Holy-wood Academy, Kolhapur.

SANJEEVAN ENGINEERING AND TECHNOLOGY INSTITUTE.

Sanjeevan Knowledge City, Somwar Peth, Panhala, Kolhapur - 416201

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

A

MINI PROJECT REPORT

ON

"*Video to Audio Converter*"

Submitted in the partial fulfillment of the requirements for the award of degree of

Bachelor of Technology (T.Y.B. Tech)

IN

COMPUTER SCIENCE AND ENGINEERING

Of

Dr. Babasaheb Ambedkar Technological University
डॉ. बाबासाहेब आंबेडकर तंत्रशास्त्र विद्यापीठ

SUBMITTED BY

Project Group ID:

Sr. No	Student Name	Roll No.	PRN
1	Digvijay Namdev Kambale	47	2063151242011
2	Samruddhi Sudhir Sasavade	48	2063151242030
3	Trupti Tanaji Sutar	49	2063151242044
4	Karan Kundalik Barale	50	2163151242552

UNDER THE GUIDANCE OF

Prof. V.S.PAWAR

Department of Computer Science &Engineering
Sanjeevan Engineering and Technology Institute, Panhala

Academic Year: 2022-2023

Holy-wood Academy, Kolhapur.

SANJEEVAN ENGINEERING AND TECHNOLOGY INSTITUTE.

Sanjeevan Knowledge City, Somwar Peth, Panhala, Kolhapur - 416201

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that Project entitled **"Video to Audio Converter"** is a bonafied Mini Project work carried out by:

Sr. No	Student Name	Roll No.	PRN
1	Digvijay Namdev Kambale	47	2063151242011
2	Samruddhi Sudhir Sasavade	48	2063151242030
3	Trupti Tanaji Sutar	49	2063151242044
4	Karan Kundalik Barale	50	2163151242552

in partial fulfillment of the requirements for the award of ***Bachelor of Technology (B.Tech)in Computer Science and Engineering*** by **Dr. Babasaheb Ambedkar Technological University, Lonere** for the academic year 2022-23. It is certified that all the corrections / suggestions indicated for internal assessment have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of ***Mini Project work*** prescribed for the said Degree.

Guide	HOD	Principal
Prof. V.S. Pawar	**Prof. R. S. Nejkar**	**Dr. S. N. Jain**

Name of the Examiners:	External Viva	Signature with Date
1		

2.

DECLARATION

We here by declare that this project work entitled “Video to audio Converter” has been prepared by us during the year 2022-23 under the guidance of Prof V. S. Pawar, Department of Computer Science, Sanjeevan Engineering and Technology Institute, Panhala in the partial fulfilment of the T.Y. B. Tech . We also declare that this project is the outcome of our effort, that it has not been submitted to any other university for the award of any degree.

ACKNOWLEDGEMENT

It gives me an immense pleasure to present a report on the successful completion of our mini project work on **"Video to audio Converter"** We express our deep sense of gratitude to our guide **"Asst. prof. V. S. Pawar"** for his valuable guidance rendered in all phases of seminar. We are thankful of his wholehearted assistance, advice and expert guidance towards making my seminar success.

My special thanks to respected Principal and Head of the Department for their keep interest, encourage and excellent support.

Sr. No	Student Name	Roll No.	PRN
1	Digvijay Namdev Kambale	47	2063151242011
2	Samruddhi Sudhir Sasavade	48	2063151242030
3	Trupti Tanaji Sutar	49	2063151242044
4	Karan Kundalik Barale	50	2163151242552

ABSTRACT

In the real world, where the biggest workplace issues are resolved, the necessity for a video and audio to text converter exists. This converter can be used for documentation reasons by a wide range of software firms, educational institutions, and other organizations. This is mostly used by software businesses to access notes, project details, project presentation materials, etc. We choose Google Speech Recognition for our system due to its superior accuracy and user-friendly interface. Python was even chosen by us for its ease of learning. When there is only one audio file in the system, the audio to text converter should be used to make it simple for the user to convert the audio to text.

INDEX

Mr. V. S. PAWAR
Project Coordinator

Mr. R. S. Nejkar
H.O.D(CSE)

1.Introduction

Now a days the online mode or work in online has the major part in the educational departments, jobs, and much more. As the pandemic occurred few year where the world has used to sit in their respective indoor with the personal computers, mobile devices, laptops advocate, etc.If we interview participant for various research projects. Often grant funding for such projects will cover transcription costs. Human transcribers remain the work and usually do an excellent job. You could of course transcribe your own interviews, works but this can be a very time consuming and laborious task. Some qualitative researchers also transcribing your own work as away of becoming more familiar with the data. We will start by converting the input video and audio file which we can convert it into text. Video content has become increasingly popular in recent times, with the rise of social media platforms such as YouTube and TikTok. However, sometimes we may only want to listen to the audio portion of a video rather than watching it. This is where video to audio converter comes in handy. A video to audio converter is a software tool that can extract audio from a video file and save it as an audio file format such as MP3 or WAV. This allows users to enjoy their favorite music or podcast without having to watch the video component. Those of us who love music can sometimes end up in a rabbit hole of watching pared-down performances, live recordings, or special channel recordings like Watch Listen Tell, which invites artists to perform acoustic versions of their songs in unusual public locations (creating some fascinating songs). In many instances, these unique performances were not recorded and released on music apps or platforms, meaning the only place we can find the song is in one specific video. For those of us who would love to extract songs from performances like this, a tool that handles MP4 to MP# conversion is the perfect solution. This means any strange video of an obscure musician can be saved and listened to even if you're away from a screen. There are some fantastic interviews, speeches, and podcast episodes to be found on YouTube and other video platforms. This being said, sometimes the best things are a few hours long, and we'd like to continue listening to them while we're away from the internet (or just trying to save data because our phone plan is a joke). Converting an online video into an audio file can allow you to take the file away from internet access and listen whenever and wherever you want. The above situations are just a few of the instances where you might want to convert a video file into an audio file. Once you start converting files occasionally, you'll find a whole bunch of other reasons to save up interesting or entertaining audio files. Those who enjoy making videos are always looking for new sound effects. Same with those who like pranking people. Maybe you simply want the most unique ringtone possible.

Whatever the reason you're in need of sound effects, you can extract whatever sounds they can find in video format (like exotic birds, race cars, foreign cities, etc.) and apply them to whichever video projects or funny moments they want. Similarly, you might want to collect classic movie lines, political quotes, meme-worthy interview lines, or any other tidbits of audio for serious or comical videos. If that's something that appeals to you, you're going to want to extract the audio portion of the video file so that you can include it with your own visuals or the project you're working on

What is the difference between the video and audio files

The video comes in a variety of file types including, Mp4, MOV, or WEBM. These files contain synchronized video and audio. Mp3 files, in contrast, contain only audio, no visual component at all. Because of this, mp3 take up far less space on storage devices like our phones, cloud storage services, or hard drives. As well, some devices only store and play mp3 files

The internet is full of videos that we might like to listen to on our commute while working out or cleaning the house—music with music videos, podcasts with video components, interviews, motivational speeches, just to name a few. Converting a video file into an MP3 allows you to listen to the audio without having the video element taking up unnecessary space on your device or distracting you while your eyes need to be elsewhere.

Save Space

We don't know about you, but at least one person in our friend group on any given day has a phone that's too full. As mentioned above, mp3 files take up less space which means you can fit more songs or photos, or videos on your devices if you get rid of unnecessary videos. You'd be surprised how much space will be taken up if you download a few videos onto your phone or computer or external storage device. Video files can get pretty big.

2. Problem Statement

2.1 Literature Survey

1. **Automatic music video summarization based on audio-visual-text analysis and alignment [Changsheng Xu , Xi Shao]:-** In this research, we offer an innovative method for aligning and analysing audio-visual-text for automatic music video summarization. The music track and the video track are split in the music video. Based on an analysis of the music's structure, the chorus is found in the musical track. Before extracting the phrases and identifying the most often repeated lyrics from the shots for the video track, we first split the shots and categorise them into close-up face shots and non-facial shots. The summary of the music video is produced using the music video's most frequently repeated lyrics, shot class, and the alignment of chorus boundaries. The studies employing 20 English music videos for chorus detection, shot classification, and lyrics detection are given. Personal user studies. The studies employing 20 English music videos for chorus detection, shot classification, and lyrics detection are given. Subjective user studies have been conducted to evaluate the quality and effectiveness of summary.
2. **Analysing video and audio data: existing approaches and new innovations [Elizabeth FitzGerald]:-** In order to record processes, procedures, or interactions, video and audio data are collected across a wide range of topic disciplines. In order to try and understand what was happening at the time of recording, these video and audio data are afterwards studied using a variety of methodologies, sometimes in connection to original hypotheses and other times in terms of a "post hoc" study utilising a more grounded approach. This study provides an introduction of video data analysis tools and methodologies and considers potential new approaches to discourse analysis that could be adapted from learning analytics.
3. **The Effect of Deep Learning Methods on Deepfake Audio Detection for Digital Investigation[Mvelo Mcubaa , Avinash Singh]:-** 1. A variety of applications, including video games and personalised speech interfaces for marketing, have exploited voice cloning techniques. The most advanced voice cloning technologies can learn speech features from a few samples and generate perceptually unidentifiable speech. Voice-driven interfaces now face additional security and privacy threats from these systems. It can be challenging to distinguish between actual and phoney audio during a digital forensic examination since

fake audio has been utilised for harmful intent. This study examines the problem of deep-fake audio classification and assesses the effectiveness of the available deep-fake audio detection techniques for forensic analysis.

4. **Implementation of Video and Audio to Text Converter[Dr. M. Saraswathi , VVSV Ronit , S Sai Pranav]:-** The need for a video and audio to text converter is present in the real world, where the major workplace difficulties are resolved. Numerous software companies, academic institutions, and other organisations may use this converter for documentation purposes. Software companies primarily use this to access notes, project information, project presentation materials, etc. Because Google Speech Recognition is so accurate and user-friendly, we decided to adopt it for our system. interface. We even choose Python because of how simple it is to learn. The audio to text converter should be utilised if there is only one audio file in the system to make it simple for the user to convert the audio to text.

2.3 Problem Definition

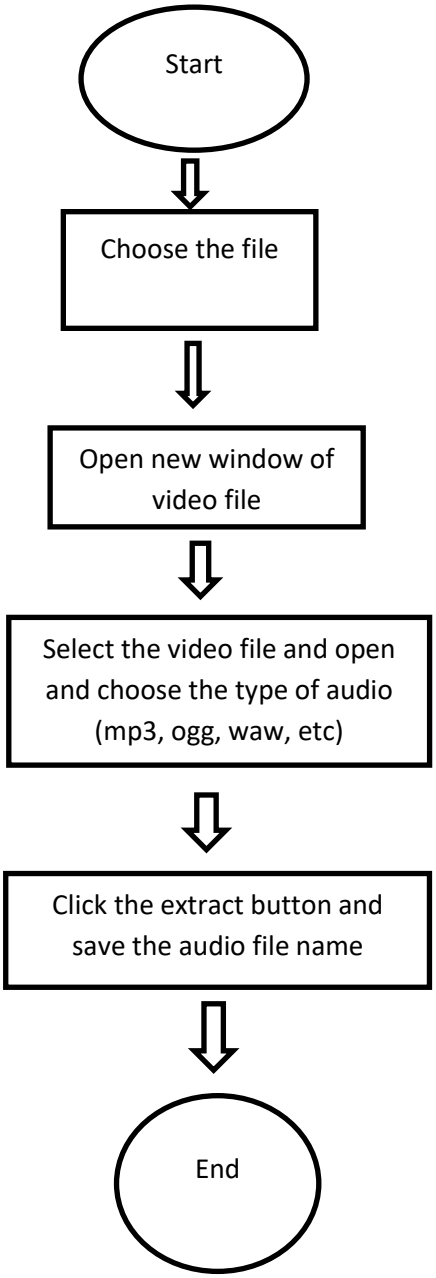
There are many systems like which just converts the audio neither the video into the text. For many transcribing work there are many live transcribing applications where we see, use and exists in the day to day work and work profiles. As there is the problem in this existing system some time it may goes the pronunciation which is wrong and is done with only the audio transcribing.The accuracy comes the problem when the video to text converting part, where we try to get the accuracy for the audio and its done with the thing to get the output .Whereas the video goes frame by frame in which it takes the time to get the output slow.

2.4 Objective

By using this application we can ease the documentation problem and get the notes from the audio and video files. We are using open CV with the given input video as a source to get the frame rate and pytesseract module to extract the text present in that frame and then adding it to the output textbox. Speech recognition for the audio file.

3. System Design and Requirement Specification Analysis

3.1 System Architecture



3.2 Modules

In this system, we have developed few modules such as

- Video to audio conversion
- Speech recognition

- **Video to audio conversion:**

  In this we are using the moviepy package in python in order to extract the audio of the uploaded video file and save it to a audio.wav file. We are using “vidfile.audio.write_audiofile("audio.wav")” to write the audio of the given video file into a file named audio.wav and saving it the parent directory for further use.

- **Speech recognition:**

  We are using the Speech Recognition module in python with the extracted audio.wav file as the source and using google speech recognition to extract the text from the audio file and return it as a stringGoogle speech recognition

3.3 System Requirements

- Supported OS: Windows 11/10/8.1/8/7(32&64bits)
- 1GHz Intel/AMD processor or above.
- RAM:256MB RAM (512MB or above recommended)
- Free Hard Disk: 70MB for installation.

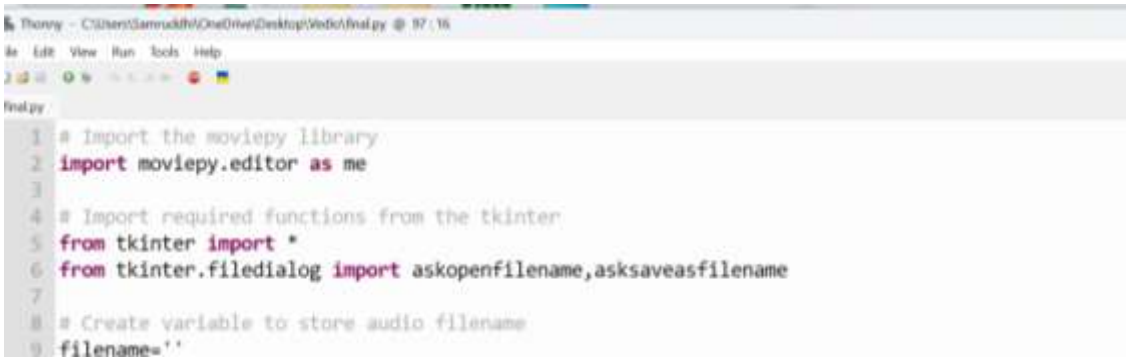
4.Implementation

4.1 Implementation Modules Description

- Let's see the step to Build Video to Audio Converter
- First import all the necessary modules.
- Create a user interface window for our converter.
- After the user selects the file, create a function that will process the file and convert the video into audio.
- File will automatically get saved in the same folder.

4.1.1 Project file structure

- **Importing the file:-** We will first import all the necessary libraries required for this project.



- **Creating Display Window:-** First, create a main class name as VideoAudioConverter, pass the root in an initiator function in which we will set the title, frames, background images and buttons for the interface.
- **Defining Function:-** In this part we'll discuss all the necessary functions which are required to complete our projects. We'll see the use of each function separately and how it'll work.

  Browse():-This function will browse the video files from the user and then process it using the convert function to convert the format of a file i.e., from .mp4 to .mp3. After converting the file it will show completed on the user interface.

  Convert():- This function will take the path of a file as an input and with the help of the VideoFileClip function of moviepy library it will read the video file and with the audio.write_audiofile function it will convert the video file into an audio file.

- **Creating Main Function:-** In the final step, we'll create a main function. In this function we declare the root and also create an object for our main class. Then we'll call this main function to execute our project.

4.2 Software Platforms, Languages &Tools Used

Software platform:- Thonny

An integrated development environment (IDE) facilitates computer programmers by integrating fundamental tools (e.g., code editor, compiler, and debugger) into a single software package. Users do not need to install the language's compiler/interpreter on their machines; an IDE provides the environment itself**.** Thonny is a free, dedicated IDE for Python designed for beginners

Features of Thonny:-

The following are some of the primary features of Thonny:

- It autocompletes code.
- It inspects code to provide bracket matching and highlight errors.
- It is easy to start with as its installer also installs Python 3.7.
- Its debugger is simple to use as no knowledge of breakpoints is required.
- It enables users to step into a function call by providing details about local variables and displaying the code pointer.
- It has an easy interface to install packages. This makes it very suitable for beginners.

Language:- Python

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics developed by Guido van Rossum. It was originally released in 1991. Designed to be easy as well as fun, the name "Python" is a nod to the British comedy group Monty Python. Python has a reputation as a beginner-friendly language, replacing Java as the most widely used introductory language because it handles much of

the complexity for the user, allowing beginners to focus on fully grasping programming concepts rather than minute details. Python is used for server-side web development, software development, mathematics, and system scripting, and is popular for Rapid Application Development and as a scripting or glue language to tie existing components because of its high-level, built-in data structures, dynamic typing, and dynamic binding. Program maintenance costs are reduced with Python due to the easily learned syntax and emphasis on readability. Additionally, Python's support of modules and packages facilitates modular programs and reuse of code. Python is an open source community language, so numerous independent programmers are continually building libraries and functionality for it.

To develop this project we need a basic knowledge of some models like tkinter, os, PIL and moviepy.

- **tkinter:-** for use Interface(UI)
- **os:-** provides functions for creating and removing a directory (folder), fetching its contents, changing and identifying the current directory, etc.
- **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.

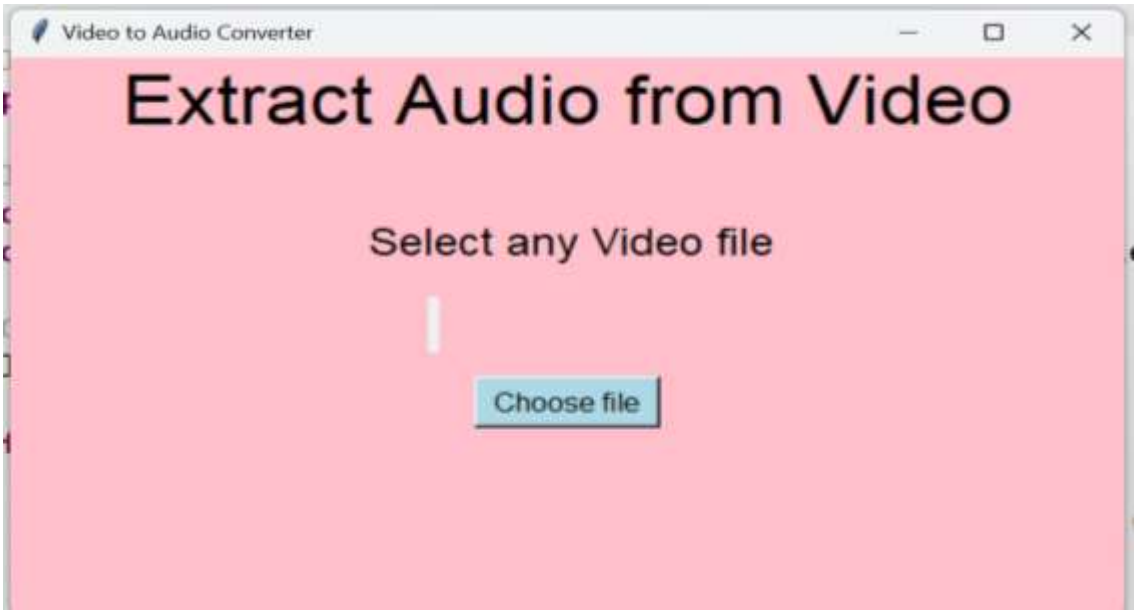
5. Testing

Test Case:

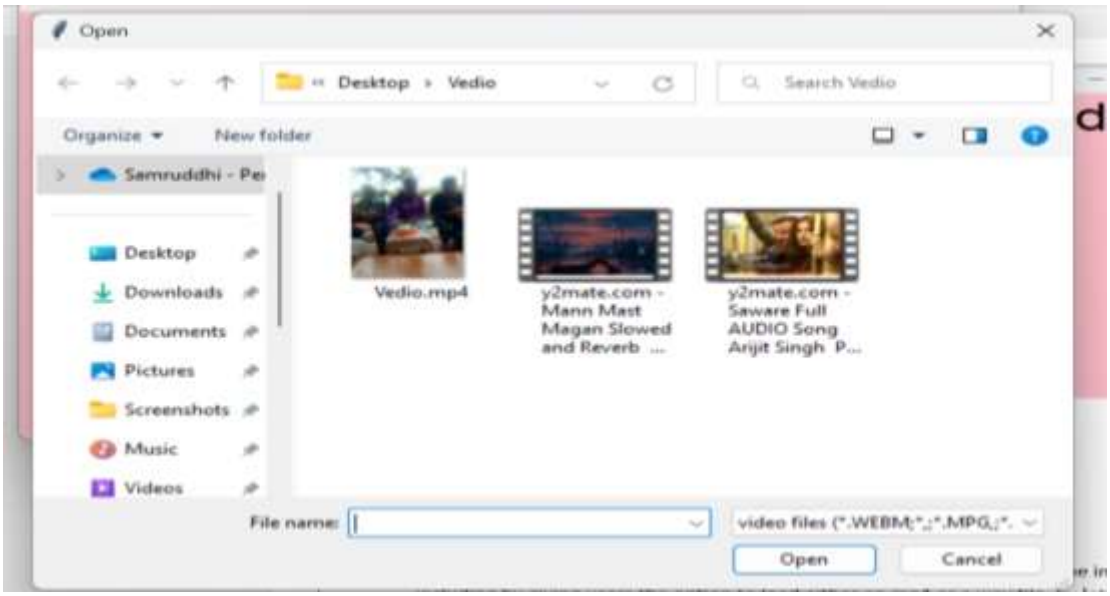
Test cases for video to audio converter:

- To check your video is downloaded in system or not.
- To check the after extracting the audio the vedio is convert or not.
- To check the all formats of audio (eg., mp3,ogg)
- To check the longer video convert the audio files
- To check the quality of audio

6.Snapshots

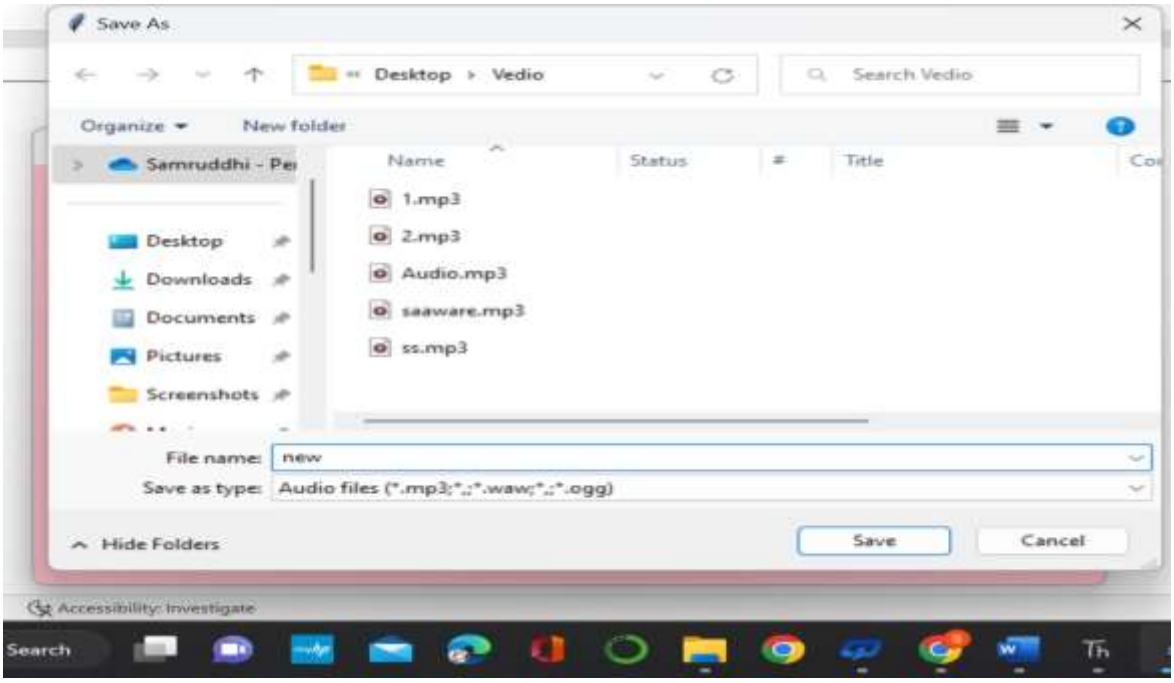


img 1.1., open the window for video to audio converter



img 1.2., select the video from this window

img 1.3., here select audio format and click the export button



img 1.4 save the audio file name

img 1.5., convert the video into audio

7.Conclusion

This program offers a very simple, barely functioning sample. This might be improved in a number of ways, including by giving users the option to load either an mp4 or a wav file, by letting them select from a variety of speech recognizers, and by showing more details like file size and length. The user experience of simple applications can be enhanced with graphical user interfaces, which are simple to add in Python using tools like Moviepy and Tkinter Designer. Performance issues can also be resolved using threading and running computationally intensive tasks in the background to minimize their impact on the user experience.

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).

[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).