Advancements in AI: A Study of English-to-Urdu Dubbing Methods and Applications

Mariam Arshad
Teesside University
Middlesbrough, UK
d3676772@tees.ac.uk

Shatha Ghareeb
Teesside University
Middlesbrough, UK
S.ghareeb@tees.ac.uk

Jamila Mustafina
Kazan Federal University
Kazan, Russia
DNMustafina@kpfu.ru

2024 17th International Conference on Development in eSystem Engineering (DeSE) | 979-8-3503-6869-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/DeSE63988.2024.10911974

Abstract— The global film industry faces significant challenges in making content accessible across diverse linguistic and cultural contexts, particularly through dubbing, which involves translating and synchronizing audio for different languages. Traditional dubbing methods, while effective, are labour-intensive, costly, and often struggle with preserving cultural nuances and emotional depth. With advancements in artificial intelligence (AI), there is a promising opportunity to address these challenges and revolutionize the dubbing process. This research explores the application of AI-driven technologies to improve the efficiency and accuracy of dubbing English-language films into Urdu. The research investigates the current state of AI-driven dubbing technologies, identifies their limitations, and proposes methods to enhance their effectiveness. Key objectives include analysing AI technologies for translation and speech synthesis, developing a streamlined process for English-to-Urdu dubbing, and evaluating the cultural and emotional fidelity of AI-generated content. A comprehensive methodology involving data extraction, speech-to-text conversion, translation, text-to-speech synthesis, and audio-video integration is employed to create an automated dubbing system. The system's performance is assessed based on accuracy, naturalness, and synchronization, revealing both successes and areas for improvement. The findings demonstrate that AI-driven dubbing can significantly enhance the efficiency and scalability of the dubbing process, although challenges related to accent variability, cultural sensitivity, and voice naturalness remain. The study concludes with recommendations for future research, including the need for improved transcription accuracy, advanced translation models, and real-time dubbing capabilities. This research contributes valuable insights into the potential of AI to transform film localization and other sectors by making content more accessible and culturally relevant to diverse audiences.

Keywords – AI for real-world applicaions; dubbing;cultural context.

I. INTRODUCTION

The global film industry is a vital medium for cultural exchange and storytelling, yet international movie distribution is often impeded by language barriers. Traditional methods such as dubbing and subtitling are commonly used to make films accessible to diverse audiences. Dubbing provides a way for viewers to experience films in their native language, thereby preserving the emotional and narrative integrity of the original work. However, the conventional dubbing process is complex, costly, and time-consuming, and frequently struggles to capture cultural nuances and emotional subtleties inherent in the source material [1].

Recent advancements in artificial intelligence (AI) have the potential to transform multilingual movie dubbing by enhancing efficiency and accuracy while addressing some of the challenges associated with traditional methods. AI-driven technologies promise to streamline the dubbing process, reduce costs, and improve the preservation of cultural and emotional aspects of the original content [2]. This is particularly relevant for language pairs like English and Urdu, where the effectiveness of localized content relies heavily on capturing cultural nuances. Despite these advancements, the implementation of AI-driven dubbing solutions remains underexplored, especially for less commonly supported languages such as Urdu. There is a critical need to identify best practices for effectively leveraging AI to overcome the limitations of traditional dubbing and enhance the quality of multilingual content.

This study aims to investigate best practices for implementing AI-driven dubbing from English to Urdu. The primary objectives include analysing the current state of AI-driven dubbing technologies, developing methods to improve efficiency and automation, evaluating the preservation of cultural nuances, comparing cost and time efficiency with traditional approaches, and assessing the linguistic accuracy and quality of AI-generated content. By addressing these objectives, the research seeks to advance the application of AI technologies in the dubbing industry, particularly for English-to-Urdu content, and provide insights into enhancing the localization process across various sectors.

II. LITERATURE REVIEW

A. Traditional Dubbing Methods and Their Limitations

Dubbing has historically been the preferred method for making films accessible to non-native speakers. Human dubbing involves several stages, including translation, adaptation, and voice recording. This process must balance various constraints, such as isochrony (timing), isometry (character length), and lip-sync accuracy, while also maintaining naturalness and translation quality [3]. Studies reveal that human dubbers often prioritize naturalness and translation fidelity over strict adherence to timing or lip-sync constraints, leading to deviations from perfect synchronization [3].

These limitations have sparked a growing interest in exploring alternative methods for multilingual content localization. For example, [4] discusses the rise of subtitling as a more cost-effective way, though it is not without its drawbacks, such as distracting viewers and limiting accessibility for those with visual impairments or low literacy levels. The need for more innovative solutions has led to the exploration of AI-driven approaches. The table below highlights key differences between traditional and AI-driven dubbing methods, illustrating the advantages of AI in addressing the challenges faced by traditional methods.

Authorized licensed use limited to: UNIVERSIDADE FEDERAL DO PARANA. Downloaded on November 05,2025 at 19:34:52 UTC from IEEE Xplore. Restrictions apply.

Table 1: Comparative Advantages of AI-Driven Dubbing Over Traditional Dubbing

Feature	Traditional Dubbing	AI Dubbing (Automatic Dubbing)
Speed	Time-consuming; requires multiple takes, voice actor coordination, and post-production editing.	Much faster; automates the process, significantly reducing turnaround time.
Cost	Expensive; involves hiring voice actors, studio time, and editing professionals.	More cost-effective; reduces the need for human talent and studio time.
Consistency	Variability due to human factors; maintaining consistent tone and quality can be challenging.	High consistency across outputs, ensuring uniformity in large-scale projects.
Scalability	Limited by the availability of voice actors and studio resources.	Easily scalable; can handle large volumes of content quickly and efficiently.
Language Support	Limited by the availability and expertise of voice actors in different languages.	Supports a wide range of languages, often beyond what is available with human voice actors.
Customization	Requires re-recording or additional editing for changes, which can be time-consuming and costly.	Allows rapid adjustments and parameter changes without needing additional takes or actors.
Availability	Dependent on the availability of voice actors and studio time, which can cause delays.	Operates without time constraints, available 24/7, unlike human actors.
Integration	Less integrated into digital workflows; often requires manual coordination and editing.	Easily integrated into digital workflows, enabling streamlined production processes.
Adaptability	Limited by human talent; updating or retraining is time-consuming and costly.	AI models can be quickly updated or retrained to improve performance or add new languages.
Human Error	Prone to human errors like mispronunciations, inconsistent delivery, and timing issues.	Minimizes errors associated with manual processes, ensuring more consistent quality.

B. AI-Driven Dubbing: Current State and Applications

Artificial intelligence has emerged as a transformative force across various industries, including media and entertainment. AI-driven dubbing, which utilizes machine learning algorithms and neural networks to automate the dubbing process, is a relatively recent development that promises to address many of the challenges associated with traditional methods. [1] and [5] discuss the use of neural machine translation (NMT) and self-learning approaches to generate translations that closely match the source text's length, crucial for synchronization. The Isometric MT model, as introduced by [5], directly learns to produce outputs with a length that matches the source, achieving better results than previous methods requiring multiple hypotheses and re-ranking.

Voice cloning technologies and prosodic alignment are critical for enhancing the naturalness of AI-generated dubbing. [1] and [6] explore how prosodic alignment and synchronization techniques improve the quality of machine-generated dubs. [6] propose a method for prosodic phrase synchronization, leveraging neural machine translation outputs to align translated dialogue with the source's prosody, achieving improvements in lip-syncing and speech rate.

The AI dubbing pipeline typically involves translation, speech synthesis, and synchronization. [5] emphasizes optimizing for isochrony (timing) rather than isometry (length) to enhance synchronization. [7] highlight how integrating machine learning techniques into video translation processes can reduce manual work and production time, although challenges remain in achieving perfect synchronization and naturalness. To provide a clearer picture of the advancements and limitations in AI-driven dubbing, the following table summarizes key studies in this field, highlighting their contributions and limitations. This overview helps illustrate the current state of research and the ongoing efforts to address various challenges in AI-driven dubbing technologies.

Table 2: Overview of Key Studies in AI-Driven Dubbing Technologies

Title of Paper	Methods or Techniques	Dataset	Language Focused	Key Contributions	Limitations
From Speech-to-Speech Translation to Automatic Dubbing	Neural Machine Translation (NMT), Neural Text-to-Speech (TTS), Audio Rendering	TED Talk excerpts	English to Italian	Enhances prosodic alignment and audio rendering for natural dubbing.	Variability in synchronization and naturalness; Italian listeners reported some unnaturalness.
ANGLABHARTI: A Multilingual Machine Aided Translation Project on Translation from English to Indian Languages	Pattern Directed Approach, Context-Free Grammar, Rule-Based System	Parallel corpora	English to Hindi, other Indian languages	Uses pattern-directed approach and human post-editing for multiple Indian languages.	Limited to certain Indian languages; relies on rules and requires significant human post-editing.
ISOMETRIC MT: NEURAL MACHINE TRANSLATION FOR AUTOMATIC DUBBING	Neural Machine Translation (NMT), self-learning approach, transformer model	TED Talks	English to French, Italian, German, Spanish	Generates translations with matched source length, improving synchronization and quality.	Needs optimization for diverse languages and longer texts; primarily tested on TED Talk excerpts.
Machine Translation from English to Malayalam Using Transfer Approach	Rule-Based Machine Translation (RBMT), Transfer Approach	Cricket domain-related articles from Sportstar magazine	English to Malayalam	Achieves high accuracy in cricket domain with domain-specific RBMT system.	Limited to cricket domain; accuracy issues with longer sentences; needs further refinement.
Prosodic Phrase Alignment for Machine Dubbing	Neural Machine Translation (NMT), attention mechanism, TTS synchronization	Heroes Corpus	English to Spanish	Synchronizes prosodic phrases and improves lip-syncing for long dialogue lines.	May struggle with complex sentence structures; limited to specific datasets.

Authorized licensed use limited to: UNIVERSIDADE FEDERAL DO PARANA. Downloaded on November 05,2025 at 19:34:52 UTC from IEEE Xplore. Restrictions apply.

Dubbing in Practice: A Large-Scale Study of Human Localization With Insights for Automatic Dubbing	Human dubbing analysis, empirical study	Human-dubbed content	English to Spanish, English to German	Provides insights into human dubbing practices and challenges assumptions in automatic dubbing.	Limited to human dubbing; insights may not apply to AI dubbing without further research.
Transformative Perspectives: AI-Enabled Dubbing Software for Multilingual Content Localization	NLP, ASR, Google Text-to-Speech (gTTS), Google Translator, Python Programming	Not specified	English to Indian regional languages	Proposes AI-based real-time translation with voice modulation and lip synchronization.	Challenges in achieving perfect synchronization and naturalness; limited to Indian regional languages.
The Development of Syllable Based Text to Speech System for Tamil Language	Syllabication, Letter-to-Sound mapping, Unit Selection Speech Synthesis	Not specified	Tamil (target language; source language not mentioned)	Develops a syllable-based TTS system for Tamil, enhancing speech synthesis.	Tailored for Tamil; techniques may not transfer easily to other languages.
An Intelligent System for Automated Translation of Videos from English to Native Language Applying Artificial Intelligence Techniques for Adaptive eLearning	Machine learning, Neural translation, Cloud services, MoviePy, FFmpeg	50 videos selected from various languages	English to various native languages	Automates video translation process, improving efficiency and accuracy.	Limited to 50 videos; broader applicability and performance not extensively evaluated.
Bilingual Prosodic Dataset Compilation for Spoken Language Translation	Prosody annotation, Automated bilingual corpus creation, Speech-to-text alignment	Parallel corpus (7000 utterances, ~10 hours of data)	English to Spanish	Creates prosodically annotated bilingual corpora for training spoken language models.	Manual subtitle corrections may still result in mismatches.

C. *Challenges and Ethical Considerations in AI-Driven Dubbing*

Despite its potential, AI-driven dubbing is not without its challenges. One of the primary concerns is the quality of the output, particularly in preserving the cultural and emotional integrity of the original content. [3] find that the transfer of semantic properties such as emphasis and emotion is a significant challenge for automatic dubbing systems, which are crucial for maintaining the naturalness and emotional impact of the original content.

Another significant challenge is the ethical implications of AI-driven dubbing. The use of AI in creative processes has sparked debates about authorship, creativity, and the potential displacement of human workers in the dubbing industry. [3] discuss the potential misuse of voice cloning technologies and its impact on employment in the dubbing industry, stressing the need for responsible AI usage and the protection of intellectual property rights.

D. *Gaps in the Literature*

While substantial research has been conducted on AI-driven dubbing technologies, several gaps remain, particularly in the application to Urdu dubbing. Existing studies often concentrate on widely spoken languages or general techniques, overlooking lesser-studied languages, regional dialects, and the unique linguistic and cultural characteristics of languages like Urdu. This limitation restricts the applicability of AI systems across diverse linguistic landscapes. Furthermore, current research inadequately addresses the challenges of adapting AI-driven dubbing to cultural nuances, idiomatic expressions, and emotional contexts. There is also a lack of standardized metrics for assessing the quality of AI-generated dubbing, and more research is needed on user experience and preferences. Practical integration into production workflows, scalability of technologies, and ethical considerations, such as the implications of voice cloning on copyright and employment, remain underexplored.

Additionally, the effectiveness of AI dubbing in handling complex dialogues and adapting to various genres and interactive content requires further investigation. Addressing these gaps could significantly advance the development and application of AI-driven dubbing technologies.

III. RESEARCH METHODOLOGY

This section outlines the research methodology employed in the development of the video dubbing system designed to automatically dub English videos into Urdu. The primary objective of this project is to enhance accessibility for Urdu-speaking audiences by replacing the original English audio with accurate and natural-sounding Urdu speech. The methodology encompasses various phases including dataset description, environment setup, audio extraction, speech transcription, text translation, speech synthesis, and audio-video merging.

A. *Dataset Description*

The dataset utilized for this project comprises 881 MP4 videos in English, obtained from Kaggle, a renowned platform for datasets and data science competitions (https://www.kaggle.com/datasets/konstantinitsi/youtube-video-trailer/data). This dataset includes a diverse range of entertainment-focused content, such as movie clips and animated material. The videos vary in length from 15 seconds to 2 hours and 30 minutes, and their file sizes range from 150 KB to 65,440 KB. Each video is formatted in MP4, with audio channels typically being stereo, although variations such as mono or 5.1 surround sound may also be present. The dataset's comprehensive nature ensures a rich variety of content for robust video dubbing analysis.

B. *Environment Setup*

To facilitate the project, the following libraries and tools were installed, and Google Cloud credentials were configured using a JSON file:

1. *MoviePy*: A versatile library for video and audio processing. It enables video editing, audio extraction, and audio replacement, which is crucial for integrating new Urdu audio with the original video.
2. *Pydub*: A powerful library for audio manipulation. It allows for converting audio formats and adjusting audio channels, such as converting stereo audio to mono to ensure compatibility with the speech recognition system.

Authorized licensed use limited to: UNIVERSIDADE FEDERAL DO PARANA. Downloaded on November 05,2025 at 19:34:52 UTC from IEEE Xplore. Restrictions apply.

3. *Google Cloud Speech-to-Text*: A cloud-based API for converting spoken language into written text. This service transcribes English audio into text, which is necessary for subsequent translation and synthesis steps.
4. *Google Cloud Translate*: A cloud-based translation service for translating text from English to Urdu. It ensures that the English transcripts are accurately and naturally translated into Urdu for the dubbing process.
5. *Google Cloud Text-to-Speech*: A cloud-based API that generates human-like speech from text. It synthesizes Urdu audio from the translated text, providing the dubbed audio track to be integrated with the original video.
6. *LanguageTool*: A multilingual grammar and style checking tool used to evaluate the linguistic quality of the text. It detects grammatical errors and style issues, ensuring the quality and correctness of the translated and synthesized text.
7. *TextStat*: A library for readability and intelligibility analysis. It calculates readability scores, such as the Flesch Reading Ease Score, to assess how easy or difficult the text is to understand, ensuring clarity in both English and Urdu transcripts.

These libraries and tools are essential for the various stages of the video dubbing process, including extraction, translation, synthesis, and evaluation of the dubbed content.

C. Execution Strategy

The methodology employed in this project involves a series of well-defined steps to ensure the accurate and efficient dubbing of English videos into Urdu. The process is designed to enhance accessibility for Urdu-speaking audiences by replacing the original English audio with natural-sounding Urdu speech. The key steps are as follows:

1. *Extracting Audio from Video:*
   The first step involves isolating the audio track from the video file. This was accomplished using the MoviePy library, which enabled the extraction of the audio track from the MP4 video files, facilitating subsequent processing and manipulation.
2. *Converting Audio to Mono:*
   To prepare the audio for speech recognition, it was necessary to standardize its format. Initially, the audio was in stereo, but it was converted to mono using the Pydub library. This conversion ensures compatibility with the speech recognition system, which requires mono audio input.
3. *Converting Speech to Text:*
   The next step involves transcribing the spoken content into written text. This was achieved using the Google Cloud Speech-to-Text API. The mono audio was processed to produce an English text transcript, which is crucial for the translation into Urdu.
4. *Translating Text to Urdu:*
   Following transcription, the English text was translated into Urdu to maintain the meaning and context of the original content. This translation was performed using the Google Cloud Translate API, which facilitated accurate conversion of the text into Urdu.
5. *Converting Text to Speech:*
   The translated Urdu text was then synthesized into speech using the Google Cloud Text-to-Speech API. This step resulted in a new audio track featuring natural-sounding Urdu speech, ready to be integrated with the original video.
6. *Merging Audio with Video:*
   Finally, the new Urdu audio was integrated with the original video. Using the MoviePy library, the synthesized Urdu audio replaced the English audio track, and the final video was saved in MP4 format with the new Urdu audio track.

This comprehensive methodology ensures that the English videos are dubbed accurately and efficiently into Urdu, thereby improving content accessibility and engagement for Urdu-speaking audiences. Each step has been carefully designed to maintain the quality and coherence of the original video content while adapting it for a new linguistic audience.

D. Performance Evaluation

To ensure the quality and effectiveness of the dubbed videos, a comprehensive performance evaluation was conducted focusing on two primary aspects: linguistic quality and intelligibility.

Linguistic Quality Evaluation was performed using the LanguageTool, an automated grammar checking tool. This evaluation involved assessing the grammatical correctness and stylistic quality of the transcripts. The process includes analyzing the text for grammatical errors, style issues, and other language-related problems. The linguistic quality score is calculated based on the ratio of detected errors to the total number of words in the text, providing a percentage that reflects the overall quality of the transcript.

Intelligibility Analysis was carried out using TextStat to determine the readability and ease of understanding of the text. This analysis involves calculating the Flesch Reading Ease Score, which quantifies how easy or difficult the text is to read. The score helps in evaluating how well the translated Urdu text and the original English text can be understood by their respective audiences. Higher scores indicate better readability, while lower scores suggest that the text may be challenging to comprehend.

Overall, these evaluations are crucial for assessing the effectiveness of the dubbing process and ensuring that the final output is both linguistically accurate and easily understandable for the target audience.

E. Implementation Summary

The project aimed to dub English videos into Urdu by leveraging various tools and libraries to ensure a high-quality output. The implementation involved several key steps, each meticulously executed to achieve the desired result. The first step involved extracting audio from the video files using the MoviePy library. This allowed for the isolation of the audio track, which was then converted from stereo to mono using the Pydub library. This conversion was necessary to ensure compatibility with the speech recognition system.

Next, the mono audio was transcribed into English text using the Google Cloud Speech-to-Text API. This transcription process provided the foundation for the subsequent translation step. The English text was then translated into Urdu using the Google Cloud Translate API, maintaining the original content's meaning and context. The

Authorized licensed use limited to: UNIVERSIDADE FEDERAL DO PARANA. Downloaded on November 05,2025 at 19:34:52 UTC from IEEE Xplore. Restrictions apply.

translated Urdu text was synthesized into speech using the Google Cloud Text-to-Speech API, producing an audio track in Urdu. This new audio was then merged with the original video using MoviePy, replacing the English audio track with the synthesized Urdu speech.

To evaluate the effectiveness of the dubbing process, two key metrics were assessed: linguistic quality and intelligibility. Linguistic quality was evaluated using LanguageTool to check for grammatical errors and stylistic issues in the transcripts. Intelligibility was analysed using TextStat to measure the readability of the text, ensuring that both the English and Urdu versions were accessible and understandable.

In summary, the implementation involved a well-coordinated process of audio extraction, transcription, translation, speech synthesis, and merging, complemented by thorough performance evaluation. This approach ensured that the final dubbed videos met high standards of accuracy and clarity, enhancing content accessibility for Urdu-speaking audiences.

IV. RESULTS AND DISCUSSION

A. *Evaluation Results*

The evaluation of the dubbing process focused on assessing both the linguistic quality and intelligibility of the translated and synthesized content. To evaluate linguistic quality, automated grammar tools were used to assess the accuracy and correctness of the translated text. The Flesch Reading Ease Score was employed to gauge the readability and understandability of both the English and Urdu transcripts. This comprehensive evaluation provides insights into the implementation of the dubbing process and the effectiveness of the translation and synthesis steps.

1. *Linguistic Quality Evaluation*

The linguistic quality of the English text was assessed using the LanguageTool library. The results from several tested videos revealed a mean quality score which indicates the ratio of grammatical errors relative to the number of words, highlighting the overall accuracy of the translation process.

2. *Intelligibility Analysis*

Intelligibility was measured using the Flesch Reading Ease Score, which assesses the ease of reading and understanding the text. For the English transcripts, the readability scores ranged from moderate. For the Urdu transcripts, the scores were categorized as high to moderate. These results illustrate how effectively the translated text communicates with the target audience and how well the new Urdu audio aligns with the clarity of the original content.

B. *Performance Discussion*

The results from the linguistic quality and intelligibility evaluations highlight several key points about the dubbing process. The high-quality score for the translated Urdu text suggests that the translation and synthesis methods employed were effective in maintaining grammatical accuracy. However, there may still be room for improvement, particularly in enhancing the readability of the Urdu text to better match the original English content.

The intelligibility scores provide further insights into how well the dubbed content is likely to be understood by Urdu-speaking audiences. The Moderate readability of the Urdu text indicates that the synthesized speech is generally clear, but attention may be needed to address any areas where readability is lower.

C. *Challenges and Improvements*

The evaluation results highlight the effectiveness of the implemented dubbing methodology regarding linguistic quality and intelligibility. Despite the successes, there is potential for further improvement. Future efforts could focus on exploring advanced translation models and additional text refinement techniques to enhance readability. Moreover, incorporating feedback from native Urdu speakers could offer valuable insights into the practical effectiveness of the dubbed content. Continuous refinement and evaluation remain crucial for achieving the highest quality and accessibility for Urdu-speaking audiences.

Several challenges were encountered during the implementation of the video dubbing system. One significant issue was the impact of accent variability and background noise on transcription accuracy. To address these challenges, it is recommended to implement advanced noise reduction techniques and expand the training dataset to include a diverse range of accents. Additionally, incorporating accent-specific models could further enhance the reliability of transcription.

Another challenge was the limited voice options available for synthesis, which affected the naturalness and personalization of the dubbed content. To improve this, it would be beneficial to introduce multiple voice options, including different genders, ages, and tones. Developing a voice selection feature and incorporating voice modulation technologies could help create a more personalized and natural-sounding dubbing experience.

V. CONCLUSION AND FUTURE WORK

This research successfully developed an automated video dubbing system that translates English-language videos into Urdu by leveraging advanced speech technologies. The integration of Google Cloud's speech-to-text, translation, and text-to-speech services facilitated the creation of a workflow that enhances content accessibility for Urdu-speaking audiences. The system demonstrated effective transcription of English audio, accurate translation into Urdu, and clear synthesis of the dubbed audio. However, some challenges were encountered, including handling diverse accents, translating idiomatic expressions and cultural nuances. Despite these challenges, the system represents a significant step forward in automated dubbing technology, providing a practical tool for broadening the reach of multimedia content.

Future work should focus on improving transcription accuracy by addressing accent diversity and background noise is essential for achieving more precise text conversion. Expanding the range of synthesized voices to include varied tones and accents will contribute to more natural-sounding dubbing. Additionally, exploring the feasibility of real-time dubbing could significantly advance the system's capabilities, allowing for immediate translation and dubbing of live or streaming content. Extending the system to support multiple languages would also broaden its usability and impact, making it a more versatile tool in global media production.

VI. REFERENCES

Authorized licensed use limited to: UNIVERSIDADE FEDERAL DO PARANA. Downloaded on November 05,2025 at 19:34:52 UTC from IEEE Xplore. Restrictions apply.

[1] Marcello Federico, Robert Enyedi, Roberto Barra-Chicote, Ritwik Giri, Umut Isik, Arvindh Krishnaswamy, and Hassan Sawaf, “From Speech-to-Speech Translation to Automatic Dubbing,” in Proc. of IWSLT, Online, July 2020, pp. 257–264, ACL.

[2] Nami Susan Kurian, Kalaivani S, Sinchana K S, Subha T D, Sornisha N L, Sai Charan G, Kundana Sree P, “Transformative Perspectives: AI-Enabled Dubbing Software for Multilingual Content Localization.” African Journal of Biological Sciences, Volume 6, Issue Si2, 2024.

[3] William Brannon, Yogesh Virkar, Brian Thompson, “Dubbing in Practice: A Large-Scale Study of Human Localization with Insights for Automatic Dubbing,” Transactions of the Association for Computational Linguistics, Vol.11,2023, pp. 419–435.

[4] Anna Matamala, Elisa Perego and Sara Bottiroli, “Dubbing versus subtitling yet again? An empirical study on user comprehension and preferences in Spain” Babel, Volume 63, Issue 3, Jan 2017, p. 423 – 441.

[5] Surafel M. Lakew, Yogesh Virkar, Prashant Mathur, and Marcello Federico. 2022. ISOMETRIC MT: Neural machine translation for automatic dubbing. In ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pages 6242–6246, Singapore.

[6] Alp Oktem, Mireia Farr ¨ us, and Antonio ´ Bonafonte. 2018. “Bilingual prosodic dataset compilation for spoken language translation.” In IberSPEECH 2018, pages 20–24. ISCA.

[7] Shenbagaraj, and Sailesh Iyer,” An Intelligent System for Automated translation of Videos from English to Native Language applying Artificial Intelligence Techniques for Adaptive eLearning,” International Journal of Intelligent Systems and Applications in Engineering 12.3s (2024), pp. 620-640.

[8] Sinha, Rohit Mahesh K., et al.” ANGLABHARTI: a multilingual machine aided translation project on translation from English to Indian languages,” IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century, Vol. 2, IEEE, 1995.

[9] Aasha, V. C, and Amal Ganesh,” Machine translation from English to Malayalam using transfer approach, ”2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)”, IEEE, 2015.

[10] Karthikadevi M, and K. G. Srinivasagan,” The development of syllablebased text to speech system for Tamil language,” 2014 International Conference on Recent Trends in Information Technology, IEEE, 2014.

[11] Alp Oktem, Mireia Farr ¨ us, and Antonio Bonafonte, “Prosodic ` Phrase Alignment for Machine Dubbing,” in Proc. Interspeech, 2019

Authorized licensed use limited to: UNIVERSIDADE FEDERAL DO PARANA. Downloaded on November 05,2025 at 19:34:52 UTC from IEEE Xplore. Restrictions apply.