

# ”Captions With YouTube Transcripts”

Manvendra Singh  
IIIT Raichur, Karnataka, India

Shubham Soni  
IIIT Raichur, Karnataka, India

**Abstract**—YouTube is the world’s largest video-sharing platform, with billions of users and videos covering a wide range of topics. With the rise of video content, there has been an increasing need for accessibility of video content. Capturing YouTube captions is an essential part of creating accessible and inclusive video content. However, producing accurate captions can be a challenging task for content creators, particularly when it comes to handling accents, regional dialects, and technical jargon. This is where YouTube transcription can provide a solution.

YouTube transcription involves converting the audio of a video into a text format. This text can then be used to create captions that accurately reflect the spoken words in the video. Transcription can also help identify any errors or inaccuracies in the captions, allowing creators to correct them before publishing the video. It also helps students to focus more on class than to note down the content being delivered. Moreover, YouTube transcription can make the video more detectable by search engines as the transcript provides searchable text that can improve the video’s Search engine optimization. This can lead to increased views and engagement for the video.

However, YouTube transcription can be a time-consuming process, especially for longer videos. This is where automated transcription services can provide a solution. These services use advanced speech recognition technology to quickly and accurately transcribe videos, saving creators time and effort.

Another challenge with producing captions is ensuring that they are properly synced with the video. This is where YouTube’s automatic timing feature can provide a solution. This feature automatically synchronizes the captions with the video, making the process much simpler for creators.

In conclusion, capturing YouTube captions with YouTube transcription offers a solution to the challenges of creating accurate captions and enhancing accessibility for a wider audience. Automated transcription services and YouTube’s automatic timing feature can help to simplify the process and can save time and effort of both the creators and viewers.

Finally, the paper presents some best practices for creating transcripts, including identifying relevant keywords and using proper formatting to improve readability. Overall, this paper aims to highlight the importance of YouTube transcripts and how they can improve the accessibility and searchability of video content on the platform.

**Index Terms**—Write important keywords here

## I. INTRODUCTION

Capturing YouTube captions is an essential part of creating accessible and inclusive video content. It helps viewers who are hearing impaired or have language barriers to access the content more easily while pursuing it from text converted from audio. This text can then be used to create captions that accurately reflect the spoken words in the video. However, producing accurate captions can be a challenging task for content creators, particularly when it comes to handling accents, regional dialects, and technical jargon.

Apart from enhancing accessibility, YouTube transcription also has other benefits. Firstly, it can make the video more detectable by search engines as the transcript provides searchable text that can improve the video’s Search engine optimization. This can lead to increased views and engagement for the video. Secondly, it helps students to focus more on class than to note down the content being delivered.

While YouTube transcription offers a solution to the challenges of creating accurate captions, it can be a time-consuming process, especially for longer videos. However, automated transcription services can provide a solution. These services use advanced speech recognition technology to quickly and accurately transcribe videos, saving creators time and effort.

Another challenge with producing captions is ensuring that they are properly synced with the video. This is where YouTube’s automatic timing feature can provide a solution. This feature automatically synchronizes the captions with the video, making the process much simpler for creators.

In conclusion, capturing YouTube captions with YouTube transcription offers a solution to the challenges of creating accurate captions and enhancing accessibility for a wider audience. Automated transcription services and YouTube’s automatic timing feature can help to simplify the process and can save time and effort of both the creators and viewers.

## II. LITERATURE REVIEW

Automated transcription services are now widely available, which use advanced speech recognition technology to quickly and accurately transcribe videos, saving creators time and effort. However, the accuracy of these services can be affected by background noise, accents, and technical terms. For this reason, some creators choose to transcribe videos manually, which can be time-consuming but can produce more accurate transcripts.

Proper formatting of transcripts, such as using line breaks, capitalization, and punctuation, can also improve the readability of the captions. The use of relevant keywords in the transcript can also enhance the video’s search engine optimization, leading to increased views and engagement.

In addition to accessibility, YouTube transcripts also offer other benefits. Transcripts can provide a searchable text version of the video’s content, which can improve the video’s search engine optimization. This can lead to increased views and engagement for the video. Additionally, transcripts can be used as a reference tool for viewers who want to revisit

a particular section of the video or for content creators who want to revisit the video's content in other formats.

However, YouTube transcription can be a time-consuming process, especially for longer videos. This is where automated transcription services can provide a solution. These services use advanced speech recognition technology to quickly and accurately transcribe videos, saving creators time and effort. However, the accuracy of these services can be affected by background noise, accents, and technical terms.

Another challenge with producing captions is ensuring that they are properly synced with the video. This is where YouTube's automatic timing feature can provide a solution. This feature automatically synchronizes the captions with the video, making the process much simpler for creators.

In conclusion, capturing YouTube captions with YouTube transcription offers a solution to the challenges of creating accurate captions and enhancing accessibility for a wider audience. Automated transcription services and YouTube's automatic timing feature can help to simplify the process and can save time and effort of both the creators and viewers.

Finally, the paper presents some best practices for creating transcripts, including identifying relevant keywords and using proper formatting to improve readability. Overall, this paper aims to highlight the importance of YouTube transcripts and how they can improve the accessibility and searchability of video content on the platform.

### III. PROBLEM STATEMENT

#### Problem Statement:

The problem at hand encompasses the lack of universally accessible and comprehensive YouTube video transcripts. Many videos on YouTube either lack subtitles altogether or have inaccurate and incomplete captions. This poses a significant barrier for individuals who rely on subtitles for understanding the content, including those with hearing impairments, language barriers, or learning disabilities. Furthermore, the absence of accurate transcripts hampers the searchability and discoverability of video content. Therefore, there is an urgent need to develop a robust solution that ensures the availability of accurate and comprehensive transcripts for all YouTube videos, regardless of whether they have pre-existing subtitles or not. This solution would enhance accessibility, improve user experience, and promote inclusivity on the YouTube platform.

### IV. ARCHITECTURE

#### YouTube API Integration:

The system integrates with the YouTube API, allowing access to video metadata, captions, and other relevant information. An API key is used to authenticate and authorize the system's requests to the YouTube API.

#### Video Processing and Classification:

The system receives a YouTube video URL as input. It determines the type of video by checking if pre-existing captions are available or not. If pre-existing captions are available, the system proceeds to extract the captions using the YouTube API directly. If no captions are available, the

system downloads the video using a library like "pytube" and converts it to a WAV audio format using "moviepy".

#### Caption Extraction and Transcription:

For videos with pre-existing captions, the system extracts the captions using the YouTube API and converts them into a text format. For videos without pre-existing captions, the system performs audio-to-text transcription using a speech recognition library like "SpeechRecognition". The audio file in WAV format is processed, and the speech recognition library converts the spoken words into text.

#### Transcript Generation and Formatting:

The extracted or transcribed text is processed and formatted to generate a coherent transcript. The system cleans the text, removing unnecessary characters, timestamps, or other artifacts. Formatting options, such as separating text into paragraphs or adding speaker labels, can be applied as required.

#### Output and Storage:

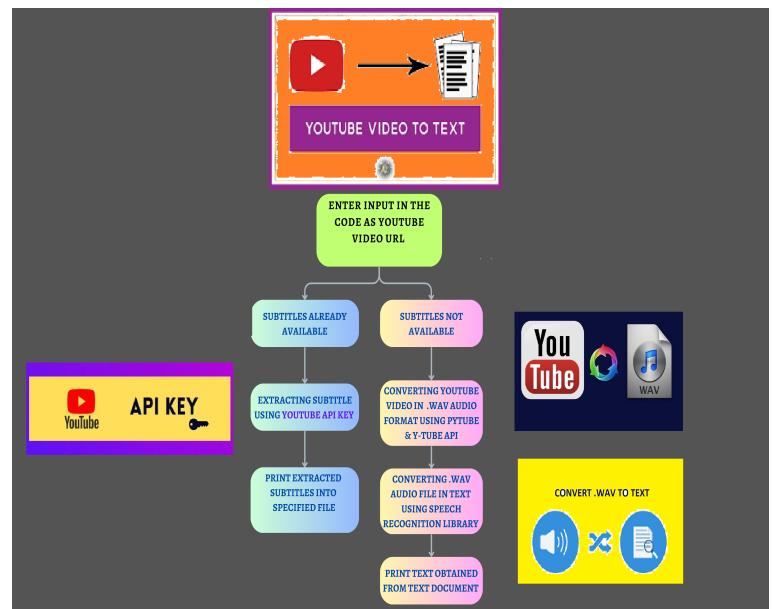
The final transcript is saved in a text file or a designated storage location. The system may also provide options to directly display or download the transcript to the user.

#### Error Handling and Logging:

The system includes error handling mechanisms to address exceptions or failures during the API requests, caption extraction, or audio-to-text transcription. Detailed logs are maintained to track the progress, errors, and status of the transcript generation process.

#### User Interface (Optional):

If the system includes a user interface, users can input a YouTube video URL and initiate the transcript generation process. The interface may display the progress, status updates, and provide options to download or view the generated transcript.



## V. ALGORITHM

This is a Python code that does the following:

1.Imports the necessary libraries: YouTubeTranscriptApi for extracting the transcript of a YouTube video, pytube for downloading the video, moviepy. editor for converting the video to audio, and speech recognition for converting the audio to text.

2.important function :

a.download video(URL) : downloads the YouTube video with the given URL and returns the filename of the downloaded video.

b.convert video to audio : converts the video file at video path to an audio file at output path, using moviepy. The output audio file is in WAV format.

c. convert audio to text converts the audio file at audio file to text using Google's speech recognition API through the speech recognition library. The function returns the recognized text.

3.Specifies the URL of a YouTube video to process.

4.Tries to extract the transcript of the YouTube video using YouTubeTranscriptApi. If successful, it writes the transcript to a file named xyz.txt and prints a message indicating that the transcript extraction is complete. If it fails, it prints an error message and continues.

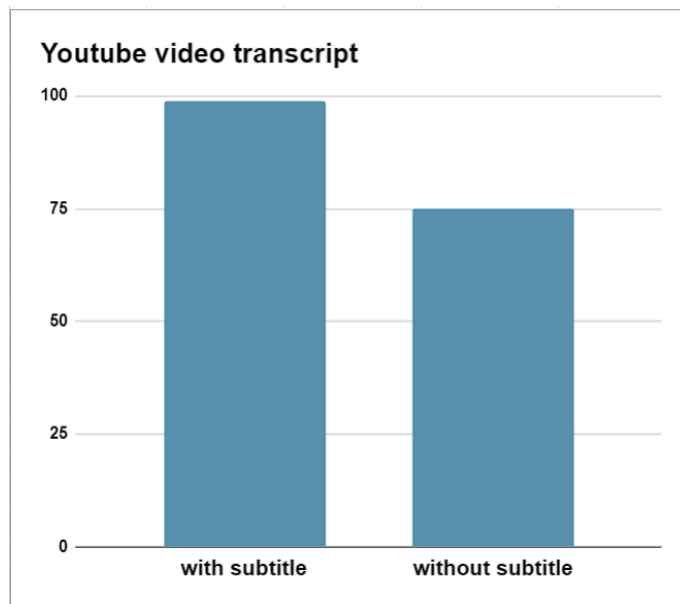
5. If transcript extraction fails, the code downloads the video, converts it to audio, and then converts the audio to text using speech recognition. The output text is written to a file named output text .txt, and the recognized text is printed to the console.

Overall, this code demonstrates how to extract the transcript of a YouTube video and how to convert a video to text.

## VI. SIMULATION AND RESULTS ANALYSIS

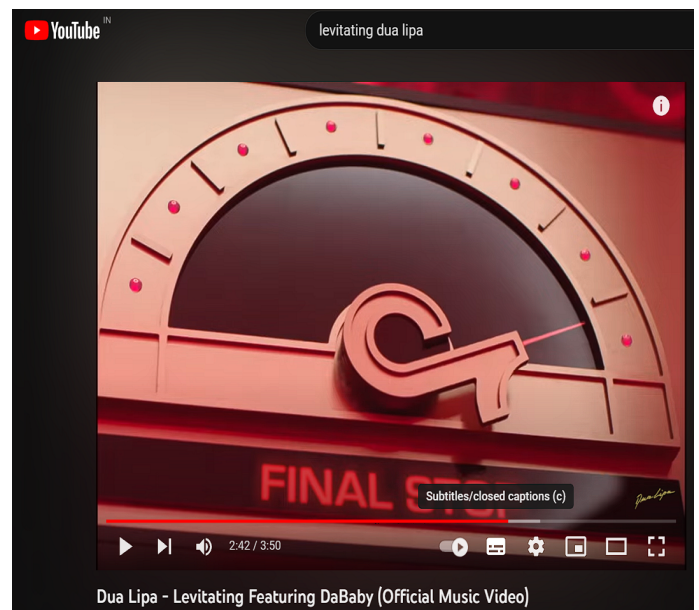
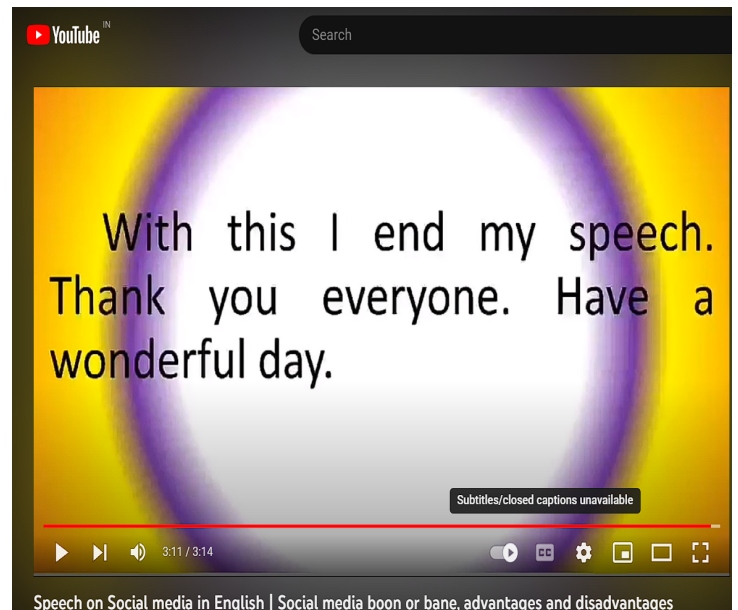
OUTPUT GRAPH:

WITHOUT SUBTITLE V/S WITH SUBTITLE



here some screenshot of input output

- (1) without subtitle YouTube video
- (2)with subtitle YouTube video
- (3)output of without subtitle video
- (4)output of with subtitle video



PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

```
PS C:\Users\Golu\OneDrive\Desktop\Engineering Academics\SECOND SEMISTER ACEDEMICS\HARDWARE DISCRIPTIVE LANGUAGE\project> python -u "c:\Users\Golu\OneDrive\Desktop\Engineering Academics\SECOND SEMISTER ACEDEMICS\HARDWARE DISCRIPTIVE LANGUAGE\project\final_project.py"
MoviePy - Writing audio in C://Users//Golu//OneDrive//Desktop//hdl voice//download npt.wav
MoviePy - Done.
Video to audio conversion complete!
Converted Text:
social media speech by smile please world good morning to the respected principal teachers parents and my dear friends today I am here to share my opinion a
bout social media social media is an online platform where we can share your thoughts information pictures and videos we or use it in our day to day life it
has become an integral part of our lives there are various social media platforms like Facebook Twitter WhatsApp accept Ram we all use them to communicate
and socialize locally as well as Global social media has both advantages and disadvantages I would like to mention advantages of social media first the bigg
est benefit of social media is that we can share your thoughts and ideas with people across the world now even a common man can express his opinion on any n
ational or International issue another benefit of social media is that people can showcase their talent on this platform many people have gained Fame by sho
wing that Ireland through social media it is also helping people to earn money people sell a lot of things like clothes books electronic items except throug
h social media social media is also used to create awareness amongst people people with good knowledge about topics like help or nutrition use this platform
to create awareness amongst people now I would like to talk about the drawbacks of social media addiction is the biggest drawback of social media people ha
ve got used to of checking their social media accounts every now and then they spend a lot of time in checking various post on social media spending a lot o
f time on social media also brings a lot of health disorders like stress anxiety and sleeplessness another drawback of social media is that we cannot trust
all the information available here some information can be true and some can be fake at times false news spread like Wildfire through this platform our pers
onal information becomes easily available too many on the net hence we need to stay a lot from hackers apart from this cyber bullying is another major drawb
ack of social media some people personality and Heath comments on others accounts and harass them social media has both advantages and disadvantages the adv
antage like socializing has made our life simpler and convenient there are some disadvantages to we need to use this platform wisely and in a limited way so
that we can make the most out of this new platform with this I and my speech thank you everyone have a wonderful day
PS C:\Users\Golu\OneDrive\Desktop\Engineering Academics\SECOND SEMISTER ACEDEMICS\HARDWARE DISCRIPTIVE LANGUAGE\project>
```

if you want to change the world startoff by making your bed if you make yourbed every morning you will haveaccomplished the first task of the dayit will give you a small sense of prideand it will encourage you to do another task and another and another and by theend of the day that one task completedwill have turned into mini taskcompleted making your bed will also reinforce the fact that the littlethings in life matter if you can't dothe little things right you'll never beable to do the big things right and ifby chance you have a miserable day youwill come home to a bed that is madethat you made and a made bed gives youencouragement that tomorrow will bebetter I've been a Navy SEAL for 36yearsevery morning in SEAL training myinstructors who at the time were allVietnam veterans would show up in mybarracks room and the first thing theydo is inspect my bed if you did it rightthe corners would be square the coverswould be pulled tight the pillowscentered just under the headboard andthe extra blanket folded neatly at thefoot of the rack it was a simple taskmundane at best but every morning wewere required to make our bed toperfection it seemed a little ridiculousat the time particularly in light of thefact that we were aspiring to be realwarriors tough battle-hardened sealsbut the wisdom of this simple act hasbeen proven to me many times over[Music]it matters not whether you ever servethe day in uniform it matters not yourgender your ethnic or religiousbackground your orientation or yoursocial status our struggles in thisworld are similar and the lessons toovercome those struggles and to moveforward changing ourselves and changingthe world around us will apply equallyto all if you think it's hard to changethe lives of ten people change theirlives forever you're wrongI saw it happen every day in Iraq andAfghanistan but changing the world canhappen anywhere and anyone can do it sowhat starts here can indeed change theworld and you will likely fail often itwill be painful it will be discouragingat times it will test you to your verycore at that darkest moment of themission is a time when you need to becalm when you must be calm when you mustbe composed when all your tacticalskills your physical power and yourinner strength must be brought to bearif you want to change the world you mustbe your very best in the darkest momentsif I have learned anything in my timetraveling the world it is the power ofhope the power of one person theWashington a Lincoln King Mandela andeven a young girl from Pakistan Malalaone person can change the world bygiving people hopestart each day with a task completedfind someone to help you through liferespect everyone know that life is notfair that you will fail often but if youtake some risks step up on the timesyou're the toughest face down thebullies lift up the downtrodden andnever ever give up if you do thesethings the next generation and thegenerations that follow will live in aworld far better than the one we havetoday and what started here will indeedhave changed the world for the betterfinally a seal training there's a bell abrass bell that hangs in the center ofthe compound for all the students to seeall you have to do to quit is ring thebell ring the bell and you no longerhave to wake up at 5 o'clock ring thebell and you no longer have to be in thefreezing cold swims ring the bell andyou no longer have to do the runs theobstacle course the PT and you no longerhave to endure the hardships of trainingall you have to do is ring the bell toget out[Music]if you want to change the world don'tever ever ring the bell[Music][Music]



## VII. CONCLUSION

in this project we discussed our proposed method of detecting and extracting text from the youtube video. The system automates the manual process of extracting text from videos and hence is economical in terms of time and human efforts. The project implemented in Python language. The project can be used for educational and news videos which contain information in form of text.

## VIII. ACKNOWLEDGMENT

We would like to start by sincerely thanking our project guide Prof. Mr. Suresh chavan for his suggestions in choosing the right topic, constant and timely advices and useful inputs in making our project “Youtube Transcript” a great success. I appreciates their moral support and timely contribution made the journey fruitful. Lastly, special appreciation for our team members for their complete coordination, honest efforts and willingness to complete the project.

## REFERENCES

- [1] I. Awasthi, K. Gupta, P. S. Bhogal, S. S. Anand and P. K.Soni, "Natural Language Processing (NLP) based Text Summarization - A Survey," 2021 6th International Conference on Inventive
- [2] Journal of King Saud University – Computer and Information Sciences: "Review of automatic text summarization techniques methods" By - Adhika Pramita Widyassari, Supriadi Rustad, Guruh Fajar Shidik. Edi Noersasongko Abdul Syukur, Affandy Affandy", De Rosal IgnatiusMoses Setiadi" Faculty of Computer Science, Dian Nuswantoro University, Semarang, Indonesia Department of Electrical Engineering, STT Ronggolawe Cepu, Blora, Indonesia doi:
- [3] P. R. Dedhia, H. P. Pachgade, A. P. Malani, N. Raul and M. Naik, "Study on Abstractive Text Summarization Techniques," 2020 International Conference on Emerging Trends in Information Technology and Engineering (icETITE), 2020, pp. 1-8, doi: 10.1109/icETITE47903.2020.087.
- [4] A. Dilawari and M. U. G. Khan, "ASoVS: Abstractive Summarization of Video Sequences," in IEEE Access, vol. 7, pp. 29253-29263, 2019, doi: 10.1109/ACCESS.2019.2902507
- 6. Liang-Tay Lin et al. (2009) 'A new intelligent traffic control system for Taiwan', 2009 9th International Conference on Intelligent Transport Systems Telecommunications, (ITST), Intelligent Transport Systems Telecommunications,(ITST),2009 9th International Conference on, pp. 138–142. doi: 10.1109/ITST.2009.5399369.
- Intelligence Research Advances Volume 1, Issue 2, 2014. [7] Lifang Gu "Text Detection and Extraction in MPEG Video Sequences" In Proceedings of the International Workshop on Content-Based Multimedia Indexing, 2001 pages 233-240.
- [8] Punit Kumar, P. S. Puttaswamy "VIDEO TO FRAME CONVERSION OF TV NEWS VIDEO BY USING MATLAB". IJARSE, Vol. No.3, Issue No.3, March 2014.
- [9] Punit Kumar, P. S. Puttaswamy "Moving text line detection and extraction in TV video frames".IEEE International Advance Computing Conference (IACC) 2015.
- [10] Cerenkov, Z Greece Nikou, C. Pitas, I."Shot detection in video sequences using entropy based metrics" Proceedings. International Conference on Volume: 3 2002.