Video Subtitle Generation

Aishwarya Bakale
Department of Computer Engineering
K.K. Wagh Institute of Engineering Education And Research, Nashik

Pratiksha Baviskar
Department of Computer Engineering
K.K. Wagh Institute of Engineering Education And Research, Nashik

Radhika Bharambe
Department of Computer Engineering
K.K. Wagh Institute of Engineering Education And Research, Nashik

Ashwini Borse
Department of Computer Engineering
K.K. Wagh Institute of Engineering Education And Research, Nashik

Abstract

Video plays an important role to help people understand and comprehend the information in much easier way, for example movies, songs or video lectures relevant to user. This can be best done by generating subtitles of the video. At present there are some systems which are able to generate only speech to text, not the video to subtitles in efficient way. Also downloading subtitles of any video from the internet is complex process. To generate subtitles of video automatically through software is solution for this. Hence, in this proposed system subtitles can be generated through three different phases namely audio extraction means separating audio from video and converting audio in suitable format if necessary. The second phase proceeds to the recognition of speech contained in the audio, before giving audio for speech recognition audio is divided into number of chunks. The subtitle generation phase generate subtitle file from the recognition results of the previous step. Subtitle generation is the final process to get subtitles as output of video in which a .txt/.srt file is generated synchronized with the input file. This system can help people to remove the gaps of their native language and the auditory problem. It can also help children with word identification, meaning, acquisition and retention, and it can establish a systematic link between written word and the spoken word.

Keywords: Audio extraction, Audio Segments, Chunks, Speech Recognition, Time Synchronization, Subprocess , Video text merging

I. INTRODUCTION

In VLC Media Player Automatic Subtitle Files Are Generated By Using Our System, No Manual Intervention Is Required. As We Have Used Video To Text Generation Using Google API.

Input for the system is downloaded video files with .mp4 format. After processing, .srt file will be generated which is used for adding subtitles for playing video. Video file with .mp4 format is input to system using FFMPEG module, audio is extracted from video. This whole audio file is divided into number of chunks. After this by using speech recognizer, text file is generated for each audio chunk. Text output of each audio chunk is sequentially added and .srt file is generated. Synchronizing .srt file while playing video, subtitles are generated on screen. The system is divided into four modules:

1) Video to Audio Extraction
2) Make Audio Chunks
3) Speech Recognition
4) .srt File Generation

II. METHODOLOGY

A. *Video to Audio Extraction:*

For extracting audio from video, FFMPEG module is used. This include subprocess command in python system which is used to execute command prompt command, then install FFMPEG module in system by doing this it will able to extract audio from video. After executing FFMPEG command it will get audio from .mp4 video format[2]. In this .mp4 is converted to .wav audio file format.

1) Subprocess command:

This module allows to spawn new processes, connect to their input/output/error pipes. This module will replace to several old modules like os.system, os.spawn* in python.

2) FFMPEG:

It stands for “Fast Forward MPEG”. It is a software which is used for producing libraries and programs for handling multimedia data. It is a free software published under GNU Lesser General Public License.

B. Make Audio Chunks:

The whole extracted audio file cannot be given to speech recognizer for speech recognition that gives error. As system is using Google API for Speech Recognition, it does not recognize whole file at a time. So this extracted audio is converted into proper number of chunks. For making chunks it uses AudioSegment command where extracted audio is given as input i.e. (.wav file format) and length of chunks is provided in miliseconds. After this chunks are created.

C. Speech Recognition:

For recognition of converted audio that is converting it into text file format, system is using Google API for speech recognition using speech recognizer API. Then applying that to different audio chunks, audio is converted into text. That text file is created in the form of .srt file format from which subtitles are generated while playing video.
Speech Recognition API: It is a Google Cloud Speech API which enables developers to convert audio to text by applying neural networks to it. The audio to be recognized uploaded in form of request.

D. .srt File Generation:

For playing proper subtitles of video, system must have time synchronization mechanism for that system need to import (date, time class) and then after synchronization subtitles are added to the video. For this, system is using VLC Media Player where for generating subtitles, it adds that subtitle (.srt) file then video played with subtitles[1].

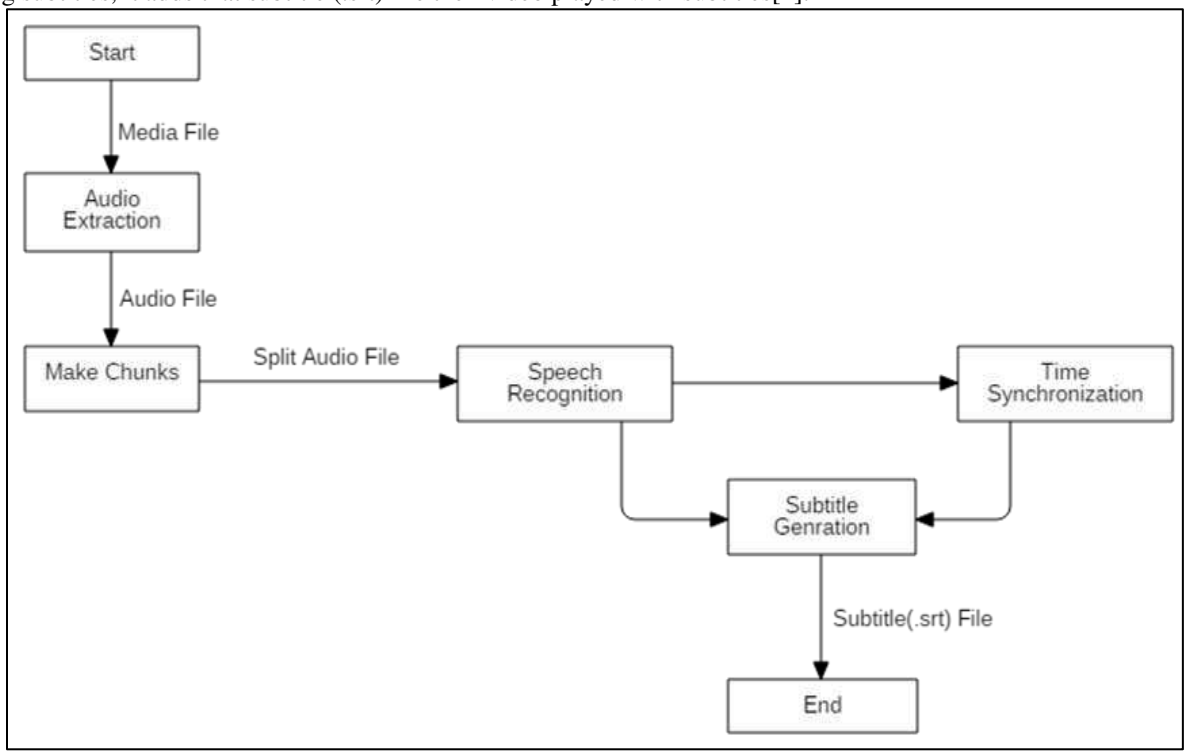


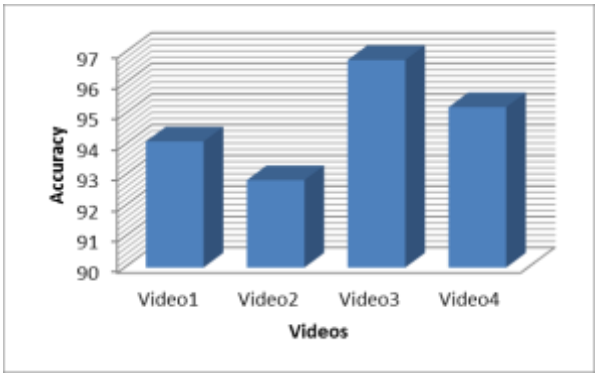
Fig. 1: Block Diagram of System

III. RESULTS

In VLC Media Player automatic Subtitle files are generated while using system, no manual intervention is required. As subtitle generation is done by using Google API, processing through different modules. Subtitles are displayed while playing the video[3]. The below graph shows accuracy of generated subtitles for different videos.
Formula for Accuracy= Correctly recognized words / Total no. Of words[4]

Table - 1
Accuracy of Videos

NO. OF VIDEOS	*TOTAL NO. OF WORDS*	*CORRECTLY RECOGNIZED WORDS*	*ACCURACY (%)*
VIDEO 1	*102*	*96*	*94.11*
VIDEO 2	*56*	*52*	*92.85*
VIDEO 3	*186*	*180*	*96.77*
VIDEO 4	*161*	*154*	*95.23*



Graph – 1: Accuracy of Videos

IV. CONCLUSIONS

Last some years have been the witnesses the tremendous use of any kind of video content. There is great demand for different useful videos. At the same time, certain individuals with auditory problem or people with language gap cannot understand the meanings of such videos because there is no text transcription available. Hence, it is important to find solutions for the purpose of making these media artefacts accessible for most people. For this different software are available but it is not working properly and requires more user participation. Hence, an automated system is to be created.

This system is going to help in various ways i.e. social, educational. It will help students with word identification, meaning, acquisition i.e. to get complete and more understanding of that video. It will help people with auditory problem to understand video. And there is no need to specially download subtitle files of that video from internet which is complex process.

ACKNOWLEDGMENT

We would like to take this opportunity to thank our internal guide Prof. S. T. Patil for giving us all the help and guidance we needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful. We are also grateful to Prof. Dr. S. S. Sane, Head of Computer Engineering Department K. K. Wagh Institute of Engineering Education and Research for his indispensable support, suggestions.

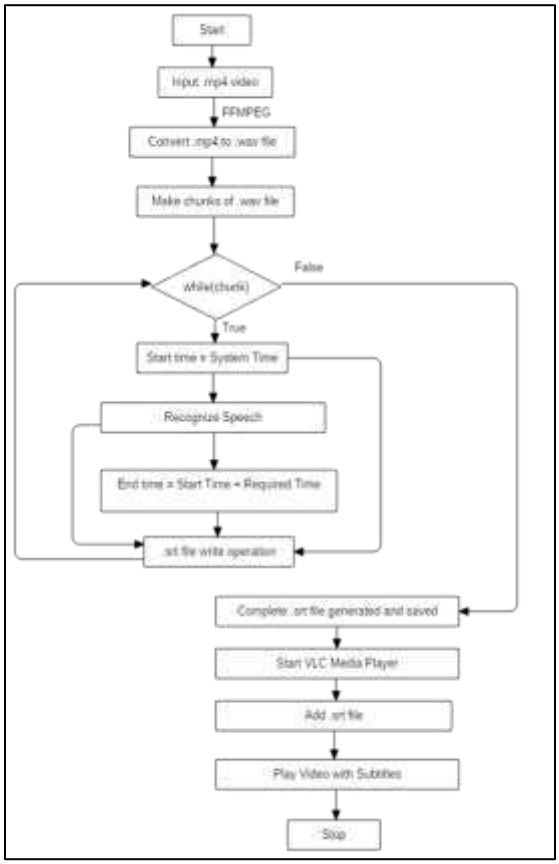


Fig. 2: Architecture of System

REFERENCES

[1] A. Mathur, T. Saxena and R. Krishnamurthi, "Generating Subtitles Automatically Using Audio Extraction and Speech Recognition," 2015 IEEE International Conference on Computational Intelligence & Communication
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7078779&isnumber=7078645

[2] X. Che, S. Luo, H. Yang and C. Meinel, "Automatic Lecture Subtitle Generation and How It Helps," 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT), Timisoara, 2017, pp. 34-38.
URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8001709&isnumber=8001692

[3] R. Sridhar, S. Aravind, H. Muneerulhudhakalvathi and M. Sibi Senthur, "A hybrid approach for Discourse Segment Detection in the automatic subtitle generation of computer science lecture videos," 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI),New Delhi, 2014, pp. 284-287. URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6968422&isnumber=6968191

[4] https://blog.mycon.com/blog-post/assessing-reading-fluency—word-recognition-accuracy