VOCALBRIDGE – BRIDGING LANGUAGES WITH VOICE PRESERVATION

Project Reference Number: 49S_BE_5726

College : Shri Madhwa Vadiraja Institute of Technology and Management, Udupi
Branch : Department of Computer Science & Engineering
Guide(s) : Ms. Sahana
Student(s) : Mr. K Sudarshan Hebbar
Mr. Kaushik
Mr. Pratap Naik
Mr. Prathvi S Jathan

Keywords:
Artificial Intelligence, Voice Cloning, Neural Machine Translation, Automatic Speech Recognition, Content Localization

Introduction:

The rapid growth of digital media platforms has significantly increased the demand for multilingual content. Educational videos, lectures, tutorials, and entertainment media are now consumed globally. However, language remains one of the biggest barriers to accessibility. Traditional dubbing techniques require professional voice actors, recording studios, and extensive manual editing. This makes the localization process expensive and time-consuming. Moreover, manual dubbing often fails to preserve the original speaker's tone, emotion, and speaking style. To address these challenges, this project introduces VocalBridge, an AI-powered automated dubbing system. The system translates speech from one language to another while maintaining the original speaker's vocal identity. By integrating Automatic Speech Recognition, Neural Machine Translation, and Voice Cloning technologies, the system ensures seamless multilingual content creation. The project focuses on preserving speech characteristics such as pitch, tone, and emotional context. This approach enhances viewer engagement and authenticity. The solution benefits educators, content creators, and businesses aiming for global reach. The system also reduces production cost and time. Additionally, it promotes inclusivity by making content accessible to non-native speakers. The project demonstrates the power of AI in content localization. The design emphasizes modular architecture and scalability. The system supports future extensions such as real-time translation. Overall, VocalBridge aims to bridge linguistic gaps using intelligent automation.

Objectives

To develop an automated pipeline for speech-to-speech translation.
• To preserve the original speaker's voice using AI-based voice cloning.
• To perform accurate speech recognition from video input.
• To implement neural machine translation for multilingual conversion.
• To synchronize generated speech with original video timing.
• To reduce cost and effort in manual dubbing.
• To improve accessibility of educational and multimedia content.
• To design a scalable architecture for future real-time deployment.

Methodology

The methodology consists of multiple stages including audio extraction, speech recognition, translation, voice cloning, and audio-video integration. Initially, the input video is processed using FFmpeg to extract the audio stream. The extracted audio is cleaned and normalized to improve transcription accuracy. Next, Automatic Speech Recognition is performed using Whisper model to convert speech into text. The transcribed text is then passed to Neural Machine Translation using IndicTrans to convert source language into target language. After translation, the translated text is synthesized into speech using zero-shot voice cloning with Coqui TTS. The generated speech maintains speaker identity. Temporal alignment is achieved using Librosa and Pydub. Finally, the synthesized audio is merged with original video.

Methodology Flow Diagram:

Input Video → Audio Extraction → Speech Recognition → Text Translation → Voice Cloning → Temporal Alignment → Audio-Video Integration → Output Dubbed Video

Results & Conclusions

The developed system successfully translated and dubbed video content. The ASR module produced accurate transcription for English speech. The translation module generated context-aware multilingual output. The voice cloning module preserved speaker characteristics. The synchronization module aligned generated speech with original video. The output video maintained natural pacing and intelligibility. The system reduced manual effort significantly. Testing showed improved accessibility for non-native audiences. The results confirm feasibility of AI-based automated dubbing. Overall, the system demonstrates practical usability and effectiveness.

Project Outcome & Industry Relevance

1. The project provides a scalable automated dubbing solution.
2. It benefits content creators, educators, and OTT platforms.
3. The system reduces production cost and time.
4. It enables global distribution of localized content.
5. Industries such as e-learning, media, and entertainment can adopt this technology.
6. The approach improves accessibility and engagement.
7. The solution supports multilingual communication.

Project Outcomes and Learnings

1. The project enhanced understanding of AI-based speech processing.
2. Students learned integration of ASR, NMT, and TTS modules.
3. Experience was gained in audio processing libraries.
4. The team improved debugging and system design skills.
5. The project strengthened knowledge of machine learning workflows.

Future Scope

1. Future enhancements include real-time translation support. Additional languages can be integrated.
2. Lip-sync technology can improve visual synchronization. Cloud deployment can enable large-scale processing.
3. Noise-robust models can improve performance.
4. The system can be extended to live streaming.
5. Mobile application integration is possible.
6. The project can support subtitle generation.
7. Advanced emotion transfer can be implemented.
8. Industry-level optimization can be added.