"Extract the Audio from Video by using python"

Samruddhi Sasavade [1], Trupti Sutar [2], Karan Barale[3], Digvijay Kambale[4]

[1,2,3,4]*Students of*
Department of Computer Science & Engineering,
Sanjeevan Engineering and Technology Institute, Panhala - 416201

---***---

Abstract - *A video to audio converter is a programme that converts video to audio and video to mp3. In this application, you must first select a file and then press the convert button once you have finished selecting your video file. This video to audio converter tool will then begin converting video to audio, and your audio file will be saved in the same folder. Let's create a Python Video to Audio Converter project.*

*Key Words***:** Python- moviepy, tkinter

1 INTRODUCTION

Since the emergence of social media sites like YouTube and Instagram, video content has grown in popularity. However, there are times when we prefer to just listen to a video's audio rather than watch it. A video to audio converter is useful in this situation. A software programme known as a "video to audio converter" may extract audio from a video file and store it in an audio file format like MP3 or WAV. By doing this, users can listen to their favourite podcasts or music without having to see the accompanying video. There are several popular multimedia formats available today, and each one is fortunate to have a specific application (Video Home System, Internet Video Streaming, Video Editing, etc.) or was invented specifically for portable multimedia devices. Some programmes and gadgets can only accept formats that were specifically designed for them. It is well known that there are numerous multimedia devices (such as DVD players, mobile phones, iPods, PSPs, Zune, PlayStations, etc.) and various audio/video data carriers (such as DVDs, the Internet, Blue Ray, HDDs, and Flash USB storage), making it relatively simple for even seasoned users to encounter difficulties when selecting the best one. Because virtually all devices are produced by universal brands like Microsoft, Apple, and Sony, which can be considered competitors, it appears to be extremely complicated. Additionally, the products that these companies are promoting occasionally appear to be incompatible not only with PC but also with one another. These limitations are brought about by the development of unique video formats for specific devices. The main difference between the video and audio is The video is available in a number of file formats, including Mp4, MOV, and WEBM. The audio and video in these files are synced. Mp3 files, on the other hand, just have audio and no visuals. As a result, mp3 files consume far less space on storage devices like smart phones, online storage platforms, and hard discs. Additionally, certain devices can only play and store MP3 files. The internet is rife with videos that we might enjoy watching while working out, exercising, or doing housework. Music with music videos, podcasts with video segments, interviews, and motivational speeches are just a few examples. We don't know about you, but at least one of our friends has a phone that is overflowing on any given day. As was already mentioned, mp3 files take up less space, so if you remove any superfluous videos, you can fit more songs, images, or films on your devices. You'd be shocked at how much space a few videos downloaded to your phone, computer, or external storage device will use. Video files can grow rather large.

Using basic operations (such cuts, concatenations, and title insertions), video compositing (also known as non-linear editing), video processing, or complex effects, MoviePy is a Python module for video editing. The majority of popular video formats, including GIF, can be read and written.

1.1 Problem Statement

There are several systems out there that only translate audio or video into text. There are numerous live transcribing programmes for various transcribing tasks that we see, use, and are present in our regular job and professional profiles. Since this system only uses audio transcription, there is a chance that occasionally the pronunciation will be incorrect.The accuracy issue arises throughout the video to text conversion process when we attempt to obtain audio accuracy in order to provide the desired output.In contrast, a video plays frame by frame, which takes time and results in a slow output.

1.2 Objective

Using this application, we may get notes from audio and video sources and solve the problem of documentation. In order to extract the text that is present in each frame and add it to the output textbox, we are utilising the open CV programme with the provided input video as a source to determine the frame rate. The audio file's speech recognition technology.

2 Literature Survey

In this study, we present a novel technique for automatically aligning and evaluating audio-visual text for music video summarization. The music video separates the video track from the music track. The chorus can be located in the musical track, according to an analysis of the music's structure. We initially divide the photos and group them into close-up face shots and non-facial shots before extracting the phrases and determining the most often repeated lyrics from the shots for the video track. The phrases from the music video that are most frequently repeated, the shot type, and the alignment of chorus boundaries are used to create the synopsis. Studies for chorus detection, shot categorization, and lyrics detection using 20 English music videos are provided. individual user research. Studies for chorus detection, shot categorization, and lyrics detection using 20 English music videos are provided. The effectiveness and calibre of summary have been assessed by a series of subjective user studies.

Video and audio data are gathered across a broad range of issue disciplines in order to document processes, procedures, or encounters. These video and audio data are afterwards investigated using a number of approaches to try to understand what was happening at the time of recording; sometimes in relation to initial hypotheses and other times in terms of a "post hoc" analysis utilising a more grounded approach. This paper introduces the tools and methodology for analysing video data and explores prospective innovations in discourse analysis that can be inspired by learning analytics.

In this third paper voice cloning techniques have been used in a range of applications, such as video games and customised speech interfaces for marketing. The most sophisticated voice cloning technology may produce perceptually unidentifiable speech by learning speech features from a small sample of speech. These systems pose new security and privacy risks to voice-driven interfaces. Since fake audio has been used with malicious intent, it can be difficult to tell the difference between real and fake audio during a digital forensic analysis. This paper looks at the issue of deep-fake audio categorization and evaluates the usefulness of forensic deep-fake audio detection methods.

In the real world, where the main workplace challenges are solved, there is a need for a video and audio to text converter. This converter can be used for documentation by many software businesses, academic institutions, and other organisations. This is generally used by software businesses to access notes, project data, project presentation materials, etc. We chose to use Google Speech Recognition for our system's UI because it is so reliable and user-friendly. Python is even our choice due of how easy it is to learn. If there is only one audio file on the system, the audio to text converter should be used to make it easier for the user to convert the audio to text.

3 Process of video to audio converter

Video Converter is a multipurpose program specially created for working with video files. The program allows you to perform practically all the operations with video.

1) Enhanced system of managing the conversion process gives you an opportunity very quickly and qualitatively to convert video from one format into the other; it is important to mention that there are a great variety of supported video formats and multimedia devices. By means of our Video Converter, a user can easily save DVD film from a disc on his/her PC, mobile phone or any multimedia device (iPhone, iPod, PSP, Zune, Creative…), or on the contrary, write the video on a DVD disc. The innovative "2pass encoding" technology enables you to save DVD movies on your computer, having drastically shrunk their size by 5–10 times. What's amazing, though, is that the video quality of the final product is still very high; it has excellent colour rendering with very bright and rich colours, and definitely drowns out small details even in the most dynamic scenes.

2) You can edit videos by adding or removing different pieces (such as titles or commercials), save audio tracks from movies in MP3 formats, and do a variety of other tasks with the aid of a professional "Video Editor."

3) The "Downloader" application allows you to browse or just download your favorite video clips from the most widely used video hosting sites on the Internet, including YouTube, MySpace, and AOL.

4) The built-in "Disc Writer" will assist you in saving the video clips to a disc for later viewing, such as on a DVD player.

5) A panel called "Profile Editor" that was built specifically for professional users allows you to manually select the desired features of the finished video.(Use in google image)
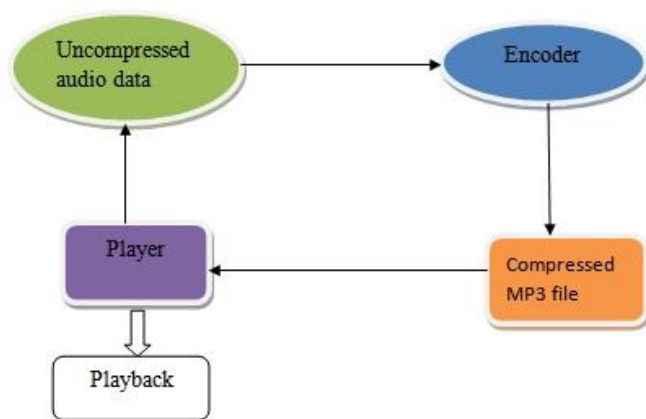
Fig -1: The Process of Video to Audio Converter

3.1 Ways of video to audio converter:

A video file can be turned into an audio file in a variety of ways. Here are some techniques you can use:

1. Online video to audio converters: These services are available on a wide variety of websites. You can find a variety of solutions by searching for "online video to audio converter" in your favorite search engine. Typically, these services let you upload your video file and convert it to several audio formats.

2 .Desktop Software: A variety of desktop software applications are available that enable audio to video conversion. VLC Media Player, Audacity, and Freemake Video Converter are a few well-liked choices. These programmes frequently support a large number of video and audio formats and offer an intuitive user interface.

3. Using FFmpeg (Command Line): For managing multimedia files, FFmpeg is a potent command-line application. You can download FFmpeg and use it to convert a movie to an audio file if you're comfortable using command-line tools.

3.2 Recording Storing and Sharing video :

A wide range of devices, including dedicated video cameras (at varied price points), digital cameras, and popular handheld devices like smartphones and other mobile phones, are available to conveniently and inexpensively collect video data. The Diver project (Digital Interactive Video Exploration & Reflection, used a set of 5 cameras to acquire a 360-degree record of activity) is one example of how head-mounted cameras might be used in fieldwork situations or for documenting surgical operations. In addition, a variety of websites and smartphone "apps" specifically designed to enable the creation, hosting, and sharing of videos have appeared in recent years (such as YouTube, Tumblr, etc.). Many of them are associated with the "social media" phenomena, in which individuals produce and distribute multimedia for use in their personal lives or

for entertainment. Although the focus of this study is not on these resources, they show how pervasive the notion of video as a publicly-created and shared communication medium has grown. Additionally, it demonstrates how simple it is to make and share video utilising a variety of devices and thus, a variety of platforms. It is hoped that the general public's increased familiarity with the concept of recording and sharing video will promote greater acceptance and more naturalistic settings where video is recorded by researchers and used in studies where those subjects in the video are the main points of interest. In any case, video diaries and video-recorded observations/user trials have a long history in many disciplinary research areas, so it is hoped that a wide audience of academics and researchers who regularly record and analyse video data will find this paper to be of interest.

3.3 Steps to Extract Audio from Video using Python with MoviePy library

It will be difficult to handle the video in its raw binary format, so we will use an external library named moviepy for this task. MoviePy works with Python versions 2.7 and up and supports a wide range of operating systems, including Windows, Linux, and macOS. It can read and write all of the most popular audio and video formats, including GIF.

1. Setup the moviepy library

Since this library is not built-in, we must install it. To achieve this, we will utilize the package manager pip. Open a terminal and type the following command to install the moviepy Library.

pip install moviepy

2. Add the moviepy library

Let's begin by importing the libraries right away. The moviepy package or the editor class only must be imported specifically.
# Import moviepy
import moviepy.editor
3. Upload the movie
Let's learn about video formats before moving on to the next level. To perform the conversion without any issues, we must be aware of the format of our video. The most popular video formats include WMV (wmv, wma, asf*), OGG (ogg, oga, ogv, ogx), 3GP (3gp, 3gp2, 3g2, 3gpp, 3gpp2) and MP4 (mp4, m4a, m4v, f4v, f4a, m4b, m4r, f4b, mov).
By addressing the video file via the argument, we must create a VideoFileClip object.
#Load the Video Clip
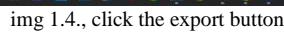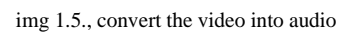video = moviepy.editor.VideoFileClip("Video.mp4")
Inside the method, we are adding the path of the video. The video we are going to convert is in mp4 format.
4. Remove the audio.
Knowing a few audio formats is also a good idea in addition to the video format. These popular file types include MP3, AAC, WMA, and AC3. We now have some thoughts on both

forms. Now is the time to use the MoviePy library to perform the conversion. The audio from the video that we defined will now be extracted. Accessing the audio component of the VideoFileClip object we constructed will allow you to easily extract the audio. The extracted audio must then be written into a new file with a specific filename.

#Extract and export the Audio
video.audio.write_audiofile("Audio.mp3")

3.4 Tkinter

The term "graphical user interface" (GUI) refers to a type of user interface that enables people to communicate with computers visually through the use of objects like windows, menus, and icons. It has advantages over the Command Line Interface (CLI), which is more difficult to use than GUI and requires users to type instructions into computers exclusively using the keyboard. The built-in Python module

Tkinter is used to develop GUI applications. Given that it is straightforward and quick to use, it is one of the most often used modules for developing GUI applications in Python. Since Tkinter is already included with Python, you don't need to worry about installing it separately. It provides the Tk GUI toolkit with an object-oriented interface.

Tkinter can be used to construct windows and dialogue boxes, which enable users to interact with your programme. These can be used to show data, collect feedback, or provide the user options. Making a GUI for a desktop application: The interface for a desktop application, which includes buttons, menus, and other interactive elements, can be made using Tkinter.

A command-line programme can have a GUI added to it using Tkinter, which makes it simpler for users to interact with the programme and enter arguments.

Tkinter allows you to build your own custom widgets in addition to a wide range of built-in widgets like buttons, labels, and text boxes.

A GUI prototype can be easily created using Tkinter, allowing you to test and iterate on various design concepts before committing to a full implementation. The bottom line is that Tkinter is a helpful tool for developing a wide range of graphical user interfaces, including windows, dialogue boxes, and unique widgets. It is especially suitable for creating desktop apps and giving command-line programmes a GUI.

4.Architecture

Let's look at how to build a video to audio converter using the description of implementation modules is

1.Import each of the required modules first.

2.Make a window for our converter's user interface.

3.Create a function that processes the file and converts the video to audio after the user selects the file.

4.The file will be saved automatically in the same folder.

5. Snapshot of the project

Img 1.1., open the window for video to audio converter

img 1.2., select the video from this window

img 1.3., Select audio format

img 1.5., convert the video into audio

6. CONCLUSIONS

This programme provides an incredibly basic, hardly usable sample. This might be enhanced in a number of ways, such as by allowing users to load either an mp4 or a wav file, by letting them choose from a range of speech recognizers, and by displaying more information such as file size and length. Graphical user interfaces can improve the user experience of straightforward programmes, and they are easy to implement in Python using programmes like MoviePy and Tkinter Designer. To minimise their negative effects on the user experience, performance concerns can also be fixed utilising threading and conducting computationally demanding operations in the background.

7. REFERENCES

[1]. Marcello Federico, Robert Enyedi, Roberto Barra-Chicote, Ritwik Giri, Umut Isik, Arvindh Krishnaswamy,Hassan Sawaf" From Speech-toSpeech Translation to Automatic Dubbing" Proceedings of the 17th International Conference on Spoken Language Translation July 2020

[2]. GOPA - International Energy Consultant INTEC & Hamm-Lippstadt University of Applied Sciences 2022

[3]. Published in: IEEE Journal of Selected Topics in Signal Processing (Volume: 16, Issue: 6, October 2022)

[4]. Published in: IEEE Transactions on Software Engineering (Volume: 48, Issue: 1, 01 January 2022)

[5]. J. Pradeep, E. Srinivasan, S. Himavathi, Neural network based handwritten character recognition system without feature extraction, in 2011 International Conference on Computer, Communication and Electrical Technology (ICCCET), pp. 40–44 (2011).

img 1.4., click the export button

img 1.4 save the audio file name

[6]. R. Mittal, A. Garg, Text extraction using OCR: A systematic review, in 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 357–362 (2020).

[7] Brantley-Dias, L., Dias, M., Frisch, J.K. and Rushton, G., The Role of Digital Video and Critical Incident Analysis In Learning to Teach Science. Proceedings of the American Educational Research Association Annual Meeting, (New York City, New York, 2008), 2008.

[8] Brundell, P., Tennent, P., Greenhalgh, C., Knight, D., Crabtree, A., O'Malley, C., Ainsworth, S., Clarke, D., Carter, R. and Adolphs, S., Digital Replay System (DRS) - a tool for interaction analysis. Proceedings of the 2008 International Conference on Learning Sciences (Workshop on Interaction Analysis), (Utrecht, 2008), 2008.

[9] Carroll, J.M., Koenemann-Belliveau, J., Rosson, M.B. and Singley, M.K., Critical incidents and critical themes in empirical usability evaluation. Proceedings of the HCI-93 Conference: People and Computers VIII (British Computer Society Conference Series), (1993), Cambridge, U.K.: Cambridge University Press, 279-292, 1993.

[10] De Liddo, A., Sándor, Á. and Buckingham Shum, S. Contested Collective Intelligence: rationale, technologies, and a human-machine annotation study. Computer Supported Cooperative Work (CSCW) (in press), 2012.

[11] Erickson, F. Definition and analysis of data from videotape: Some research procedures and their rationales. In J.L. Green, G. Camilli and P.B. Elmore, editors, Handbook of complementary methods in education research. Erlbaum, Mahwah, NJ, 2006, 177-205.

[12] Ferguson, R. and Buckingham Shum, S., Learning analytics to identify exploratory dialogue within synchronous text chat. Proceedings of the 1st International Conference on Learning Analytics & Knowledge (LAK 2011), (Banff, Alberta, 2011), 2011.

[13] Flannigan, J.C. The critical incident technique. Psychological Bulletin, 51 (4), 327-358, 1954.

[14] Killion, J.P. and Todnem, G.R. Capturing complexity: A typology of reflective practice for teacher education. Teaching and Teacher Education, 18, 73-85, 2002.

[15] Mercer, N. Sociocultural discourse analysis: analysing classroom talk as a social mode of thinking. Journal of Applied Linguistics, 1 (2), 137-168, 2004.