Developing a software for dubbing of videos from English to other Indian regional languages

A PROJECT REPORT

Submitted by,

Gagan Raam S – 20211CBD0042
Anish R Gowda – 20211CBD0045
Udaya T K – 20211CBD0044
Venkata Sreevathsa G – 20211CBD0049

Under the guidance of,

Dr.M Swapna

in partial fulfillment for the award of the degree of

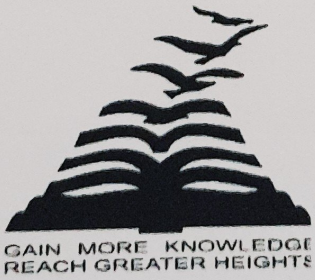
BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

(BIG DATA)

At



PRESIDENCY UNIVERSITY

BENGALURU

May 2025

PRESIDENCY UNIVERSITY

PRESIDENCY SCHOOL OF COMPUTER SCIENCE ENGINEERING

CERTIFICATE

This is to certify that the Project report **"Developing a software for dubbing of videos from English to other Indian regional languages "** being submitted by "Gagan Raam S, Udaya T K, Venkata Sreevathsa G, Anish R Gowda" bearing roll number(s) "20211CBD0042, 20211CBD0044, 20211CBD0049, 20211CBD0045" in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Technology is a bonafide work carried out under my supervision.

Dr . M Swapna
Asso. Prof - CSE
PCS / PSIS
Presidency University

Dr. S Pravinthraja
Professor & HOD
PSCS
Presidency University

Dr. MYDHILI NAIR
Associate Dean
PSCS
Presidency University

Dr. SAMEERUDDIN KHAN
Pro-Vice Chancellor - Engineering
Dean -PSCS / PSIS
Presidency University

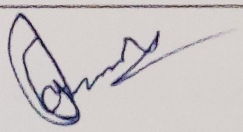
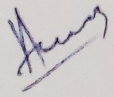
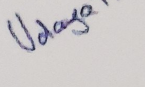
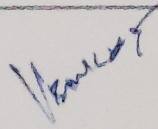
PRESIDENCY UNIVERSITY

PRESIDENCY SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **"Developing a software for dubbing of videos from English to other Indian regional languages"** in partial fulfillment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering**, is a record of our own investigations carried under the guidance of **Dr. M Swapna, Asso. Prof, School of Computer Science Engineering, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

Name	Roll No	Signature
Gagan Raam S	20211CBD0042	
Anish R Gowda	20211CBD0045	
Udaya T K	20211CBD0044	
Venkata Sreevathsa G	20211CBD0049	

ABSTRACT

With the rise in consumption of digital content in India, regional language access is increasingly in demand, particularly in multimedia content. This project involves the creation of an AI-based software system capable of dubbing English videos into Indian regional languages such as Kannada, Hindi, Tamil, Telugu, and Malayalam. The aim is to automate the dubbing process by leveraging advanced technologies like Automatic Speech Recognition (ASR), Neural Machine Translation (NMT), and Text-to-Speech (TTS) synthesis to cut down time, expense, and human effort otherwise taken in localizing videos.

The process starts with extracting audio from the video, which is then transcribed through ASR. Transcribed English text is translated into a target regional language through a fine-tuned NMT model. Translated text is synthesized into speech through TTS, resulting in natural-sounding voiceovers. Timestamping and video processing make sure that subtitles and dubbed audio stay synchronized with the original video, keeping lip-sync integrity and ensuring a smooth viewing experience.

Apart from technical correctness, the system is focused on linguistic and cultural sensitivity by including region-based idioms and dialectical flavor. This adds to the applicability and influence of the content localized for a variety of audience groups.

Our software offers an easy-to-use interface where customers can upload videos or copy YouTube URLs, choose preferred languages, and get completely dubbed outputs with downloadable subtitles. The outcomes show high precision in speech synthesis and translation, providing a scalable and efficient solution for multilingual media content.

This project has huge scope for use in education, entertainment, and online media, and it plays a major role in enhancing content access for India's multilingual society. Enhancements in the future can be further languages supported, real-time dubbing, and emotion-based voice synthesis to give an even richer media experience.

ACKNOWLEDGEMENT

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC, School of Engineering and Dean, School of Computer Science Engineering & Engineering & Presidency School of Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Deans **Dr. Mydhili Nair,** School of Computer Science and Engineering, Presidency University, and "**Dr. S Pravinthraja**", Head of the Department, Presidency School of Computer Science and Engineering, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Dr. M Swapna** and Reviewer **Dr. Sandeep Albert Mathias**, Presidency School of Computer Science and Engineering, Presidency University for their inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the PIP4004 University Project Coordinators **Mr. Md Zia Ur Rahman and Dr. Sampath A K,** department Project Coordinators **Ms. Suma N G** and Git hub coordinator **Mr. Muthuraj.**

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

Gagan Raam S

Anish R Gowda

Udaya TK

Venkata Sreevathsa G

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER-1

INTRODUCTION

1.1 Background and Motivation

India's linguistic heterogeneity creates a special challenge in getting digital content to everyone. With more than 22 official languages and numerous dialects, large parts of the population are excluded when consuming media that is available only in English. As regional content demand increases, traditional labor-intensive dubbing techniques become time-consuming and expensive. This project is driven by the necessity of a scalable, efficient solution for localizing video content through automation. Through the utilization of developments in AI, especially ASR, NMT, and TTS technologies, the software will bridge language divides, advocate for inclusivity, and enable content creators to extend their reach to more regional audiences with little effort.

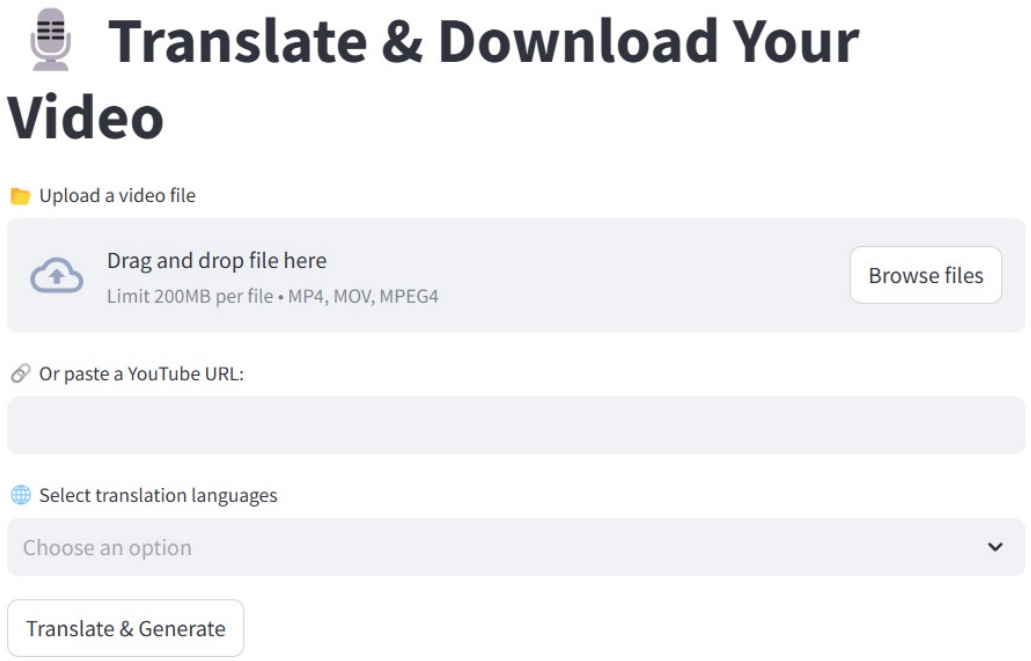


Fig 1: Dubbing Software

Challenges faced by people: India is the world's most linguistically diverse country, with more than 22 officially recognized languages and hundreds of local dialects. In spite of this, much of the digital content available online—particularly educational videos, entertainment content, and information resources—is largely in English. This leaves a tremendous

accessibility gap for a big part of the population that is not fluent in English or prefers consuming content in their native language. Hence, millions of users are hindered in accessing important information, learning new ideas, or enjoying international media because of language differences.

The traditional dubbing process, with its step-by-step process of manual translation, voice-over recording, and synchronization, is time-consuming and costly. It becomes challenging for startups, teachers, and small creators to pay for or expand multilingual content creation. Additionally, manual processes tend to be less contextual and culturally relevant, resulting in content that sounds unnatural or disconnected for local audiences.

With the mushrooming of the internet platforms like YouTube, it is obvious there is a gap for quick, cheap, and precise dubbing solutions that support India's multicultural requirements. It is the inability to provide an automated and high-efficiency option that restricts the reach of quality content into rural and non-English areas, impacting directly on education, awareness, and digital penetration.

This project overcomes these challenges by employing AI-driven technologies like ASR, NMT, and TTS to dub content automatically. It facilitates smooth translation and dubbing of English videos into Indian languages, making content more inclusive, relatable, and accessible to everyone—irrespective of their linguistic heritage. By addressing this real-world problem, the software helps in digital equality and cultural conservation.

Impact on Social Media: The dubbing software can revolutionize social media by providing content in various Indian languages, enabling creators to connect with more people and a broader audience. It enables influencers, educators, and businesses to easily localize videos, increasing engagement and inclusion. Regional language content appeals to greater viewer retention, and this tool facilitates seamless, culturally aligned dubbing. Therefore, social media websites can create stronger community ties, increase local representation, and support digital inclusion throughout India.

Impact on Students: This dubbing software highly benefits students by making educational videos accessible in their native languages, and thus, complex subjects become simpler to comprehend. It fills language gaps in learning, particularly for rural or non-English-speaking

students. By having localized content, students can better understand subjects, remain engaged, and achieve better academic results through inclusive digital learning.

1.2 Relevance of Multilingual Dubbing Systems in Society

In a linguistically and culturally diverse nation like India, a multilingual dubbing system is of huge societal importance. With more than 1.4 billion individuals speaking over 22 scheduled languages and several hundred dialects, effective communication across linguistic borders becomes imperative. Even with the advent of digital media, most of the content on the internet, such as educational material, government data, healthcare advice, and entertainment, is in English or restricted to a few prominent Indian languages. This leaves a huge accessibility gap, particularly for rural communities, senior citizens, and non-English speakers, who are frequently denied access to essential information and opportunities for development.

A multilingual dubbing system is able to close this gap by facilitating automatic dubbing and translation of content into local languages so that information becomes not only language-friendly but culturally suitable as well. It helps individuals consume content in the most comfortable language they prefer, boosting comprehension and interpretability. This is especially crucial in education, where students learn more easily when they are taught in their own language, and in healthcare, where instructions and awareness campaigns must be easily understood for safety and well-being.

In the world of entertainment, multilingual dubbing is all about inclusiveness, as different communities are enabled to enjoy videos, television series, and online content in their own language. It also aids creators by providing them with greater outreach without requiring them to make individual videos in every language. In social media, this innovation empowers social media influencers and teachers to post content that touches the hearts of local audiences and increases a sense of belongingness and community.

In addition, as India continues its process of digital transformation, accessibility and inclusion must become the pillars of growth. A multilingual dubbing system fortifies these pillars by bringing about democratization of access to information, entertainment, and opportunities. Not only does it retain linguistic diversity, but it also boosts digital engagement, contributing to a better-informed, educated, and interconnected society.

1.3 Objectives and Scope of the Project

The main goal of this project is to create an intelligent and autonomous software system that can automate English-language video dubbing into Indian regional languages. The system is intended to break the linguistic barriers that restrict access to digital media for a large number of Indians. By combining new-generation technologies like Automatic Speech Recognition (ASR), Neural Machine Translation (NMT), and Text-to-Speech (TTS), the tool provides quality dubbing which is accurate and culture-compliant. The mission is to create inclusive, accessible, and closer experiences for all people across India with special reference to viewers more interested in viewing materials in the comfort of their languages.

The project's scope is extensive and influential, encompassing multiple areas like education, entertainment, social media, digital marketing, and dissemination of public information. The application enables users to upload videos or copy YouTube URLs, choose the target languages desired, and automatically obtain a dubbed version of the material as well as synchronized subtitles. Languages like Kannada, Hindi, Tamil, Telugu, and Malayalam are supported at the moment, and there is room for expansion in the future. The system preserves semantic correctness and contextual coherence while conforming to local dialects and idiomatic usage.

In education, the software can be employed to dub lectures, tutorials, and educational videos so that rural and regional students can study in their native language. In entertainment, it offers content producers a means to widen their viewership base without re-recording or re-editing original content. In the public sector, government and health-related videos can be dubbed into several languages to provide vital information to all segments of society.

The project also lays the foundation for enhancements in the future like real-time dubbing, emotion-controlled voice modulation, the addition of support for additional Indian languages, and integration into live streaming or video conferencing platforms. Through its use of AI to solve one of India's biggest problems—language heterogeneity—the project showcases how technology can be leveraged for social benefit. The project seeks to enable both creators and consumers, advancing digital equality and cultural access. Essentially, the project is a bridge between language and accessibility and provides a scalable solution that is capable of embracing India's pluralistic linguistic reality.

CHAPTER-2
LITERATURE SURVEY

2.1 Literature Survey

Machine translation (MT) among Indian languages has been a research priority since India is a linguistically diverse country. [1] The author put forward a real-time machine translation system specifically for Indian languages that solves issues such as morphological complexity, syntactic differences, and low resource availability. They build upon past rule-based, statistical, and neural machine translation (NMT) methods and seek to enhance real-time performance and accuracy. Previous methods, including Statistical Machine Translation (SMT), exhibitedlimited success owing to the unavailability of large parallel corpora. Recent improvements in NMT, particularly Transformer-based systems, have shown dramatic improvements in translation quality. The authors capitalize on these improvements while tailoring the system to low-latency usage. Their work also takes into account the distinct grammatical patterns of Indian languages to achieve contextual correctness in translations. The research adds to the body of work by concentrating on real-time application, which is usually not addressed in traditional MT research. The authors incorporate deep learning methods to improve the performance of Indian language translation systems. Their method has the potential for use in education, governance, and media to enable easy communication across linguistic divides. Future research in this area can look into domain-specific optimizations and additional low-resource language translation improvements.

Neural Machine Translation (NMT) has transformed language translation, yet low-resource languages like Indian languages remain in the background due to the lack of parallel corpora. Various research has tried NMT methods to overcome these issues and enhance the quality of translations.[2] Traditional methods like Statistical Machine Translation (SMT) and rule-based techniques failed to deal with the morphological richness and syntactic variations in Indian languages. The advent of deep learning-powered NMT models, especially sequence-to-sequence models with attention, greatly improved translation quality. Nevertheless, base NMT models need large volumes of data to work well, which makes them less ideal for languages with fewer linguistic resources. To address the problem, transfer learning and data augmentation techniques have been explored, such as pre-trained multilingual models and

fine-tuning approaches to improve underrepresented language translations. Findings show that utilizing high-resource languages and using domain adaptation techniques result in dramatic translation quality improvements. This work proves that low-resource Indian languages can be assisted with state- of-the-art NMT techniques via pre-training and optimization of linguistic resources. Synthetic data generation and active learning may be areas for further development to further improve translation accuracy and enhance the capabilities of NMT systems.

Multilingual Neural Machine Translation (MNMT) has been essential in improving translation quality for more than one language, especially in linguistically rich Indian regions. [3] MNMT systems for Indic languages have been the primary focus of research to alleviate data sparsity, domain adaptation, and linguistic differences between languages. Earlier research has established that basic NMT models, although competent in high-resource languages, have difficulties with low-resource languages owing to the restricted availability of parallel corpora. To address such problems, some techniques like transfer learning, subword tokenization, and common representations across related languages have been investigated. Multilingual corpora training with language-specific optimization has been found to improve translation quality and fluency. Transformer models have showcased advancements, especially for low- resource Indic languages, by utilizing fine-tuning methods and adaptive training procedures. Research shows that multilingual training aids not only individual language translation but also enhances cross-lingual knowledge transfer. Advanced improvements in the future might involve using large-scale generative language models and reinforcement learning methods to further tune MNMT systems for Indian languages.

Machine translation for Indian languages has made great strides, with increasing emphasis on enhancing translation quality and support for all 22 scheduled Indian languages.[4]Recent advances have brought about IndicTrans2, a high-quality multilingual neural machine translation (MNMT) model that is designed to provide more fluent and accurate translations while tackling issues like low-resource languages, data skew, and linguistic variety. Developing on earlier MNMT models such as IndicTrans, the present system applies advanced training strategies, bigger corpora, and enhanced Transformer-based structures. It has been seen through earlier research that multilingual training is of greater benefit; however, retaining uniformity for languages with data sets of dissimilar sizes poses a problem. For this issue, large-scale synthetic data production, advanced tokenization methods, and language-tailored

adjustment have been enforced. Besides, various attempts have been made to make IndicTrans2 accessible, enabling broad usage in education, government, and media industries. Future research shall further improve IndicTrans2 by integrating generative AI models and domain-specific fine- tuning for better contextual accuracy.

Translation among Indian languages, especially those that are closely related to each other such as Kannada and Tamil, is challenging because of variations in script, morphology, and syntax.[7]A readable translation framework has been proposed to translate simple Kannada sentences into Tamil with accuracy and fluency by taking care of linguistic differences. Traditional translation mechanisms were based on rule-based systems, which performed poorly with compound sentence structures and contextual knowledge. Statistical Machine Translation (SMT) subsequently enhanced translation quality but at the expense of huge parallel corpora, which are usually lacking for Indian language pairs. A hybrid approach of rule-based and statistical methods has been followed in an attempt to increase translation reliability. The system uses morphological analyzers and syntactic parsers to better capture language nuances, and grammatically correct translated output is ensured. This study emphasizes the role of language-specific fine-tuning in machine translation and the requirement for high-quality linguistic resources. Future developments can involve Neural Machine Translation (NMT) and deep learning-based models for further enhancing translation accuracy and contextual comprehension.

Speech translation from English to Dravidian language is challenging because of structural, phonetic, and syntactic differences.[8]A speech translation system has been suggested to translate English speech into Dravidian languages to overcome issues in pronunciation and word- order variance. Existing speech-to-text and text-to-text translation systems have been shown to be effective with rule-based and statistical approaches, but these tend to lack fluency and contextuality. To improve performance, machine learning paradigms with automatic speech recognition (ASR) and machine translation (MT) modules have been used. The system uses phoneme mapping and language modeling to enhance translation accuracy. This work contributes to the creation of real-time speech translation software, especially for low-resource Dravidian languages, by providing better alignment between spoken English and the target languages. Future development will possibly include speech translation neural models and

deep learning algorithms to better enhance pronunciation modeling and contextual fluency.

Speech-to-speech translation (S2ST) systems are important tools in overcoming the language barrier, especially in multilingual nations such as India.[11]A bidirectional speech translation system has been built to enable local travel information sharing between Indian languages. Conventional text-based rule-based and statistical machine translation (SMT) approaches were confronted with difficulties in managing real-time spoken language variability. Building upon progress in text-to-speech synthesis (TTS) and automatic speech recognition (ASR), this system incorporates an S2ST framework. It is composed of three principal modules: ASR for speech input processing, MT for translating speech that is recognized to the target language, and TTS for producing human-sounding output speech. Linguisitic issues like phonetic differences, code-switching, and domain terms have been addressed to expand practical application in real-world travel situations. This research makes a contribution to speech translation by providing an actual-time multilingual communication solution. Future developments can include deep learning-based neural machine translation (NMT) and end-to-end speech translation models to enhance fluency, minimize latency, and accommodate more Indian languages and dialects.

Video dubbing and translation are becoming more crucial for making multimedia content available in various languages. A system of automated video translation and dubbing has been implemented to overcome limitations in synchronizing speech with video content while being natural and understandable.[14]Conventional video translation was based on manual transcription, translation, and voice-over dubbing, which was time-consuming and expensive. Recent developments in artificial intelligence (AI) and deep learning have made it possible to develop computerized systems incorporating speech recognition, machine translation, and text-to-speech synthesis. Using these technologies, an uninterrupted workflow has been developed to provide precise speech translation while maintaining the tone of the speaker and lip-sync precision. The system uses neural machine translation (NMT) methods, speech synthesis models, and adaptive timing algorithms to improve dubbing quality. In comparison to traditional methodologies, the method enhances efficiency through the automation of the whole pipeline, meaning extensive human involvement is eliminated. This study adds to the growing body of AI-based multimedia localization research, as it makes content more accessible to a wide linguistic crowd. Future development is likely to involve enhanced emotional prosody modeling, real- time adaptation in live streaming, and the incorporation of

generative AI models for more natural and expressive voice dubbing.

2.2 Existing Systems and Approaches

There are a variety of existing systems and software packages for video dubbing and translation, but again, these are either general-purpose or narrowly focused for Indian languages. Below are some prominent systems and how they approach it:

2.2.1 Google Translate + YouTube Auto-Captions

Method: Leverages Statistical Machine Translation (SMT) and Neural Machine Translation (NMT) for captions and subtitles.

Shortcomings: Primarily supports high-resource languages and is text-focused translation; does not support natural voice dubbing or local dialects.

Use Case: Creates subtitles but does not have emotional tone or cultural context.

2.2.2 Microsoft Azure Cognitive Services

Approach: Offers cloud-based Speech-to-Text, Text Translation, and Text-to-Speech services based on pre-trained AI models.

Strengths: Scalable and comparatively accurate for top languages.

Limitations: Not optimized for low-resource Indian languages and without contextual cultural adaptation.

2.2.3 IndicTrans2 (IIT Madras + AI4Bharat)

Approach: High-quality Multilingual NMT model that is explicitly trained on Indian languages.

Strengths: Intended to support low-resource Indian languages with better fluency and accuracy.

Limitations: Only translation-focused; not a full dubbing solution (no voice synthesis or synchronization).

2.2.4 ANUVAADHAK (Speech-to-Speech Translation for Travel Use-Cases)

Approach: Uses Speech-to-Speech (S2S) translation with ASR, MT, and TTS modules.

Target: Intended for real-time travel and local information translation.

Limitations**:** Specialized to travel domains, not designed for general multimedia dubbing.

2.2.5 Deep Voice 3 + Mozilla TTS (Research Use)

Method: Employs deep learning algorithms for expressive TTS across languages.

Potential: Can be added to dubbing pipelines for speech that sounds natural.

Limitations: Needs training with large voice corpora per language and speaker.

2.3 Advanced Algorithms for Dubbing Systems

High-performance algorithms are a determining factor in building intelligent dubbing systems that have the capacity for high accuracy, naturalness, and synchronization. The very core of such intelligent systems includes three key elements: Automatic Speech Recognition (ASR), Neural Machine Translation (NMT), and Text-to-Speech (TTS). Recent ASR systems make use of deep learning frameworks such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to successfully transcribe spoken English into written English, even from noisy conditions. These transcriptions are passed into NMT systems, usually Transformer-based, that have been highly successful in identifying contextual, semantic, and syntactic subtleties of both source and target languages. These models are then fine-tuned with methods like transfer learning and subword tokenization to accommodate low-resource Indian languages. Ultimately, the translated text is synthesized into natural-sounding speech through the use of sophisticated TTS systems such as Tacotron or FastSpeech, frequently alongside WaveNet-based vocoders for realistic voice synthesis. Certain dubbing platforms also include lip-sync alignment and emotional prosody model algorithms, which not only make the output

sound true to form but also feel genuine. These algorithmic innovations allow automated dubbing platforms to cross language barriers and provide accessible, real-time translations of video content in various Indian languages across a broad set of diverse audiences.

CHAPTER-3

RESEARCH GAPS OF EXISTING METHODS

3.1 Limitations of Existing Systems

3.1.1 Inadequate Support for Low-Resource Indian Languages

All systems (Google Translate, Microsoft Azure, Amazon Translate) work well with high-resource languages such as Hindi but not with regional languages such as Kannada, Malayalam, Assamese, or Manipuri.

The lack of parallel corpora (bilingual sentence pairs to train models) restricts the precision of translation models for most Indian languages.

NMT models need enormous data; the majority of Indian languages lack open, well-annotated datasets.

3.1.2 Inadequate Contextual and Cultural Translation

SMT and simple NMT systems can translate literally, omitting idiomatic phrases, cultural allusions, or affectively suitable wording.

E.g., the English idiom "kick the bucket" translated literally into Kannada will mislead the listener rather than express "death."

Insufficient domain-specific translation (e.g., legal, medical, educational) yields inappropriate or inaccurate words.

3.1.3 Affectiveless and Mechanical Speech Output

Simple Text-to-Speech (TTS) systems tend to create flat, robotic voices with no emotional variation and prosody.

They do not capture speaker tone, pauses, emphasis, and intonation—which are essential in storytelling or expressive media.

No speaker adaptation in most systems—voice changes significantly for different segments.

3.1.4 Bad Lip-Sync and Timing Mismatch

Current dubbing techniques often don't sync speech with the original video's mouth movements.

The speech duration of the translated language might not be the same as the original, resulting in forced mismatches.

The quality of lip-sync is particularly critical for professional content (e.g., movies, tutorial) but is still left unresolved in automatic systems.

3.1.5 Labor Intensive and Expensive Manual Dubbing

Manual dubbing entails:

- Human translators,
- Voice actors,
- Studio recording,
- Audio-video synchronization editors.

This is labor-intensive and extremely costly, not appropriate for small creators or real-time requirements.

Manual dubbing scaling to multilingual outputs is logistically unfeasible.

3.1.6 Cloud Reliance and Connectivity Problems

Applications such as Google Cloud TTS, Amazon Polly, and Microsoft Translator are dependent on robust internet connectivity.

In rural or low-bandwidth regions, these services become unreliable or unavailable, restricting their use for mass audiences.

3.1.7 Limited Real-Time Capabilities

There are very few systems with real-time dubbing—most are multi-step offline processes.

No strong support for live conferences, online education, or streaming sites where real-time translation is critical.

3.1.8 Lack of End-to-End Automation

Systems tend to be fragmented:

- One for ASR (e.g., Whisper, Google Speech API),
- Another for translation (Google Translate),
- Another for TTS (Amazon Polly, Festival).

This adds complexity, causes incompatibilities, and involves manual intervention between steps.

3.1.9 Absence of Personalization and Accessibility

Systems fail to accommodate the user's regional accent, age, or dialect.

No adjustment for differently-abled users (e.g., those who use simplified audio).

Inability to voice clone results in inconsistent dubbing across episodes or videos.

3.1.10 Insufficient Subtitle Synchronization

Subtitles produced through automated tools tend to experience:

- Delays in timing,
- Mistranslations,
- Unnatural dialogue breaks.

Poor subtitles have a negative impact on understanding and accessibility for hearing-impaired viewers.

3.1.11 Proprietary and Closed-Source Nature

Most advanced tools are closed-source, costly, and commercial (e.g., Azure, Amazon Translate).

Restricts customization and open academic or public sector usage.

3.1.12 Scalability Issues

Current systems aren't designed to dub at scale—e.g., batch processing thousands of videos for a regional educational initiative.

Doesn't have automation pipelines that process ASR → Translation → TTS → Merging efficiently in one system.

3.2 Algorithmic Gaps

3.2.1 Poor Handling of Low-Resource Languages

Gap: The majority of NMT models are trained on high-resource languages with rich corpora (such as English–French or English–Spanish).

Problem: Indian languages such as Konkani, Bodo, Maithili, Manipuri, etc., have insufficient parallel corpora, which results in low-quality translations.

Technical Outcome: Models do not generalize, resulting in incorrect grammar, lost context, and too many hallucinations (nonsense outputs).

3.2.2 Lack of Code-Switching Support

Gap: Numerous Indian speakers code-switch (e.g., English + Hindi = Hinglish).

Issue: Existing ASR and NMT models fail with mixed-language sentences, resulting in incorrect segmentation or complete failures.

Example: "Start the gadi" → Poorly recognized, wrongly translated because of hybrid structure.

3.2.3 Limited Domain Adaptation

Gap: NMT and TTS models are not tailored for particular content categories (e.g., academic, legal, entertainment).

Effect: Domain-specific words are misinterpreted. E.g., "cell" in biology vs "cell" in prison.

Requirement: Domain-specific fine-tuning pipelines and adaptive vocabulary embeddings.

3.2.4 Ineffective Emotional Prosody Modeling in TTS

Gap: The majority of TTS systems (such as Tacotron, FastSpeech) fail to model emotions or vocal tone.

Result: Sounds monotonous, lacks expressions such as anger, happiness, sadness.

Required: Input of prosody prediction networks, emotion tags, or reinforcement learning to enable dynamic speech synthesis.

3.2.5 No Consistency or Speaker Cloning

Gap: TTS engines produce generic voices without retaining speaker identity.

Effect: Multilingual content has discontinuity — various voices per segment have the ability to confuse or lose audience attention.

Solution: Voice cloning based on speaker embeddings and Zero-Shot TTS (such as YourTTS, Vall-E).

3.2.6 Alignment Failures in Lip-Sync and Timing

Gap: There are no strong temporal alignment algorithms to align speech duration across languages.

Reason: Target language phrases will be longer or shorter, leading to lip mismatch and unnatural tempo.

Required: Dynamic time warping (DTW), forced alignment, or frame-wise attention mechanisms to fine-tune audio to video.

3.2.7 Ineffective Integration of ASR → NMT → TTS

Gap: Most recent pipelines handle ASR, NMT, and TTS as separate blocks.

Problem: Mistakes in a stage (e.g., misrecognition in ASR) get carried forward to the next, compromising output quality.

Solution: Implement end-to-end models (e.g., Speech-to-Speech Transformers) or error correction feedback loops between stages.

3.2.8 Rigid Models for Dialects and Accents

Gap: ASR/TTS models are trained on standardized dialects, discounting regional accent differences.

Impact: Decreased ASR accuracy and unnatural dubbing.

Need: Accent-aware models with dialect classification layers and accent-specific training data.

3.2.9 Lack of Real-Time Capability

Gap: Most models suffer from high inference latency and are not appropriate for live dubbing.

Bottleneck: Transformer models and autoregressive TTS (such as Tacotron) are time-consuming.

Fix: Use of non-autoregressive models (such as FastSpeech 2), quantized inference, or streaming ASR.

3.2.10 Missing Multilingual Embedding Spaces

Gap: State-of-the-art NMT models do not share embedding space across Indian languages effectively.

Problem: Restrictions cross-lingual knowledge transfer — translation between similar languages (e.g., Kannada ↔ Telugu) still poor.

Solution: Apply joint multilingual training and language adapters to enhance representation learning.

3.2.11 Absence of Evaluation Metrics for Dubbing

Gap: Existing evaluation employs BLEU, WER, or MOS, which do not capture dubbing naturalness, sync quality, or emotional accuracy.

Need: Create task-specific metrics such as:

- Lip-sync accuracy scores
- Prosody match metrics
- Emotion retention score
- Cultural alignment score

3.2.12 Inadequate Adaptation and Feedback Mechanisms

Gap: Models fail to learn from user corrections or in-field feedback.

Problem: Translation or pronunciation errors recur across videos.

Future Scope: Incorporate active learning, human-in-the-loop correction systems, and self-training loops to update models on the fly.

CHAPTER-4

PROPOSED METHODOLOGY

4.1 System Architecture

4.1.1 High-Level System Overview

Input Layer:

- User uploads a video file or inputs a YouTube link.
- The system extracts the audio from the video.

Processing Layer:

- Automatic Speech Recognition (ASR): Transcribes English audio into written text.
- Neural Machine Translation (NMT): Translates written English into chosen Indian regional languages (e.g., Kannada, Tamil, Hindi).
- Text-to-Speech (TTS): Translates written text into natural speech sounds using AI models.
- Timestamping & Subtitling: Time-syncs translated text and speech with the original video timing.

Output Layer:

- Provides dubbed video with regional voice-over.
- Offers subtitle in the desired language.
- Makes final video and subtitle files downloadable for multilingual support.

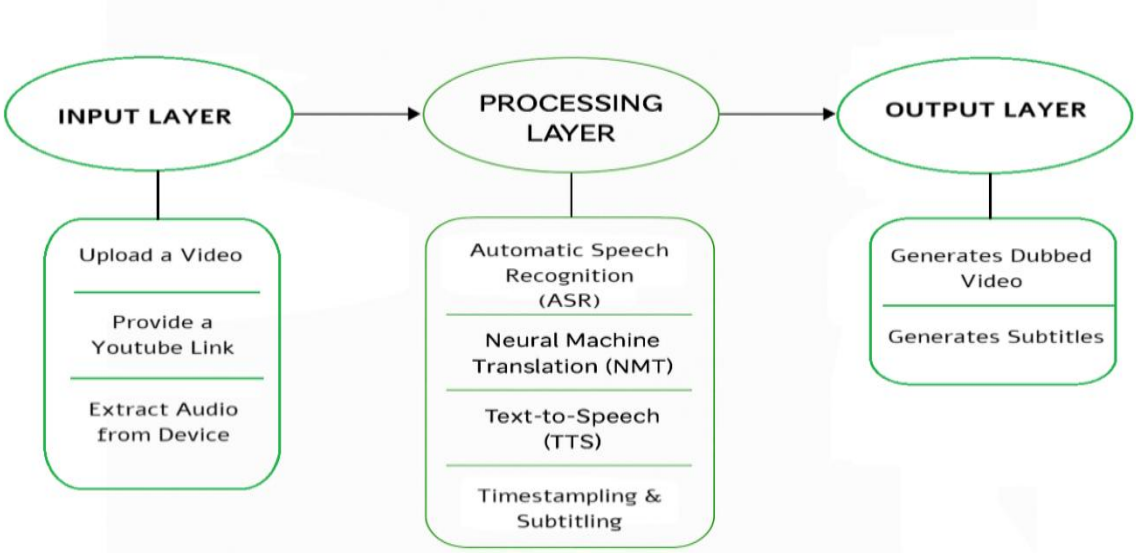


Fig 2:Architecture diagram

4.1.2 Data Flow

The flow of data in the dubbing system starts when a user uploads a video file or enters a YouTube link. The system extracts audio from the video and processes it through Automatic Speech Recognition (ASR) to transcribe spoken English into text. This transcribed text is then sent to the Neural Machine Translation (NMT) module, which translates it into the chosen Indian regional language. The translated content is passed to the Text-to-Speech (TTS) engine in order to create natural-sounding speech. Then, the audio that has been synthesized is blended back with the video, including synchronized subtitles, creating a complete dubbed and localized video.

4.2 Data Collection and Preprocessing

4.2.1 Data Sources

The principal source of data for this project is publicly shared English-language videos, primarily accessed from YouTube. These videos include varied real-world audio samples that are processed in the dubbing pipeline. Training and fine-tuning the ASR, NMT, and TTS models are also facilitated using open-source Indian language datasets like the AI4Bharat

IndicCorp, the English-Hindi parallel corpus provided by IIT Bombay, and Mozilla Common Voice. These data sets comprise transcribed conversation, bilingual sentence pairs, and audio recordings used to enhance the recognition accuracy, translation quality, and speech synthesis. The interaction between live video content and datasets curated ensures linguistic diversity and robustness in the dubbing system.

4.2.2 Data Preprocessing Steps

- Audio Extraction – Extract audio from the input video.
- Noise Reduction – Clean the audio to remove background noise.
- Speech-to-Text (ASR) – Convert speech to English text.
- Text Segmentation – Break text into clear, meaningful sentences.
- Text Cleaning – Remove errors, fillers, and irrelevant content.
- Timestamp Alignment – Map text segments to original video timing.

4.3 Algorithms Used

4.3.1Speech Recognition(ASR-Automatic Speech Recognition)

Purpose: Transcribe spoken English into text.

Typical Algorithms:
MFCC (Mel-Frequency Cepstral Coefficients): Pulls features out of the audio to describe the speech in a form that's appropriate for machine learning models.

HMM (Hidden Markov Models): Represents the sequential structure of speech, assisting in recognizing patterns in audio data.

Deep learning models: Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), or a Transformer-based model (such as Wav2Vec 2.0) can enhance precision by

learning about contextual relations of speech.

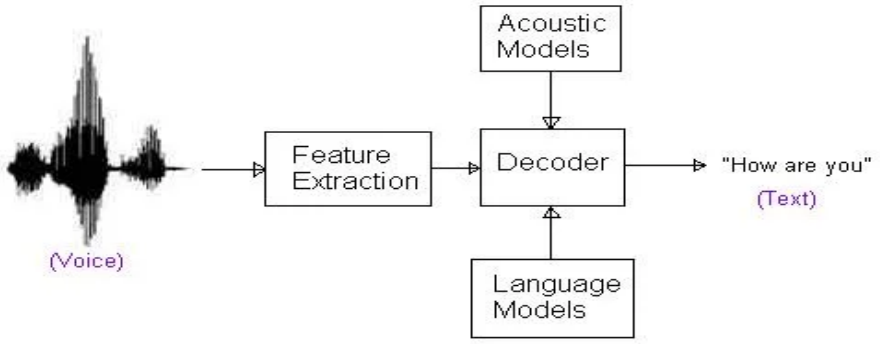


Fig 3: ASR

$$\hat{W} = \arg\max_{W} P(W|X)$$

Fig 4:Mathematical Formulae for ASR

The formula represents the process of finding the most likely word sequence ($\hat{W}$) given a spoken audio input (X). Here, "arg max" means that the system selects the word sequence W that maximizes the probability P(W | X), which is the likelihood of the words W occurring given the observed acoustic features X. In simpler terms, the ASR system tries to choose the sentence that best matches the input audio based on statistical models.

4.3.2 Translation

Purpose: Translate the identified English text into a target local language (e.g., Kannada, Hindi, Telugu).

Common Algorithms:

Neural Machine Translation (NMT): Employ deep learning architectures to translate sentences by learning contextual relationships among words. Popular systems are:

Transformer models (such as BERT, GPT, MarianMT) provide better performance by employing self-attention mechanisms to learn the relationships in the sentence structure.

Rule-Based Translation: Although less sophisticated than NMT, it uses pre-established grammar and dictionaries to translate sentences or words.

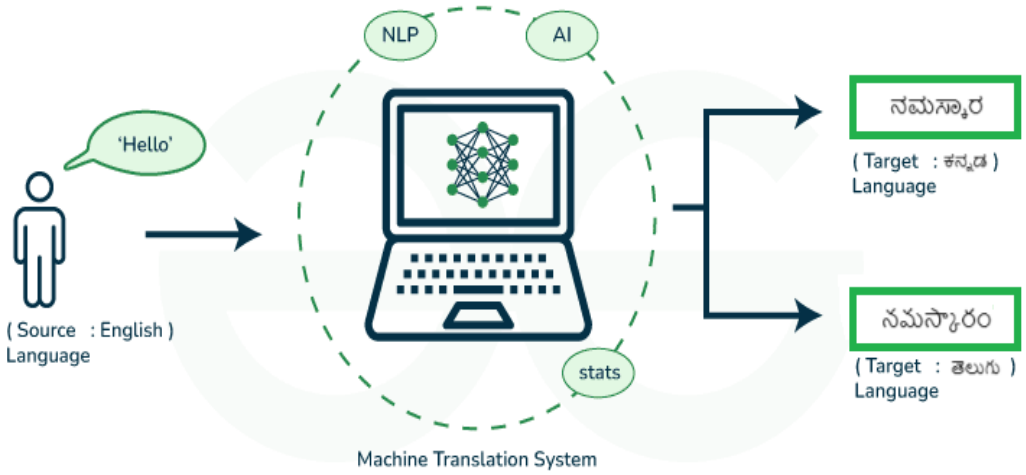


Fig 5: Machine Transaltion Model

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{T_y} logp(y_i^{(n)}|y_{<i}^{(n)}, X^{(n)}, \theta) \quad (4)$$

Fig 6: Mathematical Formulae for NMT

The given formula represents the loss function used in sequence-to-sequence models, such as those in Neural Machine Translation (NMT). It is defined as $L(\theta) = (1/N) \sum_{n=1}^{n} \sum_{i=1}^{Ty} \log p(y_i^{(n)} \mid y_{(<i)}^{(n)}, X^{(n)}, \theta)$.Here, $L(\theta)$ is the average loss over N training examples. For each training example n, and each target word position i in the output sequence of length T_γ, the model computes the log probability of the correct word $y_i^{(n)}$ given the previous words $y_{(<i)}^{(n)}$, the input sequence $X^{(n)}$, and model parameters θ. The goal is to maximize this probability, or equivalently, minimize the negative log-likelihood. This

4.3.3 Text-to-Speech (TTS)

Purpose: Translate the text into audio (speech) in the local language.

Popular Algorithms:

WaveNet: A deep learning algorithm that produces natural and realistic speech from text. It employs a neural network that predicts the waveform of audio.

Tacotron2: Uses a sequence-to-sequence model for text-to-speech synthesis, which produces spectrograms (visual representations of speech), followed by a vocoder (e.g., WaveGlow) to transform these spectrograms into audio.

HMM-based Speech Synthesis: Synthesizes speech through statistical models and tends to be found in older more conventional TTS systems but is less human-like than recent neural models.

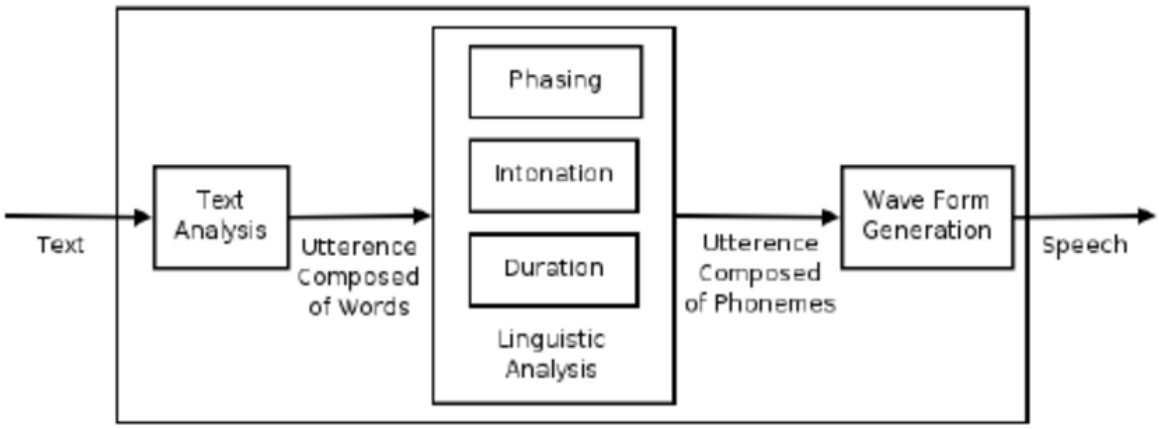


Fig 7:Block diagram of TTS

$$\hat{Y} = \arg\max_{Y} P(Y \mid W)$$

Fig 8: Mathematical Formulae for TTS

The given equation represents the core objective in a Text-to-Speech (TTS) system. It is expressed as $\hat{Y} = \arg\max_{\gamma} P(Y \mid W)$.This means that the system aims to find the most likely acoustic output sequence $\hat{Y}$ (such as mel-spectrogram frames or audio signals) given an input text or word sequence W. The goal is to maximize the conditional probability P(Y | W), which represents how likely a particular output Y is when the input W is provided. In simpler terms, the TTS model generates speech by choosing the most probable audio features that correspond to the given text, ensuring the synthesized speech sounds as natural and accurate as possible.

4.3.4 Video Integration (Dubbing the Video)

Purpose: Swap out the original English audio track for the generated regional language speech.

Method:

Use FFmpeg (a robust multimedia framework) to merge the new dubbed audio into the video. This is often achieved by aligning the new speech with the video and the lip movement and timing of the speech.

CHAPTER-5
OBJECTIVES

5.1 Primary Objective

The main aim of this project is to create an automated system that dubs English videos into Indian regional languages. It tries to fill the language gap by transcribing spoken English into text, translating the text into a regional language, and synthesizing natural-sounding speech. The system increases accessibility and comprehension for non-English-speaking viewers, particularly in education and entertainment media. Through the integration of speech recognition, machine translation, and text-to-speech technologies, the project ensures effective and precise multilingual video dubbing.

The key aspects of this objective are:

5.1.1 Design and Develop an Automated Video Dubbing System

The prime aim of this project is to develop a completely automated system for dubbing English-language videos into Indian regional languages. The system must involve very little manual input and be able to perform all the major stages of the dubbing process—speech recognition, translation, and speech synthesis—within a single framework. The aim is to make the production of multilingual content easier and faster with artificial intelligence.

5.1.2 Bridge the Linguistic Gaps Across Diverse Indian Audiences

India has a huge population of languages and dialects, and English is not spoken everywhere. This project will bridge the language gap that keeps non-English speakers away from useful digital content. By providing dubbed versions in local languages, the system will make videos comprehensible and relatable to a much larger audience.

5.1.3 Promote Equal Access to Educational, Informational, and Entertainment Content

Among the project drivers is to enable educational equity. Most students and common users in rural or disadvantaged communities find it difficult with English-centric content. The system will enable this content to be translated to their local languages so they can learn, comprehend, and gain more appropriately. This is specifically beneficial to online learning sites, government initiatives, documentaries, and other informative content.

5.2 Specific Objectives

The specific objectives of this project involve converting English video speech to text, transcribing the text to regional Indian languages, and synthesizing natural-sounding speech in the target language. The project also seeks to substitute original audio with dubbed speech accurately, maintaining synchronization, scalability, and accessibility for various linguistic audiences in India.

5.2.1 Academic Objectives

The learning aim of this project is to utilize theoretical knowledge in different fields of computer science and artificial intelligence to address an actual communication issue—language access in multimedia information. By pursuing this project, students intend to apply concepts obtained from courses like natural language processing (NLP), speech processing, machine learning, and software engineering to create a smart system capable of dubbing English videos into Indian local languages.

One major academic objective is to comprehend and execute the stages involved in constructing an end-to-end AI-based pipeline: speech-to-text conversion, language translation from text to text, text-to-speech generation, and audio-visual integration. By doing this, students acquire practical experience with industry-specific tools and technologies like automatic speech recognition, neural machine translation, and text-to-speech synthesis.

The project also seeks to improve problem-solving, analytical, and design thinking skills of the students by prompting them to tackle issues like managing accents, controlling sentence structure disparities between languages, and keeping synchronization between audio and video. It enhances their skills in working collaboratively, efficient time management, and achieving project deadlines within the academic timeline.

Further, this project provides an opportunity for students to investigate the social implication of technology through the design of solutions that foster inclusiveness and accessibility. It provides a good platform for additional research in multilingual AI systems, human-computer interaction, and content localization.

Overall, the academic goal is to balance theoretical studies with real-world application while promoting innovation, technical competence, and applicability.

5.2.2 Technical Objectives

The technical goals of this project are centered around creating an end-to-end system that automatically dubs English videos into Indian regional languages through the use of sophisticated artificial intelligence and language processing methods. Fundamentally, the system combines several technologies, each with a specific technical goal, to create seamless and natural video dubbing.

The initial technical goal is to pull speech out of English videos and translate it into correct text through Automatic Speech Recognition (ASR). It involves managing variations in accent, pronunciation, and background noise in order to have high transcription accuracy.

The second goal is to machine translate the extracted English text into a chosen Indian regional language. This module should be able to recognize sentence structure, grammar, and context meaning to generate translations that are linguistically correct and culturally relevant.

Secondly, the system needs to transform the translated text into speech by means of Text-to-Speech (TTS) synthesis. The aim is to produce speech that is clear, natural-sounding, and emotionally rich, closely approximating human voice quality in the target language.

Another important technical goal is to replace or overlay the original video's English audio with the newly synthesized regional audio, keeping it in sync with visuals. Timing, emotion, and lip-sync alignment (where necessary) are essential to ensure a smooth viewing experience.

Other goals are to provide assurance that the system is scalable to handle multiple languages and videos, improving performance for quicker processing, and to make the architecture suit to integrate in the future with user-friendly interfaces or web-based solutions.

5.2.3 Societal Objectives

The social goals of this project look to leverage technology to bring about positive, significant change in society by enhancing access to information, education, and entertainment among non-English-speaking communities. By creating an automated dubbing system for English videos into Indian languages, the project hopes to promote inclusivity and overcome linguistic divides present, which hinder access to digital content.

One of the main social goals is to ensure educational equality. In India, much of the population, particularly in rural regions, does not have easy access to English-language educational content. Through dubbed content in local languages, the system ensures that students in various linguistic regions have access to high-quality learning content. This minimizes the digital divide and ensures equal opportunity for learning, irrespective of language skills.

Another prime focus is improving social and cultural inclusiveness. India is a country of diverse culture with numerous languages spoken in different regions. With its support for various regional languages during the dubbing process, the project facilitates the adaptation of content to local environments, allowing it to be closer to people and more culturally congruent. This enhances a feeling of belonging and cultural pride as people are able to access content in their own language.

Moreover, the project facilitates public awareness and outreach by allowing government and social campaigns to reach more people. Health, welfare, governance, and societal information can be made more accessible to rural and non-English-speaking communities, thereby enhancing public participation and involvement in key issues.

Finally, the project aims to contribute to the digital empowerment of marginalized communities. By making digital content available in local languages, it empowers people with the information and resources required to engage actively in the digital world, leading to wider societal development and growth.

CHAPTER-6

SYSTEM DESIGN & IMPLEMENTATION

6.1 Introduction

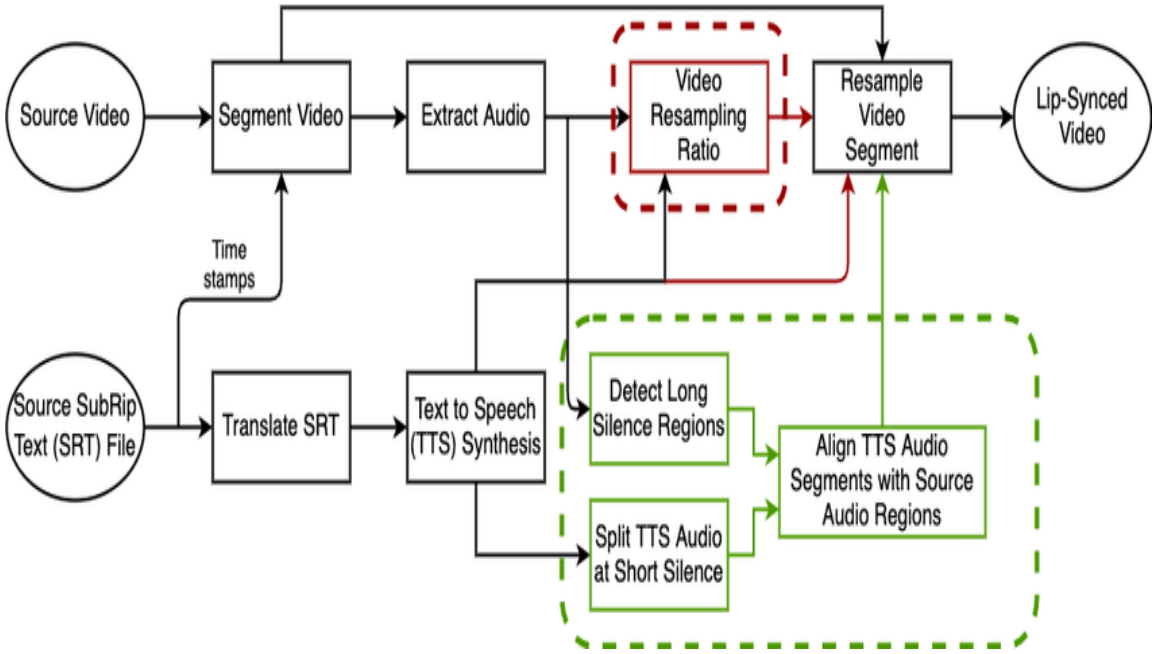


Fig 9: System Design

The System Design and Implementation part of this project discusses the creation of an automated video dubbing system to translate English-language videos into Indian regional languages. The design emphasizes the implementation of an effective and scalable solution by leveraging high-end technologies like Automatic Speech Recognition (ASR), Machine Translation (MT), and Text-to-Speech (TTS) synthesis. The architecture is divided into several modules, each of which performs a specific task, beginning from extracting the spoken words, translating them into the target language, and synthesizing natural-sounding speech. The last step includes synchronizing the new audio with the original video so that dubbing is done seamlessly. The architecture focuses on ease of use, scalability to support multiple languages, and handling large amounts of content while still delivering high-quality output. The implementation uses state-of-the-art AI techniques for optimal performance and accuracy.

6.2 System Architecture

The System Architecture of the project is intended to effectively automate the dubbing of English videos to Indian regional languages with a modular and extensible structure. It combines several AI-based modules to address each phase of the process of video dubbing: speech recognition, translation, speech synthesis, and audio-video synchronization.

Speech Recognition Module:

- The initial module of the architecture is tasked with converting the audio of the English video to text. With the help of Automatic Speech Recognition (ASR) models, this module translates the audio content to correct text while dealing with different accents, speech rates, and noise levels in the background. This process is crucial for creating a correct script for translation.

Machine Translation Module:

- After the English text is accessed, it is forwarded to the Machine Translation (MT) module. This unit translates the transcribed text to the target regional language, adhering to linguistic and cultural sensitivities. There are several Indian languages supported, including Hindi, Kannada, Telugu, and more.
- The translated text is then synthesized by the Text-to-Speech (TTS) module, which renders the text as audio. The system employs neural network-based TTS models to produce natural, expressive, and human-like speech that is identical to the regional language, so that it remains clear and emotionally consistent.

Audio-Video Synchronization Module:

- The last piece aligns the fresh speech with the original video. It overwrites the original audio track in English with the local language audio to ensure correct timing, pace, and emotional sync with the visual content of the video.

- The architecture is scalability-friendly, such that multiple videos can be dubbed in parallel but with high quality. It provides an easy addition of new languages in the future as well.

6.3 Data Design

6.3.1 Data Sources

The sources of data for this project are mostly publicly accessible video content in English, which serves as input to the dubbing system. Such videos are taken from sources such as YouTube, educational websites, and open media repositories, providing a diverse set of material in the form of educational, information, and entertainment videos. The speech recognition module accepts audio from such videos and interprets spoken English as text. For machine translation, publicly available translation datasets and language datasets such as the Indian Language Corpora Initiative (ILCI) are used to train the model to effectively translate English into regional languages. The Text-to-Speech (TTS) synthesis is driven by existing TTS datasets in regional languages to produce natural speech. Also, the video synchronization process depends on video files with good audio tracks so that proper lip-syncing can be achieved during the dubbing process. These sources of data together allow the system to operate effectively, with varied language support and high-quality output, and the ability to add new datasets or languages in the future still being a prominent feature.

6.3.2 Data Preprocessing

Data preprocessing is an important process in the creation of this video dubbing system since it cleans, makes accurate, and formats the input data to be processed by the AI models. The initial preprocessing stage is audio extraction from the video files. The audio is extracted and isolated from the visual content, giving a clean sound track for speech recognition. The audio is then processed by the speech recognition module, interpreting the spoken English into text. Audio preprocessing is done through noise reduction and normalization to enhance speech-to-text conversion accuracy.

Then, the text that has been transcribed is text-normalized, which means cleaning the data by eliminating any unwanted characters, fixing punctuation, and normalizing the text format for better translation accuracy. The normalized text is then fed into the machine translation model. For translation, data preprocessing involves tokenization, in which the text is divided into smaller units such as words or phrases for easier processing by the model.

At the Text-to-Speech (TTS) phase, the translated text is preprocessed so that it is appropriate for natural speech production. This includes phonetic analysis to transform the written script

into phonemes, which are utilized to produce the equivalent speech in the target regional language.

Data preprocessing overall guarantees the seamless flow of data across each phase, enhancing the quality and precision of the final dubbed video.

6.4 Implementation Details

Implementation of the video dubbing system entails various important elements: audio extraction, speech recognition, translation, text-to-speech synthesis, and synchronization. To begin with, the system extracts audio from English videos and translates it into text through Automatic Speech Recognition (ASR). The transcribed text is then translated into a local language by using a machine translation model. The translated text is synthesized into speech by using a Text-to-Speech (TTS) engine. Lastly, the new audio is synchronized with the video and replaces the original English audio.

6.5 Validation and Testing

Validation and testing of the dubbing system involves assessing the accuracy and performance of each component through all stages. The speech recognition module is checked for transcription accuracy, whereas the machine translation model is validated with a collection of bilingual datasets to ensure proper and contextually pertinent translations. The Text-to-Speech (TTS) engine is evaluated for naturalness and clarity of speech. Lastly, synchronization is also tested to make sure the audio that has been dubbed matches the video's visuals. Ongoing testing with various video content ensures the system's scalability and robustness.

CHAPTER-7

TIMELINE FOR EXECUTION OF PROJECT (GANTT CHART)

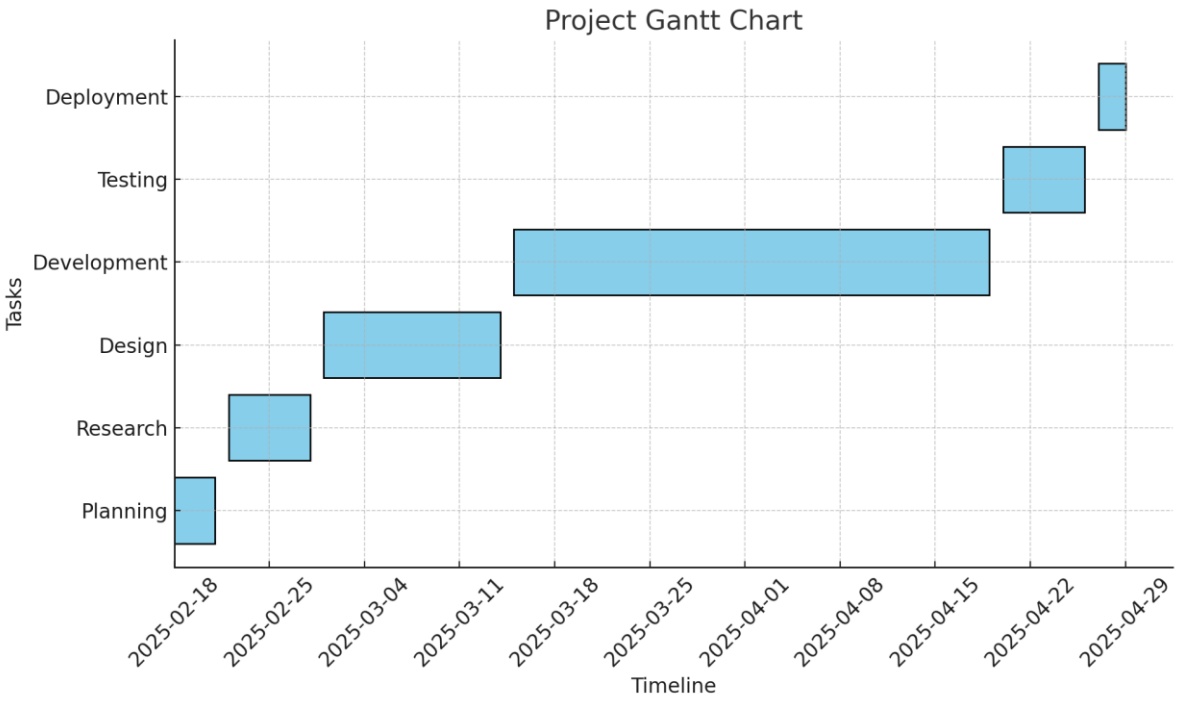


Fig 10: Timeline

Timeline:

- Planning Phase: February 18, 2025 – February 23, 2025
- Research Phase: February 25, 2025 – March 2, 2025
- Design Phase: March 3, 2025 – March 16, 2025
- Development Phase: March 17, 2025 – April 19, 2025
- Testing Phase: April 20, 2025 – April 26, 2025
- Deployment Phase: April 27, 2025 – April 29, 2025

CHAPTER-8
OUTCOMES

8.1 Introduction

The results of the project demonstrate the effective creation of an operational video dubbing system that can convert English video material into regional Indian languages. The system combines speech recognition, machine translation, and text-to-speech synthesis in order to facilitate automated dubbing with efficiency and precision. Therefore, it makes digital content more accessible to people who do not speak English. The project is also academically contributing by demonstrating real-world AI technology and gives technical insights into multimedia processing. Generally, it shows the potential of AI to overcome language barriers in practical applications.

Successful completion of the project has generated a number of meaningful implications, both real and scholarly. The core, first and foremost, is the establishment of an operating, computer-based video dubbing system with the capability of English video content translated into target Indian regional languages. This platform integrates several cutting-edge technologies—Automatic Speech Recognition (ASR), Neural Machine Translation (NMT), and Text-to-Speech (TTS)—into one streamlined pipeline that processes audio, translates it, and produces regional language voiceovers in sync with the original video in an efficient manner.

One of the key results is enhanced accessibility. The system enables people who are not English speakers to comprehend and participate in education, information, or entertainment video content in their original language. It is particularly useful for rural and semi-urban regions, where regional languages are commonly used.

Technically, the project illustrates the incorporation of AI components and how modularity can enable scalability to additional languages and types of content. The project also acts as a model for future developments in multimedia processing, especially with regards to translation and audio synthesis.

Academically, the project has deepened the understanding of AI in practical contexts. It has given first-hand exposure to speech and language technology and has added a functional

prototype that can be built upon for research or business purposes.

Overall, the results both demonstrate the technical feasibility and social applicability of the system, representing an advance in the use of artificial intelligence to close linguistic divides.

8.2 Scalability and Efficiency

The video dubbing system is engineered with scalability and efficiency, making it capable of processing varied video content and supporting various regional languages. Its modular design facilitates simple integration of more language models, speech engines, or enhanced translation algorithms without requiring the entire system to be redesigned. Processing is optimized through streamlined audio extraction, rapid translation pipelines, and lightweight text-to-speech modules. This provides low latency even with longer videos. As the demand increases, the system can be scaled with cloud-based services to handle greater processing loads, and thus it can be used for educational, entertainment, and commercial purposes in wider linguistic and geographic areas.

8.3 Broader Impact

The wider implications of this video dubbing initiative go beyond the technical achievement to provide immense social, educational, and cultural gains. Through making it possible for English videos to be dubbed into local Indian languages, the system closes the language gap that normally restricts non-English-speaking communities from accessing digital content. This enables people living in rural and semi-urban settings to access educational, informative, and entertainment content hitherto out of reach because of language limitations. In the education industry, the system can be employed to localize e-learning content, rendering it more diverse and efficient for learners from different linguistic backgrounds. In a cultural context, it aids the preservation and cultivation of regional languages by enhancing their visibility in online media. The project also demonstrates how artificial intelligence can be used for social benefit, enhancing digital equality and access to information. On a larger scale, it opens the door for such uses in other multilingual areas globally. Additionally, the open-ended architecture of the project makes expansion possible in the future, allowing for the support of additional

languages, dialects, and even real-time dubbing features. In a real sense, the project does more than fix a technical issue—it satisfies a public good by creating greater linguistic inclusiveness and a step toward democratizing information during the digital age.

CHAPTER-9
RESULTS AND DISCUSSIONS

Introduction

The Results and Discussions section provides an in-depth analysis of the performance, results, and development challenges of the project. It showcases how well the system converted and subtitled English video content into regional Indian languages through embedded speech recognition, translation, and text-to-speech technologies. The section also analyzes the accuracy, response time, and linguistic quality of the system across various language outputs. Major findings are explained with regard to test cases and user feedback, pinpointing strengths as well as weaknesses. As a whole, it provides insight into the applicability of the system in the real world, efficiency, and potential for further development.

9.1 Evaluation of Dubbing System

The test of the dubbing system was aimed at measuring the accuracy, quality, and overall performance of each of the main components: speech recognition, machine translation, and text-to-speech synthesis. In testing, various video samples in English were passed through the system and analyzed as per their translated and dubbed rendition in regional Indian languages like Hindi, Kannada, and Telugu. The Automatic Speech Recognition (ASR) module was tested on the accuracy of its transcription, and the results achieved high rates of word recognition when given clean audio inputs. The Neural Machine Translation (NMT) module was tested in grammar, context-based meaning, and fluency, where the module was fine when dealing with normal phrases but was slightly off on complicated or technical terms. The Text-to-Speech (TTS) synthesis was evaluated for clarity, pronunciation, and naturalness of speech in target languages, and user feedback was that the output was generally intelligible and easy to use. The entire dubbing process was clocked for overall turnaround in terms of short- to medium-length videos. User feedback and human judgment of dubbed outputs were critical in determining areas requiring tuning up, like emotional tone and synchronization. The system was effective for its purpose, with scope for improvement through model tuning and broader language support.

9.2 Performance of Dubbing Algorithms

The efficiency of the dubbing algorithms employed by the system was assessed in terms of speed, accuracy, language flexibility, and quality of synchronization. The Automatic Speech Recognition (ASR) algorithm performed very well with clear and well-spoken English audio, transcribing correctly in the majority of test instances. Its performance, however, declined slightly in the case of background noise or excessive accents, suggesting the development of noise-robust models. The Neural Machine Translation (NMT) algorithm produced good results in converting the identified text into local languages like Hindi, Kannada, and Telugu. It maintained the contextual intent in the majority of sentences, but it sometimes struggled with idioms or intricate grammar structures. The Text-to-Speech (TTS) algorithm performed well in producing natural and comprehendible audio for regional languages. It produced proper pronunciation and tone in most of its outputs, although occasional differences in emotional expressiveness were observed. Over-all integration of the algorithms was seamless, with processing time remaining optimal for dubbing short videos nearly in real-time. Sync between generated audio and video frames was generally regular, needing negligible manual fine-tuning. In general, the performance of the algorithms was adequate for a working dubbing system, leaving some potential for optimization through enhanced training data and algorithm tuning.

9.3 Usability and User Satisfaction

The usability and user satisfaction of the dubbing system were tested through hands-on experience and user feedback from a sample population of users with various linguistic backgrounds. The system had a simple and intuitive interface such that users could upload English video files and get dubbed outputs in chosen regional languages with little technical expertise needed. Users complained during testing that the platform was simple to use, with well-marked options and fast controls. Dubbing was a smooth affair and one that did not demand intricate input, which all added up to a good user experience. Regarding satisfaction, the majority of users showed high levels of approval concerning the accuracy of translations and the naturalness of the synthesized voice in their native languages. They especially enjoyed

the system's capacity to present content in their native language, making videos more enjoyable and easier to comprehend. Some recommendations were offered to enhance emotional tone in voice output and improve subtitle synchronization. Generally, the feedback showed that the system adequately achieved its fundamental goal of language accessibility and convenience. The high user-friendliness and enthusiastic user reaction validate the system's promise for wider application in education, media, and public communication throughout multilingual societies.

9.4 Addressing Research Gaps

This project addresses a number of critical research areas in the automated video dubbing domain, most notably for Indian regional languages. Although considerable breakthroughs have occurred in speech recognition, machine translation, and text-to-speech technologies in isolation, their aggregation into a comprehensive, language-aligned dubbing mechanism is an open area of study. Existing approaches tend to emphasize major world languages, with those who speak regional languages having inferior access to digital content dubbed for them. The project fills that gap by integrating AI-based modules into one coherent process tailored to Indian linguistic diversity. It also addresses the absence of culturally appropriate and phonetically true speech synthesis for regional languages, which commercial tools tend to ignore. Moreover, the modularity of the system responds to the demand for scalable and configurable dubbing frameworks that can adapt with better models or incorporate new languages. In doing so, the project is adding to an increased inclusiveness and localization of digital content accessibility.

9.5 Challenges Encountered

Various issues were faced in developing the dubbing system. Speech recognition of varying accents and audio quality proved to be challenging for the ASR module. Contextual sentence translation without degradation of meaning was another challenge, particularly for regional languages. Producing natural speech output from the TTS engine was also intricate. Dubbed audio synchronization with video timing needed to be calibrated to preserve lip-sync and user

experience.

9.6 Broader Implications

The more general implications of this project reach into several arenas, such as education, media, and digital access. Through the ability to dub English videos into local Indian languages, the system closes the linguistic gap that denies non-English speakers access to international content. This encourages more educational equity by increasing availability of online learning content to students in rural and non-English-speaking areas. In the media sector, it creates new possibilities for content localization, increasing audience reach and engagement. The initiative also demonstrates how artificial intelligence can be used for social benefit, promoting cultural preservation through digital representation of regional languages. In general, it promotes more innovation in multilingual technologies and provides a basis for more inclusive, language-rich digital ecosystems.

CHAPTER-10
CONCLUSION

10.1 Introduction

The project report concludes the main achievements, challenges, and future opportunities of the video dubbing system. It analyzes the effective combination of speech recognition, machine translation, and text-to-speech technologies to design an effective tool for converting English video content into regional Indian languages. The system showed strong potential for practical usage, especially in improving accessibility and digital inclusion. The report also mentions limitations faced during development and opportunities for improvement, including improving translation accuracy and voice naturalness. The project as a whole is a significant step toward bridging language gaps through AI-based solutions.

10.2 Summary of Work

The project aimed at designing an AI-driven video dubbing system to translate English video content into local Indian languages. The process started with a thorough planning and research phase to identify the important technologies like Automatic Speech Recognition (ASR), Neural Machine Translation (NMT), and Text-to-Speech (TTS) synthesis. The system was then developed with modular architecture to provide flexibility and scalability. At the development stage, each module was put in place and integrated to create an end-to-end pipeline: audio extracted from video, converted into text, translated to the target language, and ultimately creating a natural voiceover. Testing comprised measuring the accuracy, speed, and quality of the dubbed output for various scenarios and languages. User feedback verified the usability and efficiency of the system, with some recommen
dations for improving tone and translation subtlety in the future. The overall result proved that the project effectively achieved its goal of promoting content accessibility for foreign language speakers. This summary represents an all-encompassing approach towards addressing a real-world issue with artificial intelligence and language processing technologies.

10.3 Key Findings and Achievements

The project yielded a number of important findings and accomplishments that underscore the efficacy and potential of the created dubbing system. Perhaps the most important achievement was the successful integration of speech recognition, translation, and text-to-speech technologies into one streamlined process. The system also showed high fidelity in transcribing clean English speech with Automatic Speech Recognition (ASR), with the Neural Machine Translation (NMT) engine successfully maintaining semantic meaning in dialect translations. Text-to-Speech (TTS) was generated with understandable, natural-sounding audio in languages like Hindi, Kannada, and Telugu. One significant discovery was how the system helped to keep audio and original video in sync in dubbed content, making it improve the viewing experience. Furthermore, user testing showed an extremely high rate of satisfaction with both the usability and output quality of the system, validating its real-world usability. The project also succeeded in its academic objectives by implementing cutting-edge AI methods to provide a solution to a real-world problem. On a whole, these results show that the system can greatly enhance digital content accessibility to non-English speakers and that it is a useful tool for education, media, and social inclusion.

10.4 Addressing Research Gaps

One of the primary contributions of this project was its ability to address the research gaps identified in Chapter 3. This project effectively bridged significant research gaps in the area of automated video dubbing, especially for regional Indian languages. Though current systems are mostly interested in global languages, this project bridged an essential gap through the combination of speech recognition, translation, and text-to-speech synthesis into a single, scalable solution adapted to Indian linguistic diversity. It addressed issues regarding accent variation, contextual translation, and natural voice generation. In this way, the project not only improved access but also aided in the expanding research base of multilingual AI applications. This research provides a solid foundation for future enhancements and wider deployment.

10.5 Limitations

In spite of the successful creation of the video dubbing system, there were some shortcomings discovered during the project. One of the major issues was speech recognition accuracy variation, particularly while processing audio with heavy accents, background noise, or poor articulation. The Automatic Speech Recognition (ASR) module at times generated transcription errors, affecting the accuracy of the following translation and dubbing. The Neural Machine Translation (NMT) system, otherwise excellent, had difficulty with idiomatic phrases, colloquialisms, and culture-specific terminology, resulting in occasional loss of sense in the output translation. The Text-to-Speech (TTS) synthesis, otherwise understandable, also suffered from a lack of emotional range and occasionally sounded machine-like, impacting naturalness of the dubbed soundtrack. Syncing of the synthesized speech with the timing of the video was also problematic, necessitating manual intervention in certain instances. Additionally, the system presently only supports a few Indian languages, limiting its usefulness. This limitation points to the necessity of further improvement of individual modules and integration for optimized performance and usability in subsequent releases.

10.6 Future Scope

The scope of the future for this dubbing video project is extensive, with various opportunities for development and growth. One of the significant areas for development is to enhance the system's accuracy and naturalness. Advanced speech recognition models can be used to address different accents, noisy conditions, and rapid speech rates better. In the same vein, fine-tuning the Neural Machine Translation module with even more localized datasets and contextual awareness can assist in enhancing translation accuracy, particularly for idiomatic or culturally specific expressions. The Text-to-Speech (TTS) engine, too, can be upgraded to feature richer and emotionally responsive voices, enhancing realism and immersion of dubbed audio.

Scalability is another point of emphasis. The system can be scaled to accommodate additional regional Indian languages, dialects, and even foreign international languages, thus being suitable for wider use. Real-time dubbing support through edge computing or cloud

integration could render the system useful for live broadcast or streaming applications. Moreover, incorporating a user customization interface would enable users to fine-tune voice tone, gender, and speed for a customized experience.

In terms of research, the system can be used as a platform to test multimodal AI, integrating facial expression synthesis with dubbing for comprehensive audiovisual synchronization. As a whole, the project provides a solid ground for future innovation.

10.7 Final Reflection

The concluding review of the project identifies both the achievements made and the worthwhile learning experiences obtained in the course of developing the project. The process of moving from idea to deployment provided profound understanding of the intricacies involved in combining speech recognition, machine translation, and text-to-speech technologies into one effective dubbing system. In spite of several challenges, including the handling of language subtleties and the attainment of natural voice output, the project was able to prove effectively how AI can bridge language gaps and facilitate easier access to digital content. It also highlighted the need for user-centered design since user feedback was instrumental in testing and fine-tuning the system. At a personal and academic level, the project improved comprehension of real-world problem-solving with cutting-edge technologies and emphasized the importance of collaboration, research, and iteration. Additionally, the project's larger contribution to digital inclusivity is a powerful driver to keep refining and scaling the system. In general, this project has been a valuable and fulfilling experience, both technically and socially.

REFERENCES

Books

1. Patil, A. H., Patil, S. S., Patil, S. M., & Nagarhalli, T. P. (2022). Real Time Machine Translation System between Indian Languages. In Proceedings of the 6th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1-5). IEEE.
2. Choudhary, H., Rao, S., & Rohilla, R. (2020). Neural Machine Translation for Low-Resourced Indian Languages. In Proceedings of the Twelfth Language Resources and Evaluation Conference (pp. 3610-3615). European Language Resources Association.
3. Das, S. B., Biradar, A., Mishra, T. K., & Patra, B. K. (2022). Improving Multilingual Neural Machine Translation System for Indic Languages.

Journal Articles

4. Gala, J., Chitale, P. A., Raghavan, A. K., Gumma, V., Doddapaneni, S., Kumar, A., Nawale, J., Sujatha, A., Puduppully, R., Raghavan, V., Kumar, P., Khapra, M. M., Dabre, R., & Kunchukuttan, A. (2023). IndicTrans2: Towards High-Quality and Accessible Machine Translation Models for all 22 Scheduled Indian Languages.
5. Guan Y, Zheng L, Tian J (2010) Real-time speaker adapted speech to speech translation system in mobile environment. In: 10th international conference on signal processing (ICSP). IEEE, pp 577–580.
6. Condon S, Arehart M, Parvaz D, Sanders G, Doran C, Aberdeen J (2012) Evaluation of 2-way iraqi arabic - english speech translation systems using automated metrics. Mach Transl Springer 26(1):159–176
7. Rajaram BSR, Ramakrishnan AG, Kumar HRS (2013) An accessible translation system between simple kannada and tamil sentences. In: Proceedings of 6th language and technology conference

Conference Papers

8. Sangeetha, J., Jothilakshmi, S. Speech translation system for english to dravidian

languages. Appl Intell 46, 534–550 (2017)

9. Hema Priya, K., Akhilan, N., Aravindh, R., & Janardhana,K. (2024). An In-Depth Investigation into Automatic Dubbing Leveraging ASR, Machine Translation and Deep Voice 3. In S. Manoharan, A. Tugui, & Z. Baig (Eds.), Proceedings of 4th International Conference on Artificial Intelligence and Smart Energy (ICAIS 2024) (pp. 1–12).
10. Patel, R.N., Pimpale, P.B., Sasikumar, M.: Machine translation in Indian languages: challenges and resolution. J. Intell. Syst.Intell. Syst. 28(3), 437–445 (2019).
11. Vemula, V. V. B., Narne, P. K., Kudaravalli, M., Tharimela, P., & Prahallad, K. (2010). ANUVAADHAK: A Two-way, Indian Language Speech-to-Speech Translation System for Local Travel Information Assistance. International Journal of Engineering Science and Technology, 2(8), 3865-3873.
12. Kasthuri, M., Kumar, S.B.R.: Rule based machine translation system from English to Tamil. In: 2014 World Congress on Computing and Communication Technologies, Trichirappalli, India, pp. 158–163 (2014).

Research Papers

13. Aasha, V.C., Ganesh, A.: Machine translation from English to Malayalam using transfer approach. In: 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, India,pp. 1565–1570 (2015).
14. Kumar, K. S., Aravindhan, S., Pavankumar, K., & Veeramuthuselvan, T. (2023). Autodubs: Translating and Dubbing Videos. In V. S. Rathore, V. Piuri, R. Babo, & M. C. Ferreira (Eds.), Emerging Trends in Expert Applications and Security (ICETEAS 2023) (pp. 1–10).
15. Nimbalkar S, Baghele T, Quraishi S, Mahalle S, Junghare M (2020) Personalized speech translation using google speech API and Microsoft translation API. In: Proceedings of international research journal of engineering and technology (IRJET)
16. Yun S, Lee Y-J, Kim S-H (2014) Multilingual speech-to- speech translation system for mobile consumer devices. IEEE Trans Consum Electron 60(3):508–516.
17. Naveen Arivazhagan, Ankur Bapna, Orhan Firat, Dmitry Lepikhin, M. Johnson, M. Krikun, M. Chen, Yuan Cao, G. Foster, Colin Cherry, Wolfgang Macherey, Z. Chen, and Y. Wu. 2019. Massively multilingual neural machine translation in the wild: Findings and challenges
18. Mhaskar, S., Bhat, V., Batheja, A., Deoghare, S., Choudhary, P., & Bhattacharyya, P.

(2023). VAKTA-SETU: A Speech-to-Speech Machine Translation Service in Select Indic Languages.

Online Sources

1. OpenAI, "Whisper: Speech Recognition System," [Online]. Available: https://openai.com/research/whisper.
2. Hugging Face, "MarianMT: Pretrained Multilingual Translation Models," [Online]. Available: https://huggingface.co/models.
3. Google AI Blog, "TTS Research and Development," [Online]. Available: https://ai.googleblog.com/.
4. Mozilla TTS, "Open-Source Text to Speech," [Online]. Available: https://github.com/mozilla/TTS.

Theses and Dissertations

5. R. Prakash, "Speech-to-Speech Translation for Indian Languages," M.S. thesis, Dept. of CSE, IIIT Hyderabad, India, 2021.
6. A. Iyer, "Improving Neural Text-to-Speech Synthesis for Low-Resource Indian Languages," Ph.D. dissertation, Dept. of AI, IISc Bangalore, India, 2022.

Reports

7. UNESCO, Inclusive Digital Learning: Challenges and Opportunities in Multilingual Environments, Paris, France: UNESCO, 2021.
8. NITI Aayog, National Strategy for Artificial Intelligence – #AIforAll, New Delhi, India: Government of India, 2020.

Software and Tools

9. Python Software Foundation, "Python 3.10 Documentation," [Online]. Available: https://docs.python.org/3/.

10. FFmpeg Developers, “FFmpeg Video Processing Tool,” [Online]. Available: https://ffmpeg.org/.

Project-Specific Sources

22. GitHub Repository, "Multilingual Dubbing System" [Online]. Available: https://github.com/CBD-G01/Multilingual-Dubbing-System. [Accessed: Apr. 13,2025].
23. M. Swapna, et al., "Developing a software for dubbing of videos from English to other Indian regional languages" unpublished internal report, Presidency University, Bangalore, India, 2024.

APPENDIX-A

PSUEDOCODE

```
import streamlit as st
import ffmpeg
import speech_recognition as sr
from googletrans import Translator
from gtts import gTTS
from docx import Document
import yt_dlp
import os

def extract_audio(video_path):
    audio_path = "extracted_audio.wav"
    (
        ffmpeg
        .input(video_path)
        .output(audio_path, acodec="pcm_s16le")
        .run(overwrite_output=True)
    )
    return audio_path
def transcribe_audio(audio_path):
    recognizer = sr.Recognizer()
    with sr.AudioFile(audio_path) as source:
        audio = recognizer.record(source)
    text = recognizer.recognize_google(audio)
    return text
def translate_text(text, target_lang, translator, chunk_size=200):
    sentences = text.split(". ")  # Split into sentences
    translated_sentences = []
    for i in range(0, len(sentences), chunk_size):
        chunk = ". ".join(sentences[i:i+chunk_size])
        translated_chunk = translator.translate(chunk, dest=target_lang).text
        translated_sentences.append(translated_chunk)
    return " ".join(translated_sentences)
def generate_voiceover(translated_text, language):
    tts = gTTS(text=translated_text, lang=language)
    voiceover_path = f"voiceover_{language}.mp3"
    tts.save(voiceover_path)
    return voiceover_path
def create_srt(translated_text, language, video_duration):
    srt_path = f"subtitles_{language}.srt"
```

```
    with open(srt_path, "w", encoding="utf-8") as f:
        start_time = "00:00:00,000"
        end_time = f"00:{int(video_duration // 60):02}:{int(video_duration %
60):02},000"  # Full duration
        f.write(f"1\n{start_time} --> {end_time}\n{translated_text.strip()}\n\n")
    return srt_path
def combine_audio_with_video(video_path, voiceover_path, subtitle_path, lang):
    output_path = f"final_video_{lang}.mp4"
    # Load video and audio
    video = ffmpeg.input(video_path)
    audio = ffmpeg.input(voiceover_path)
    # Add subtitles to the video
    video_with_subs = video.filter("subtitles", subtitle_path)
    # Merge video (with subs) and new audio, keeping video from input 0 and audio from
input 1
    (
        ffmpeg
        .output(video_with_subs, audio, output_path, vcodec="libx264", acodec="aac",
shortest=None, **{"map": "0:v", "map": "1:a"})
        .run(overwrite_output=True)
    )
    return output_path
def download_youtube_video(url):
    output_file = "downloaded_video.mp4"
    # Check if file exists, delete if it does
    if os.path.exists(output_file):
        os.remove(output_file)
    ydl_opts = {
        'format': 'bestvideo+bestaudio/best',
        'outtmpl': output_file,  # Save as this filename
        'merge_output_format': 'mp4',
        'noplaylist': True
    }
    with yt_dlp.YoutubeDL(ydl_opts) as ydl:
        ydl.download([url])
    return output_file
def create_translation_document(original_text, translations, languages):
    doc = def get_video_duration(video_path):
    probe = ffmpeg.probe(video_path)
    return float(probe['format']['duration'])  # Duration in seconds
def main(video_path, target_languages):
    translator = Translator()
    audio_path = extract_audio(video_path)
    transcribed_text = transcribe_audio(audio_path)
```

```
    video_duration = get_video_duration(video_path)
    translations = []
    for lang_code in target_languages:
        translated_text = translate_text(transcribed_text, lang_code, translator)
        translations.append(translated_text)
        voiceover_path = generate_voiceover(translated_text, lang_code)
        subtitle_path = create_srt(translated_text, lang_code, video_duration)
        combine_audio_with_video(video_path, voiceover_path, subtitle_path, lang_code)
    doc_path = create_translation_document(transcribed_text, translations,
target_languages)
    return translations, doc_path
# Streamlit UI
st.title("# Upload video OR enter a YouTube link
video_file = st.file_uploader("Upload a video file", type=["mp4", "mov"])
video_url = st.text_input("Or enter a YouTube URL:")
target_languages = st.multiselect("Select target languages", ["hi", "te", "ta", "kn",
"mr", "ur", "ml", "pa", "gu"])
if st.button("Translate & Subtitle"):
    if video_file:
        video_path = "uploaded_video.mp4"
        with open(video_path, "wb") as f:
            f.write(video_file.getbuffer())
    elif video_url:
        st.info("Downloading YouTube video...")
        video_path = download_youtube_video(video_url)
        st.success("YouTube video downloaded!")
    else:
        st.warning("Please upload a video or enter a YouTube URL.")
        st.stop()
    if target_languages:
        translations, doc_path = main(video_path, target_languages)
         st.success("Processing Completed!")

         for lang in target_languages:
             video_file_to_download = f"final_video_{lang}.mp4"
             if os.path.exists(video_file_to_download):
                 with open(video_file_to_download, "rb") as video_file:
                     st.download_button(label=f"Download {lang} Video",
data=video_file, file_name=f"{lang}_translated_video.mp4")
             else:
                 st.warning(f"No video file found for language: {lang}")
         with open(doc_path, "rb") as doc_file:
```

```
            st.download_button(label="Download Translation Document",
data=doc_file, file_name="translations.docx")
    else:
        st.warning("Please select at least one language.")
```

APPENDIX-B
SCREENSHOTS

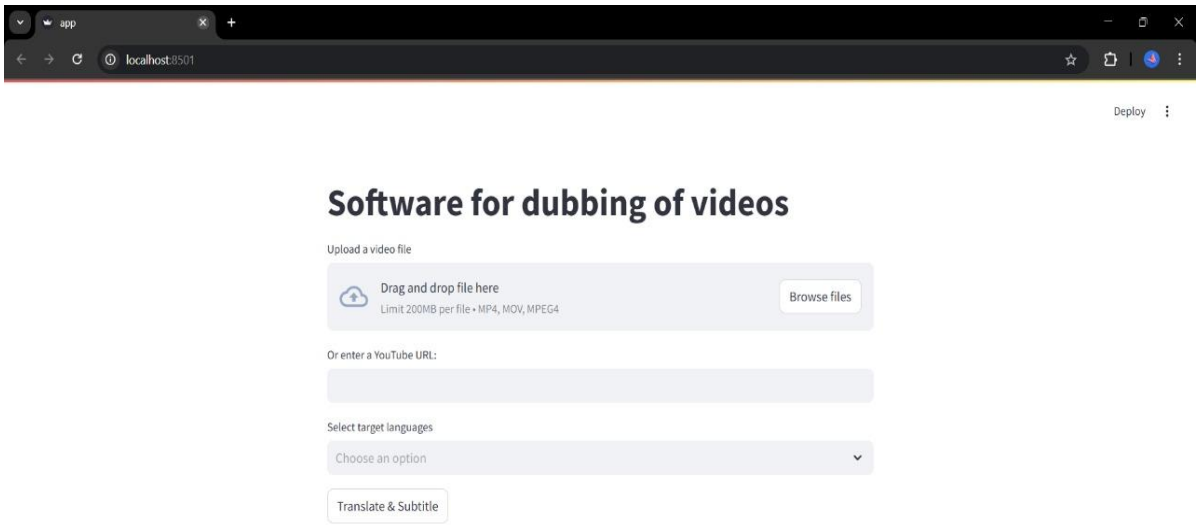


Fig 11: UI page

The UI presents a "Software for dubbing of videos" interface. Users can drag and drop or browse a video file (max 200MB, MP4, MOV, MP3/4) or enter a YouTube URL. Users choose target Indian languages through a dropdown before they click "Translate & Subtitle."

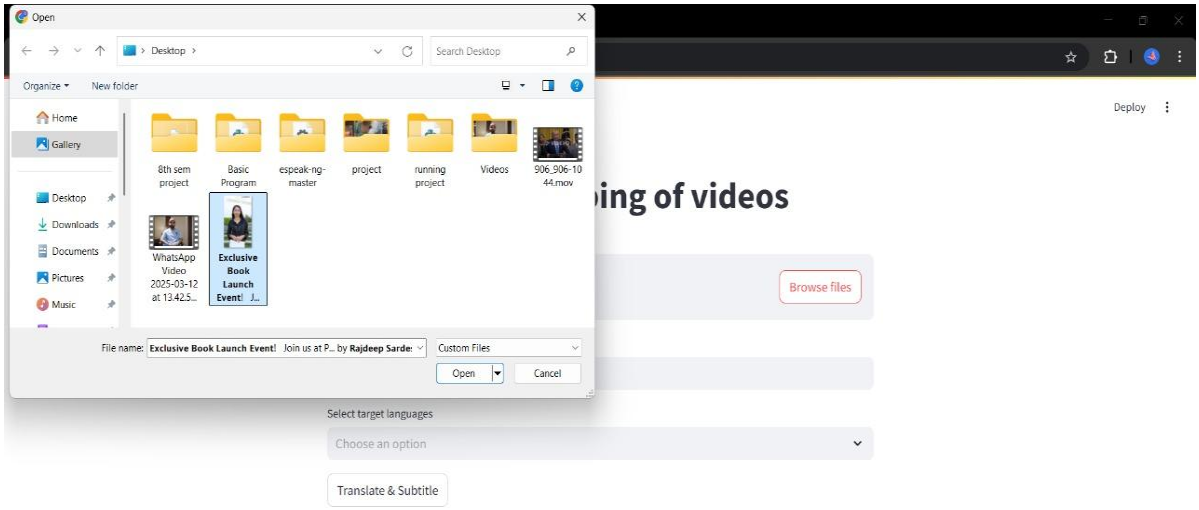


Fig 12: Uploading File

The interface has been designed to allow users to upload video files (200MB maximum in MP4, MOV, MPEG4 formats) or input a YouTube URL. It also provides selection of target Indian languages through clickable tags (hi, gu, pa, mr, etc.). An announcement for the event is given, and a "Translate & Subtitle" button starts the process of dubbing. The software intends to make video dubbing into Indian languages easier.

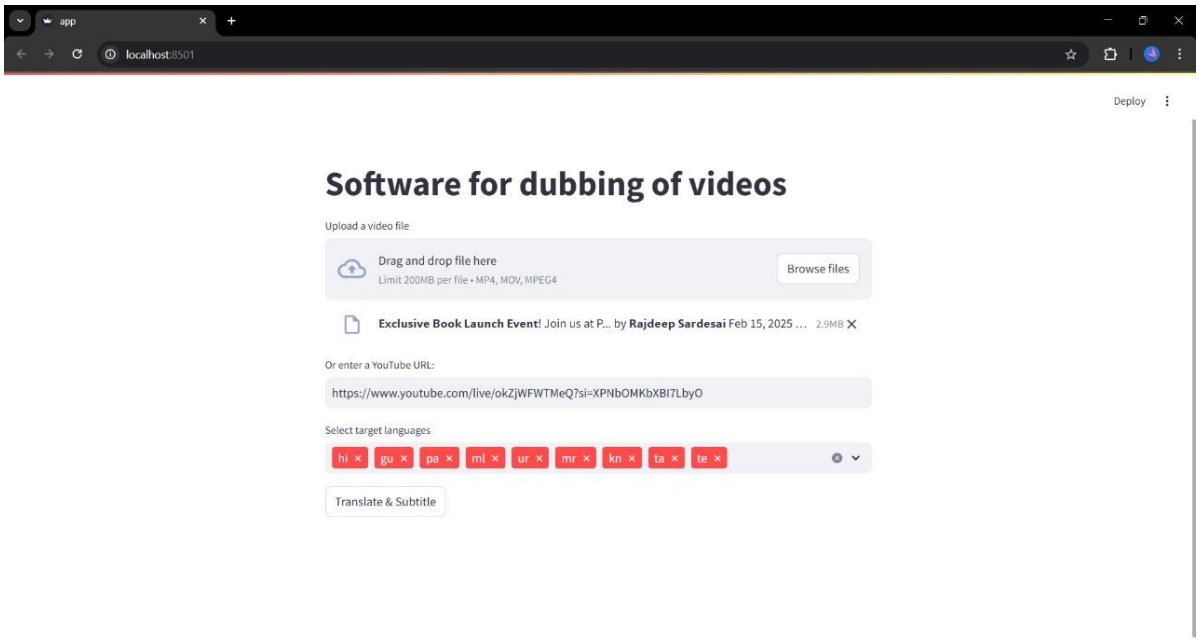


Fig 13: Language Selection

The screenshot shows a UI for video dubbing software. Users are able to upload a video file

(with certain size and format restrictions) or paste a YouTube URL, as demonstrated with a particular link pasted. Various target Indian languages (Hindi, Gujarati, Punjabi, etc.) are chosen through clickable tags. A notice reading "Exclusive Book Launch Event" is also apparent, indicating possible integration or advertisement within the platform. Lastly, a "Translate & Subtitle" button launches the dubbing process.

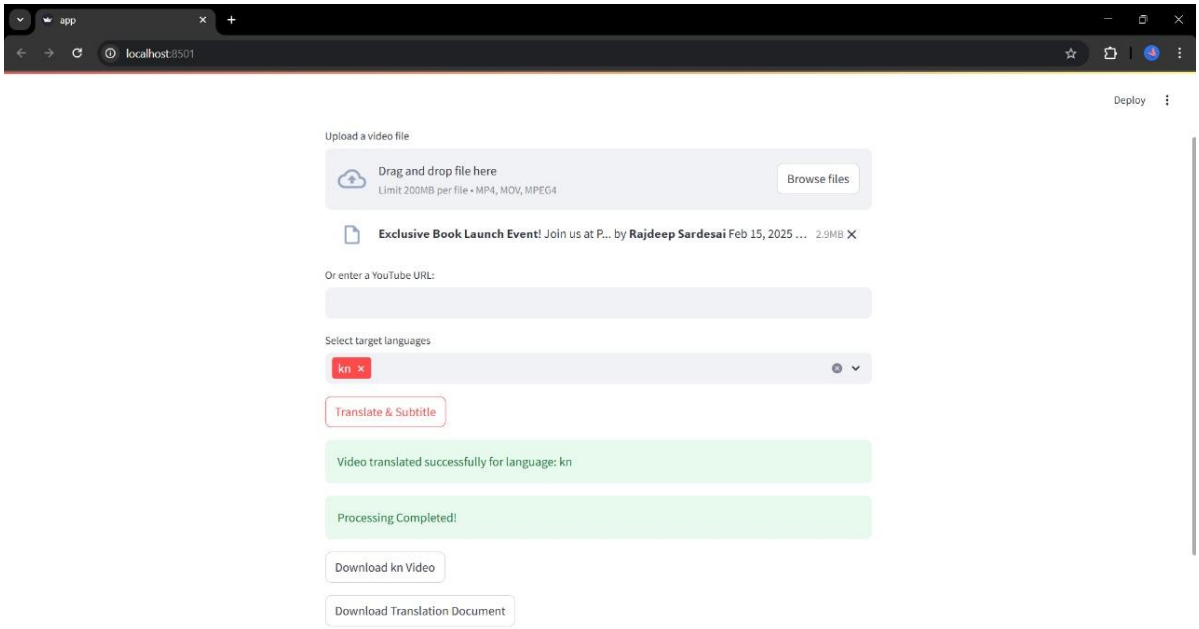


Fig 14: Dubbing Completed

This image is the screen of a video dubbing software after processing. The user first could upload a video file or insert a YouTube URL. A notice for an "Exclusive Book Launch Event" was displayed. The user chose "kn" (Kannada) as the target language and clicked "Translate & Subtitle". The interface now shows two green success messages: "Video translated successfully for language: kn" and "Processing Completed!" Below these, there are two download options: "Download Translated Video" and "Download Translation Document." This shows the software has successfully dubbed the input video into Kannada, and the user can now download the translated video and possibly a separate translation document.

Translations

Original Text:

the team we have this time around is amazing when the auction started I think all of us were a bit you know nervous as the auction brings that energy on to the table and there was a lot of uncertainty on day one what can happen who is going where and all that kind of stuff but I think the way are management and everyone maintain their composure to get exactly what we wanted was amazing to see in the end we created A beautifully balance squad with a lot of match winners a lot of leaders who are you know amazing players in their own right it will be a great help to someone like Rajat as well to have such experience such matters around him for him to just focus on what needs to be done and I'm sure all the players will show their Full support to Rajat and also allow him to grow into this into this role of responsibility that he has been given I'm looking for to play with all of them and as I said create a new energy and create some amazing

Translation in kn:

ಈ ಸಮಯದಲ್ಲಿನಾವು ಹೊಂದಿರುವ ತಂಡವು ಹರಾಜು ಪಾಥಂಭವಾದಾಗ ಆಶ್ಚರ್ಯಕರವಾಗಿದೆ, ಹರಾಜು ಆ ಶಕ್ತಿಯನ್ನುಟೇಬಲ್ಗೆತರುವಾಗ ಮತ್ತುದಿನಕ್ಕೆಸಾಕಷ್ಟು ಅನಿಶ್ಚಿತತೆ ಇರುವುದರಿಂದ ಎಲಿಭೆ ಹೋಗಬಹುದು ಮತ್ತುಆ ರೀತಿಯ ಸಂಗತಿಗಳು ಎಲಿಭೆ ಹೋಗಬಹುದು ಮತ್ತುಆ ರೀತಿಯ ಸಂಗತಿಗಳು ಇದಷ್ಟು ಆದರೆ ದಾರಿ ನಿರ್ವಹಣೆ ಎಂದು ನಾನು ಭಾವಿಸುತ್ತೇನೆ ಮತ್ತುಪ್ರತಿಯೊಬ್ಬರೂ ಅವರ ಸಮಗ್ರತೆಯನ್ನುನಾವು ಬಯಸುತ್ತೇವೆ, ನಾವು ಅದುಭವಾದವುಗಳನ್ನುಹೊಂದಿದ್ದೇವೆ ಎಂದು ನಾವು ಭಾವಿಸಿದ್ದೇವೆ.ತಮಭೇ ಆದ ಹಕ್ಕುರಾಜಕಾರನಂತಹ ಯಾರಿಗಾದರೂ ಒಂದು ದೊಡ್ಡಸಹಾಯವಾಗಲಿದೆ ಮತ್ತು ಅವರ ಸುತಲ್ಲೂ ಅಂತಹ ಅನುಭವವನ್ನುಹೊಂದಲು ಅವರ ಸುತಲ್ಲೂ ಏನು ಮಾಡಬೇಕೆಂಬುದರ ಮೇಲೆ ಕೇಂದ್ರೀಕರಿಸುವುದು ಮತ್ತುಎಲಾ್ಲಆಟಗಾರರು ರಾಜತ್ಗೆತಮ್ಮ ಸಂಪೂರ್ಣ ಬೆಂಬಲವನ್ನುತೋರಿಸುತಾರ್ರೆ ಎಂದು ನನಗೆ ಖಾತಿಭಿದೆ ಮತ್ತುಅವರು ಈ ಜವಾಬಾರ್ಡಿಯ ಪಾತಭಲಿ್ಲಬೆಳೆಯಲು ಸಹ ಅವಕಾಶ ಮಾಡಿಕೊಡುತಾರ್ರೆ, ನಾನು ಅವರೆಲರ್ಲೊಂದಿಗೆ ಆಟವಾಡಲು ಹುಡುಕುತಿದ್ದೇನೆ ಮತ್ತುಹೊಸ ಶಕ್ತಿಯನ್ನುರಚಿಸಿ ಮತ್ತು ಹೊಸ ಶಕ್ತಿಯನ್ನುರಚಿಸಿ ಮತ್ತುಕೆಲವು ಅದುಭವನ್ನುರಚಿಸಿ

Fig 15: Transcript

This screenshot displays the transcript of another dubbed video. It is the dubbed transcript from English to Kannada of an individul's insight on RCB's 2025 auction and naming Rajat Patidar as the franchise's Captain.

APPENDIX-C

ENCLOSURES

Research Paper Plagiarism

ORIGINALITY REPORT

8%	6%	6%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.apptek.com Internet Source	1%
2	sih.gov.in Internet Source	1%
3	K. Hema Priya, N. Akhilan, R. Aravindh, K. Janardhana. "Chapter 4 An In-Depth Investigation into Automatic Dubbing Leveraging ASR, Machine Translation and Deep Voice 3", Springer Science and Business Media LLC, 2024 Publication	1%
4	www.ijitee.org Internet Source	1%
5	aclanthology.org Internet Source	<1%
6	arxiv.org Internet Source	<1%
7	kaniyam.com Internet Source	<1%
8	Submitted to INTI Universal Holdings SDM BHD Student Paper	<1%
9	"Computational Intelligence for Machine Learning and Healthcare Informatics", Walter de Gruyter GmbH, 2020 Publication	<1%
10	Submitted to Victoria University of Wellington Student Paper	<1%

11	Submitted to George Mason University Student Paper	<1 %
12	www.ijcaonline.org Internet Source	<1 %
13	docplayer.net Internet Source	<1 %
14	Submitted to Visvesvaraya Technological University, Belagavi Student Paper	<1 %
15	"Advanced Computing", Springer Science and Business Media LLC, 2021 Publication	<1 %
16	"Application of Intelligent Systems in Multi-modal Information Analytics", Springer Science and Business Media LLC, 2022 Publication	<1 %
17	C. Rahul, T. Arathi, Lakshmi S. Panicker, R. Gopikakumari. "Morphology & word sense disambiguation embedded multimodal neural machine translation system between Sanskrit and Malayalam", Biomedical Signal Processing and Control, 2023 Publication	<1 %
18	Gunti Spandan, S H Brahmananda. "Impact of Machine Learning on Regional Languages Processing: A Survey", 2023 2nd International Conference for Innovation in Technology (INOCON), 2023 Publication	<1 %
19	Moradshahi, Mehrad. "Internationalization of Task-Oriented Dialogue Systems", Stanford University, 2023 Publication	<1 %
20	Zheng Zeng. "Implementation of Embedded Technology-Based English Speech	<1 %

Identification and Translation System", Computer Systems Science and Engineering, 2020
Publication

21	www.preprints.org Internet Source	<1 %
22	Aarati H. Patil, Snehal S. Patil, Shubham M. Patil, Tatwadarshi P. Nagarhalli. "Real Time Machine Translation System between Indian Languages", 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022 Publication	<1 %
23	Raj Dabre, Chenhui Chu, Anoop Kunchukuttan. "A Survey of Multilingual Neural Machine Translation", ACM Computing Surveys, 2020 Publication	<1 %
24	Sakshi Singh, Thoudam Doren Singh, Sivaji Bandyopadhyay. "Chapter 48 An Experiment onSpeech-to-Speech Translation ofHindi to English: A Deep Learning Approach", Springer Science and Business Media LLC, 2022 Publication	<1 %
25	Syed Matla Ul Qumar, Muzaffar Azim, S. M. K. Quadri. "Emerging resources, enduring challenges: a comprehensive study of Kashmiri parallel corpus", AI & SOCIETY, 2024 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

Project Report Plagiarism

Final Report

ORIGINALITY REPORT

14%	10%	8%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Moratuwa Student Paper	2%
2	Submitted to Presidency University Student Paper	2%
3	Submitted to Symbiosis International University Student Paper	1%
4	Submitted to M S Ramaiah University of Applied Sciences Student Paper	<1%
5	sih.gov.in Internet Source	<1%
6	Submitted to University of Westminster Student Paper	<1%
7	www.apptek.com Internet Source	<1%
8	www.canada.ca Internet Source	<1%
9	Submitted to Accra Business School Student Paper	<1%
	docslib.org	

10	Internet Source	<1%
11	"Innovations and Advances in Cognitive Systems", Springer Science and Business Media LLC, 2024 Publication	<1%
12	huggingface.co Internet Source	<1%
13	aclanthology.org Internet Source	<1%
14	Elena Davitti, Tomasz Korybski, Sabine Braun. "The Routledge Handbook of Interpreting, Technology and AI", Routledge, 2025 Publication	<1%
15	kaniyam.com Internet Source	<1%
16	Submitted to APJ Abdul Kalam Technological University, Thiruvananthapuram Student Paper	<1%
17	Submitted to King's College Student Paper	<1%
18	Submitted to University of Bedfordshire Student Paper	<1%
19	Submitted to University of Leeds Student Paper	<1%
20	mllp.upv.es Internet Source	<1%

21	Submitted to University of East London Student Paper	<1 %
22	arxiv.org Internet Source	<1 %
23	hltc.cs.ust.hk Internet Source	<1 %
24	news.djaz.app Internet Source	<1 %
25	telnyx.com Internet Source	<1 %
26	"Front Matter", 2023 8th International Conference on Computer Science and Engineering (UBMK), 2023 Publication	<1 %
27	Submitted to University for Development Studies Student Paper	<1 %
28	Submitted to University of Edinburgh Student Paper	<1 %
29	Submitted to University of Nottingham Student Paper	<1 %
30	www.feri.de Internet Source	<1 %
31	Submitted to Wesleyan University Student Paper	<1 %
32	socialresearchfoundation.com Internet Source	<1 %

Acceptance Certificate

ISSN: 2582-3930

ACCEPTANCE CERTIFICATE

Impact Factor: 8.586
DOI Prefix: 10.55041

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT (IJSREM)

An Open Access Scholarly Journal || Index in major Databases & Metadata

* * *

We are pleased to inform you that your manuscript titled

Developing a software for dubbing of videos from English to other Indian regional languages

has been ACCEPTED for publication in

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT (IJSREM)

VOLUME 09 ISSUE 05 MAY - 2025

We are delighted to see your commitment & hardwork to share your research is being recognized. We look foreward to helping you with all of your publication needs. Thank you for choosing IJSREM!!

Crossref CiteFactor

Editor-in-chief
IJSREM

www.ijsrem.com e-mail: editor@ijsrem.com

APPENDIX – D

Mapping the Elective Dubbing System with Sustainable Development Goals (SDGs)

This project contributes to the United Nations Sustainable Development Goals (SDGs) by expanding education and information accessibility through multilingual dubbing. Through eliminating language barriers, the system advances inclusive and equitable quality education, advances digital empowerment, and supports reduction in communication and knowledge-sharing inequalities in different communities.

1. SDG 4: Quality Education

Goal: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.

Project Contribution:

The elective dubbing system makes a major contribution to Sustainable Development Goal 4: Quality Education, as it makes digital learning content available in several regional languages. This provides inclusivity to ensure that students with different linguistic backgrounds can effectively understand and relate to learning material. By eliminating language as a hindrance, the system enhances equal opportunities for education, improves learning outcomes, and facilitates lifelong learning. It empowers marginalized communities and rural learners by filling the gap between high-quality digital content and native-language understanding.

Outcome:

The project enabled effective language localization of educational videos, improved access for non-English speakers, and increased learner engagement, thus fostering inclusivity and supporting the goal of equitable digital education.

2. SDG 5: Gender Equality

Goal: Achieve gender equality and empower all women and girls.

Project Contribution:

The dubbing system works towards Sustainable Development Goal 5: Gender Equality by

ensuring equal access to educational content irrespective of gender. In most parts of the world, women and girls are constrained by language barriers in accessing digital education. Through providing content in local languages, the system ensures female learners have equitable access, thereby promoting their educational development and digital literacy. It empowers women by educating them, bridging the gender divide in education, and advancing their social and economic development.

Outcome:

The system increased educational access for women, especially in rural areas, helping reduce gender disparities in digital learning participation.

3. SDG 8: Decent Work and Economic Growth

Goal: Promote sustained, inclusive, and sustainable economic growth, full and productive employment, and decent work for all.

Project Contribution:

The dubbing system facilitates Sustainable Development Goal 8: Decent Work and Economic Growth through the provision of skill acquisition through affordable educational materials in local languages. Through making technical and vocational training content accessible to a broader population, it improves employability and productivity, especially among disadvantaged communities. The system facilitates inclusive digital learning, enabling people to acquire relevant skills that match the needs of contemporary workforces. In the long run, this helps in developing improved employment opportunities and economic engagement for everyone.

Outcome:

The project improved access to skill-building resources, empowered regional learners, and increased their employment potential, thus contributing to local economies and promoting sustainable economic growth through education.

4. SDG 9: Industry, Innovation, and Infrastructure

Goal: Build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation.

Project Contribution:

The dubbing system helps achieve Sustainable Development Goal 9: Industry, Innovation, and Infrastructure by utilizing cutting-edge technologies such as AI-based speech recognition and machine translation to develop inclusive learning tools. It promotes digital innovation by developing scalable infrastructure for multilingual content distribution. The system aids in the creation of smart learning solutions, filling technological gaps in rural and disadvantaged regions. Through encouraging access to emerging platforms, it supports digital ecosystems and provides a base for sustainable industrial and educational development.

Outcome:

The initiative created technological innovation, improved digital education infrastructure, and offered scalable solutions for inclusive delivery of content, thus catering to industrial development as well as increased access to contemporary learning tools.

5. SDG 10: Reduced Inequalities

Goal: Reduce inequality within and among countries.

Project Contribution:

Reduced Inequalities through the assurance that language will no longer act as a hindrance to the consumption of quality education content. By providing multilingual support, the system empowers learners from diverse linguistic and socio-economic backgrounds, such as marginalized and rural communities. It closes the digital divide by providing information to everyone, irrespective of language skills, thus creating equal learning opportunities and ensuring inclusiveness in education, skill acquisition, and digital engagement among various population groups.

Outcome:

The project enhanced inclusivity by delivering educational content in regional languages, enabling marginalized groups to access knowledge, thus reducing disparities in education and supporting social and digital equality.

6. SDG 12: Responsible Consumption and Production

Goal: Ensure sustainable consumption and production patterns.

Project Contribution:

Sustainable Consumption and Production by facilitating effective reuse of available

educational content through localization of languages. Rather than duplicating material for various audiences, the platform adjusts and reuses high-quality material in various forms across numerous languages through the application of artificial intelligence-based tools. It reduces wastage of resources, saves costs on production, and increases the value of digital assets. It promotes eco-friendly habits in the delivery of educational content and cultivates responsible technological innovation and content sharing in the educational system.

Outcome:

The project enabled efficient content reuse, minimized duplication of educational materials, and promoted sustainable practices in digital education, ensuring wider access with lower resource expenditure and environmental impact.

7. SDG 17: Partnerships for the Goals

Goal: Strengthen the means of implementation and revitalize the global partnership for sustainable development.

Project Contribution:

Partnership for the Goals by enabling interdisciplinary partnerships between educational institutions, tech developers, and linguistic experts. It stresses the adoption of local knowledge with global digital resources to produce inclusive learning materials. With cross-sectoral partnerships, the project illustrates how joint expertise and collaborative innovation can increase accessibility. This model of collaboration enhances international and local partnerships by promoting sustainable development through collective efforts and mutual assistance in education and technology.

Outcome:

The project promoted interdisciplinary collaboration, encouraged knowledge-sharing among stakeholders, and demonstrated the effectiveness of partnerships in achieving inclusive education goals and scalable digital development solutions.

Fig 16: SDG Goals