

ABSTRACT

We introduce an innovative solution for automated synchronization and translation of video content. Using AI-generated voices, the system effortlessly translates and dubs video conversations into numerous languages by utilizing pre-existing SRT subtitle files. The process includes synthesizing realistic neural voices, estimating audio clip lengths based on subtitle timings, and translating text using Google Cloud or DeepL. Key features include the generation of translated SRT files, batch processing for multiple languages, and configurable settings for detailed control over translation and synthesis. External prerequisites for using this system include the installation of ffmpeg and optional utilities such as Rubber Band for stretching audio. `Cloud_service_settings.ini` allows users to set up DeepL API access, Microsoft Azure, Google Cloud, and other services. The project provides text-to-speech with neural voices using Microsoft Azure, Google Cloud, and Eleven Labs, as well as text translation using Google Translate and DeepL. The system works with Python 3.9 and previous iterations. The `main.py` script may be used by users to produce dubbed audio tracks and translated subtitle files that are kept in an output folder. The project also provides optional scripts for things like translating YouTube titles and descriptions and adding language tracks to video files. This adaptable project provides a comprehensive solution for content makers looking to easily reach a variety of worldwide audiences by supporting numerous Text-To-Speech and translation services.