Transcription Software with Translation and Audio Streaming

Dona Davis, Student, CSE, IESCE, Kerala, India
Guide: Ms. Hrudya K P, Faculty, CSE, IESCE, Kerala, India

ABSTRACT

The design and implementation of speech-to-text and text-to-speech systems to cater for language inefficiency has been seen to be highly encouraging in recent times. This enables us to incorporate speech technology to our data independent of the platform. Over the past few decades, people have tended to connect themselves to audio elements more than written contents. Its use with different speech recognition technologies paves way for advanced developments. In this project, a web-based speech-to-text system is presented where a voice recognition module builds into the system to perform the speech recognition task while a dedicated application does the speech transcription work. The transcribed text could be further translated into various other languages with the use of Python modules. This may help to overcome the language barrier all over the world and bridge the gap between different communities. It aims to develop a cost-effective service of providing audio information for a larger crowd.

KEYWORDS

Audio Extraction, Speech Recognition, Audio Transcription, Text Translation, Audio Streaming, Socket Programming, Client-Server Connection.

INTRODUCTION

The Web-based Transcription software aims to develop a platform for audio transcription. Python libraries make it very easy to handle functions relating to audio files. The Speech Recognition library available in Python allows to capture audio effectively and enables it to be transcribed into a textual form. The text content could be further translated to other desired languages based on the user's choice using the Google Translate library of Python, which allows translation to more than 100 plus languages. Further, the translated text could be given back as an audio input after converting the text into suitable audio format. The final audio could be streamed across multiple devices based on the user's need by establishing an efficient Client-Server connection.

MODULE DESCRIPTION

1. Audio Extraction

The audio could be easily extracted from a video using some models like tkinter, os, PIL and moviepy.

1. tkinter – for use Interface(UI)
2. os – provides functions for creating and removing a directory (folder), fetching its contents, changing and identifying the current directory, etc.
3. moviepy – MoviePy is a Python module for video editing, which can be used for basic operations (like cuts, concatenations, title insertions), video compositing (a.k.a. non-linear editing), video processing, or to create advanced effects.

4. PIL – Pillow is the friendly PIL fork by Alex Clark and Contributors. PIL is the Python Imaging Library by Fredrik Lundh and Contributors.

2. Audio to Text Conversion

Speech recognition is the ability of computer software to identify words and phrases in spoken language and convert them to human-readable text. We can easily convert speech to text in Python using the SpeechRecognition library.

As a result, we do not need to build any machine learning model from scratch, this library provides us with convenient wrappers for various well-known public speech recognition APIs (such as Google Cloud Speech API, IBM Speech To Text, etc.).

3. Text Translation

Google Translate is a free service that translates words, phrases, and entire web pages into more than 100 languages. We will be performing language translation in Python using Googletrans library. Googletrans is a free and unlimited Python library that makes unofficial Ajax calls to Google Translate API to detect languages and translate text. We can even detect languages using this library among others.

Main features of this library:

Auto language detection (it offers language detection as well)

Bulk translations
Fast & reliable
HTTP/2 support
Connection pooling

4. Text to Audio Conversion

There are a whole bunch of text-to-speech tools available for Python such as Google Text-to-Speech API (gTTS API for short).

Using gTTS API, we can simply feed a text document to it and get an mp3 file out with the spoken version of the text.

5. Audio Streaming

A sample cloud based Streaming System can be implemented by developing a web based service on Java/Oracle Platform to be hosted on the cloud that streams audio files with the help of an embedded media player.

CONCLUSION

An efficient Web-based Transcription software could be developed for audio transcription. We could capture audio effectively and transcribe it into a number of desired languages. Further, the translated text could be converted into suitable audio format. The final audio could be streamed across multiple devices.

ACKNOWLEDGEMENT

REFERENCES

1. Automatic Speech Recognition - **Authors:** Anchal Katyal, Amanpreet Kaur, Jasmeen Gill
2. Google Translate and Translation Quality - **Author:** Karl Griffith
3. Multimedia Streaming and Related Issues - **Authors:** Sober Lourembam, Venkatata Rajesh M, Meina Singh