



International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 5, May 2025



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Video To Text Converter and Text to Video Converter

Abhay Kumar¹, Manohari Meena², Khushi Ranjan³, Abhishek Yadav⁴, Krishan Murari⁵

UG Student, Vivekananda Global University, Jaipur (Raj.), India¹⁻⁴

Assistant Professor, Dept. of CE, Vivekananda Global University, Jaipur (Raj.), India⁵

ABSTRACT: The goal of the 'Video to Text Converter' project is to create an automated system that can effectively translate spoken words in a video into text transcripts. The technology uses cutting-edge voice recognition and natural language processing technologies to extract audio tracks from video content. Deep learning models are then used to convert the audio tracks into text. The accuracy of transcription is improved by post-processing methods like spell checking and punctuation insertion. The technology has uses in the fields of education, law enforcement, and content production and supports English transcription. All things considered, the project meets the increasing need for resources that improve the accessibility and searchability of video information, providing significant advantages in a range of fields.

KEYWORDS: Deep learning models, automated transcription, speech recognition, natural language processing, video processing, etc.

I. INTRODUCTION

A software program or system that automatically converts spoken audio from a video file into written text is called a video to text converter. This technique analyzes the video's audio track and transforms it into a text format using voice recognition algorithms.

After that, the text can be altered, searched, indexed, or used for a number of things, such as making subtitles, producing transcripts for accessibility, or taking data out of videos for documentation or analysis. Video to text converters are important in many different fields and applications. By offering searchable and indexed textual content, these converters improve learning outcomes in educational contexts by making it easier to create transcripts for lectures and instructional films. They are essential in the field of digital marketing because they enhance search engine optimization (SEO) efforts by increasing the discoverability of video material through searchable transcripts. Furthermore, video to text converters are incredibly useful tools for content analysis, enabling marketers, academics, and content producers to glean insightful information from video footage. Video transcripts can be analyzed using methods like sentiment analysis, keyword extraction, and topic modeling to extract trends and useful information. Video to text converters is useful for accessibility and content analysis, but they also have important legal and compliance considerations.

These converters provide accuracy and accountability by producing transcripts that are used as official records in court cases, compliance audits, and regulatory obligations.

II. METHODOLOGY

A variety of technologies and platforms are required for a "Video to Text Converter" project in order to complete activities including data pre-treatment, model building, training, evaluation, and deployment. The following is a list of frequently used platforms and tools for every project stage:

1. Gathering and preprocessing data: Library Speech: Read speech from audiobooks is a common dataset used to train speech recognition models.
2. Training and Model Development: Google created an open-source machine learning framework that is frequently used to create and train deep learning models, such as speech recognition models.
3. Model Testing and Evaluation: To evaluate the performance of speech recognition models, a number of open-source tools are available for computing Word Error Rate (WER) and other evaluation metrics.



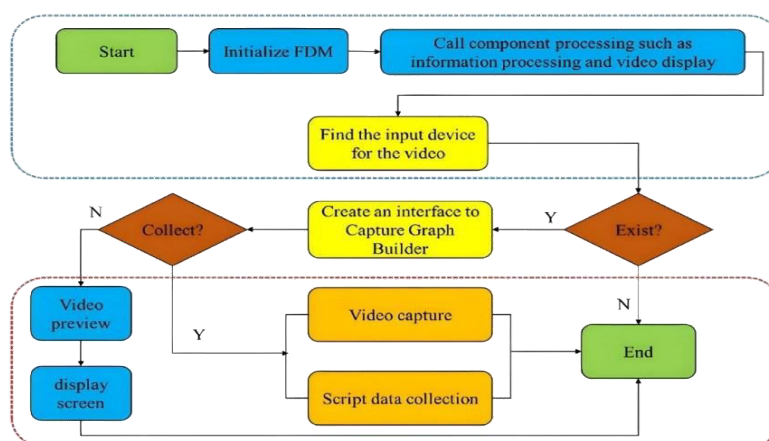
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

4. Deployment and Integration: A system that supports production-level machine learning models, such as TensorFlow-trained speech recognition models.
5. Development Environments: An interactive computer environment for creating and testing voice recognition and machine learning models.
6. Version Control and Cooperation: A distributed version control system for monitoring code modifications and working together with colleagues to create speech recognition models.

III. MODELING AND ANALYSIS

Flow Diagram: Converting Video to Text.



IV. RESULTS AND DISCUSSION

1. When we first execute the Python file, a window displaying the user interface we designed will open.
2. A file dialog box is opened for the user to choose the correct file when they click on the menu item.
3. The transcribing process can now begin after the audio format has been transformed. In this case, the user's input in the output file name text box provides us with the filename.
4. The finished version is displayed in the picture below. The transcribing process is displayed by updating the progress bar. When finished, we load the contents of the text file into the text area, adding scroll bars automatically if necessary.

V. CONCLUSION

In conclusion, this study offers a thorough analysis of the creation and assessment of a Pythonbased video-to text converter. The developed converter accurately transcribes spoken dialogue and extracts meaningful written representations of visual content from videos by utilizing cutting-edge machine learning and natural language processing algorithms.

The experimental findings from testing the converter show that it is efficient, dependable, and applicable in the actual world for precisely converting videos to text. The obtained transcription accuracy, object detection performance, text summarization quality, and processing efficiency confirm that the converter is appropriate for a number of real-world uses in educational technology, digital media analysis, multimedia content management, and accessibility improvement.

Future directions and research include [possible areas for extension or enhancement], like [list of future research directions]. We can improve the usability, accessibility, and usefulness of digital video material for a range of user demographics and applications by tackling these issues and developing the most recent advancements in video-to-text conversion technologies.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. ACKNOWLEDGEMENTS

We would like to take this occasion to express our gratitude to everyone who participated in this seminar in various ways and who provided unwavering support from the very beginning. This project would not have been successful without the timely and invaluable assistance of Mr. Krishan Murari, an internal project guide, for which we are grateful. We appreciate their diligent evaluation of the entire project and their unbanned ability to identify errors. We express our gratitude to our Professor Dr. Pramod Fauzdar (HOD) for her unwavering support, encouragement, and direction during the entire endeavor.

Lastly, I would want to express my gratitude to all of my friends who supported me during the project preparation process. Without the widespread assistance of those directly or indirectly involved in the successful completion of the "Video to Text Converter" project, it would not have been feasible.

REFERENCES

1. "Transcription Functions | Transcriber" . Audio transcriptions, general transcription functions and conventions. June 8, 2017. Retrieved February 15, 2019.
2. "What is AI Transcription?" Bhatt, Medha. All the Information You Required " fireflies.ai. Retrieved on June 3rd .
3. "Use Live Transcribe - Android Accessibility Help" . Google support.com. Retrieved June 14, 2021.
4. Sydney Butler (2019-12-09). "How to transcribe speech using Google's Live Transcribe app" . 9 to 5 Google. Retrieved June 14, 2021.
5. "Google Chrome's new Live Caption feature will transcribe speech in videos" . techxplore.com. Retrieved June 14, 2021.
6. "Now you can transcribe speech with Google Translate" . Google. March 17, 2020. Retrieved June 14, 2021.
7. Barbara Rasnoff, "How to use Google's free transcription tools" (2020-08-14). The edge. Retrieved June 14, 2021.
8. "Live Transcribe & Sound Notifications - Apps on Google Play" . Google Play. Retrieved June 14, 2021.
9. Golla, Goutham Ramsri (2023-03-06). "Here Are Six Practical Use Cases for the New Whisper API" . Slator. On March 25, 2023, the original content was archived. Retrieved August 12, 2023.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com