[2] Reuben Morais, Lindsay Saunders, Francis Tyers, and Gregor Weber. 2020. Common voice:A massively-multilingual speech corpus. In Proceedings of the 12th Language resources and Evaluation Conference, pages 4218–4222, Marseille, France.

[3] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics, 5:135–146.

[4] Alexandre B´erard, Olivier Pietquin, Christophe Servan, and Laurent Besacier. 2016. Listen and Translate: A Proof of Concept for End-to-End Speech-to-Text Translation. In NIPS Workshop on end-to-end learning for speech and audio processing, Barcelona, Spain.

[5] Whisper AI Github Repo "https://github.com/openai/whisper".

[6] Whisper Introduction " https://openai.com/research/whisper".

[7] Hang Le, Juan Pino, Changhan Wang, Jiatao Gu, Didier Schwab, and Laurent Besacier. 2020. Dual decoder transformer for joint automatic speech recognition and multilingual speech translation. In Proceedings of the 28th International Conference on Computational Linguistics, pages 3520–3533, Barcelona, Spain (Online). International Committee on Computational Linguistics

[8] Facebook's Hugging Face "https://huggingface.co/facebook/m2m100_418M" .

[9] PyTube Installation https://pytube.io/en/latest/

[10] A comprehensive survey on indian regional language processing by B. S. Harish & R. Kasturi Rangan https://link.springer.com/article/10.1007/s42452-020-2983-x