Vividha Hub: AI-Powered Multilingual Video Dubbing and Subtitling

Project Report

Team Members :

Ch.Pushkar 2303A52362,
Ch.Shivasai 2303A52488,
M.Siddhartha 2303A52443,
K.Adithya 2303A52490

Mentor : Dr Pramod Kumar Poladi

Department : CS&AI

University : Sr University

Contents:

Abstract

Vividha Hub is an innovative, open-source AI platform that automates video dubbing and subtitling into multiple languages, reducing costs by up to 90% and processing times by a factor of ten compared to traditional methods. Leveraging technologies like OpenAI Whisper, Coqui TTS, PyAnnote, and FFmpeg, it transcribes, translates, and dubs videos while preserving speaker voices and background music. This report details the technical framework, key innovations, practical applications in education and media, and future directions for enhancing global content accessibility.

1 Introduction

In an increasingly connected world, video content is a vital medium for education, entertainment, and communication. However, language barriers often limit its accessibility, particularly in linguistically diverse regions like India. Vividha Hub addresses this challenge by offering an AI-powered platform that automates video dubbing and subtitling into multiple languages, such as Hindi, Telugu, and Tamil. By leveraging advanced technologies like automated transcription, voice cloning, and intelligent synchronization, Vividha Hub delivers high-quality localized videos with up to 90% cost savings and tenfold faster processing compared to traditional methods. Its open-source nature and modular architecture make it adaptable for integration into various platforms, from educational tools to streaming services, fostering inclusive communication and global reach.

2 Technical Overview

Vividha Hubs backend integrates advanced AI and multimedia processing technologies into a seamless pipeline:

1. **Video Upload and Audio Extraction**: Users upload videos (e.g., MP4) via a web interface. The audio is extracted using MoviePy and saved as a WAV file.
2. **Transcription**: OpenAI Whisper transcribes audio into English text with high accuracy, handling multi-speaker and noisy inputs.
3. **Translation**: Google Translator converts text into the target language, generating an SRT file for subtitles.
4. **Speaker Diarization**: PyAnnote identifies distinct speakers, extracting voice samples for dubbing.
5. **Voice Cloning and Dubbing**: Coqui TTSs XTTS-v2 model generates dubbed audio, preserving speaker characteristics.
6. **Background Music Synchronization**: Pydub mixes dubbed audio with the original audio (at -20 dB) to retain background music.
7. **Subtitle Rendering**: FFmpeg embeds subtitles using DejaVu Sans font, styled with white text, black outline, and a semi-transparent background.

   It processes a 1-minute video in approximately 2-3 minutes, with GPU

utilization optimized to 70-80%.

3 Key Innovations and Advantages

Vividha Hub offers significant advantages over traditional and other AI-based localization solutions:

- **Cost Efficiency**: Automates processes to reduce costs by up to 90%.
- **Speed**: Processes videos ten times faster than manual methods.
- **Authenticity**: Preserves original speaker voices and background music for an engaging experience.
- **Open-Source**: Enables integration into platforms like YouTube or Moodle, with community-driven improvements.
- **Scalability**: Supports diverse sectors with a robust API.

Table 1: Comparison with Traditional Methods

Feature	Vividha Hub	Traditional Methods
Cost	Up to 90% savings	High (manual labor, studios)
Speed	10x faster	Slow (days to weeks)
Voice Authenticity	Preserves speaker style	Often uses different voice actors
Background Music	Retained	Often replaced or lost
Integration	Open-source, API-driven	Proprietary, limited flexibility

4 Use Case Examples and Practical Impact

Vividha Hubs versatility enables applications across multiple sectors:

- **Education**: Platforms like BYJUS and Vedantu can provide multilingual lessons, making education accessible in native languages.
- **Media and Entertainment**: Streaming services like YouTube and Netflix can dub and subtitle content to reach global audiences.
- **Film Industry**: Production houses can localize movies cost-effectively.
- **Podcasting and Music**: The platform can translate podcasts or subtitle music videos, broadening their appeal.
- **Business**: Companies can localize training and marketing materials for diverse markets.

The platform enhances accessibility, reduces costs, and expands global reach, fostering cultural exchange and inclusive communication.

5 Conclusion and Future Directions

Vividha Hub is a transformative tool for video localization, offering cost-effective, fast, and authentic multilingual content creation. Its successful processing of videos like ‘minute.mp4‘ demonstrates its potential. Future enhancements include expanding language support, adding lip-sync with Wav2Lip, and developing a standalone desktop application with PyQt5 or Electron.

6 Links to Videos

- Original video before editing: [https://drive.google.com/file/d/1_0MYpuf1q5I0ejuMqSuwleHaXlPGk0UO/view?usp=sharing]
- Video after editing: [https://drive.google.com/file/d/1mvI1GSNVGwWEc4lI-mKEJkM7_z6H4Y6X/view?usp=sharing]
- Video explaining how the project works: [https://drive.google.com/file/d/1s_eqG27NaCb3P-DJUBktJ-MqhyEpcayC/view?usp=sharing]
- Other tested videos:
- 1 . [https://drive.google.com/file/d/1na3TeoVDLi3Y1RN6ce9xXeRGQUkWmMr-/view?usp=sharing] English to spanish_Audio-french_Subtitles
- 2.[https://drive.google.com/file/d/1TRIvA4lwb-qW34T1tT6bRJSgIBmu7lKu/view?usp=sharing] English to Hindi_Audio-Hindi_Subtitles
- 3.[https://drive.google.com/file/d/1eEs1oy7fqDenPfY9U1u_5bZ2SqDwwGb9/view?usp=sharing] Original video of 2
- 4.[https://drive.google.com/file/d/1KWpD2hEocZYu9REk1Ywj9ctXR8AzzwE5/view?usp=sharing] English to Hindi_Audio-Hindi_Subtitles
- 5.[https://drive.google.com/file/d/1_Q-1GQkd1EdqdDq7HY5H7UvufUEpbuVZ/view?usp=sharing] Original video of 4

7 References

- OpenAI Whisper: https://github.com/pyannote/pyannote-audio
- Coqui TTS: coqui/XTTS-v2 · Hugging Face
- PyAnnote Audio: pyannote/pyannote-audio: Neural building blocks for speaker diarization: speech activity detection, speaker change detection, overlapped speech detection, speaker embedding
- FFmpeg: https://ffmpeg.org
- MoviePy: https://zulko.github.io/moviepy
- Pydub: https://github.com/jiaaro/pydub
- Google Translator: https://github.com/nidhaloff/deep-translator