Laboratory Teaching Manual
Workshop 1

SCIE 1006
Big Data and Smart Technology

Hong Kong Baptist University
Faculty of Science

Latest AI tasting

Learning Outcomes

By finishing this session, you should be able to

- Create subtitles from video voice using the speech recognition system Whipser with Subs AI.
- Change a person's voice into another person's voice using the singing voice conversion model SoftVC VITS (so-vits-svc).
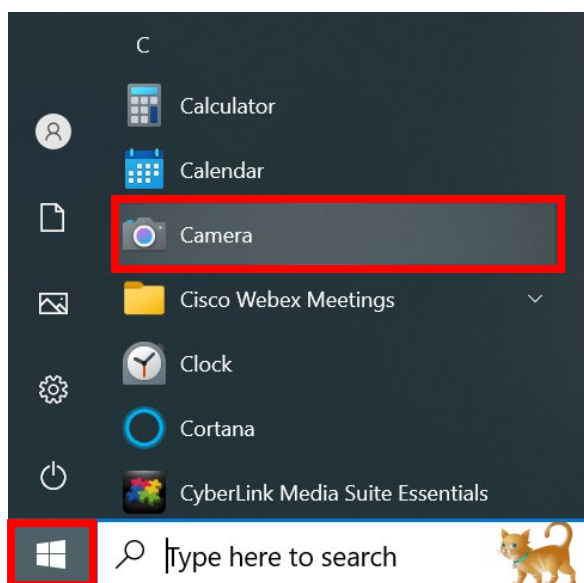
Part 1: Video Recording

Before we convert the voice, we need a video! Let's record a video now. You are already divided up into 4-person groups. Please make one video for each group.

### A. Read the following requirements carefully before recording the video

- Use the lab laptop computer we provided to record the video, there are built-in camera and microphone in the laptop.

- If you want to use your phone to record the video, please make sure you know how to transfer the video to the laptop computer.

- Save the video to desktop.

- The video should be **around 30 seconds** long.

- **Everyone** in the group should be in the video.

- When speaking or singing, please be loud and clear. You have to speak or sing in English.

- In every moment of the video, there should be only one person speaking or singing.

- Please be polite to other groups, try to avoid making loud noises when other groups are recording, you can try to coordinate with other groups to take turns for recording.

- Have fun!

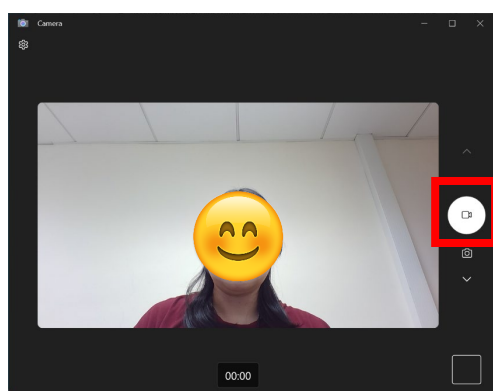## B. Recoding video using the lab laptop computer

1. Click the Windows Start button, then scroll down and select Camera.
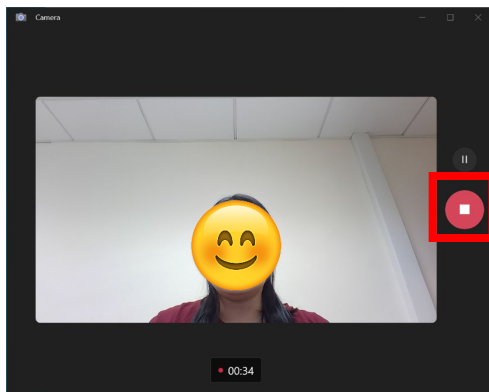

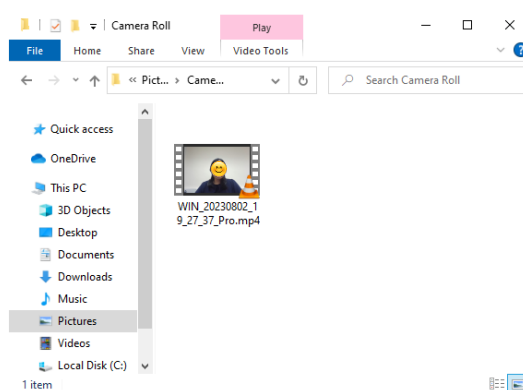2. Slide the shutter to the right to uncover the camera (if required).
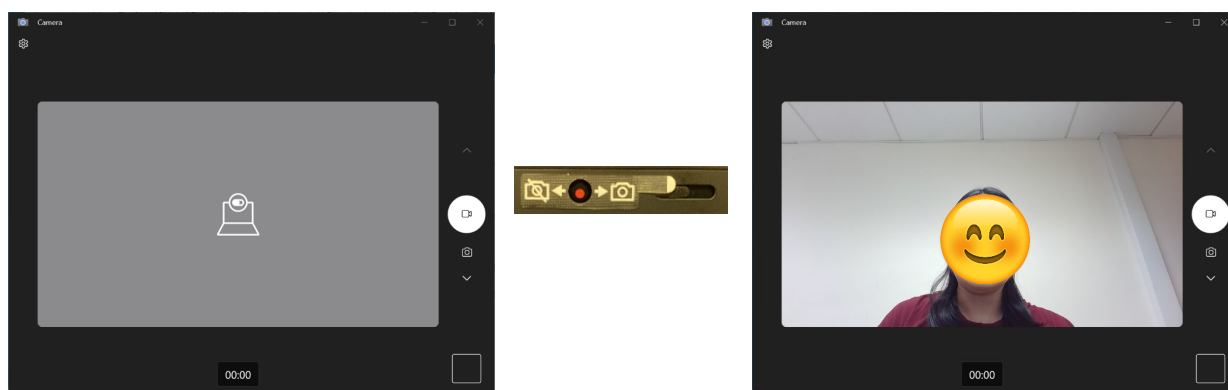
3. Click the Take video button to start recording. Now, take turns singing or speaking within the video.
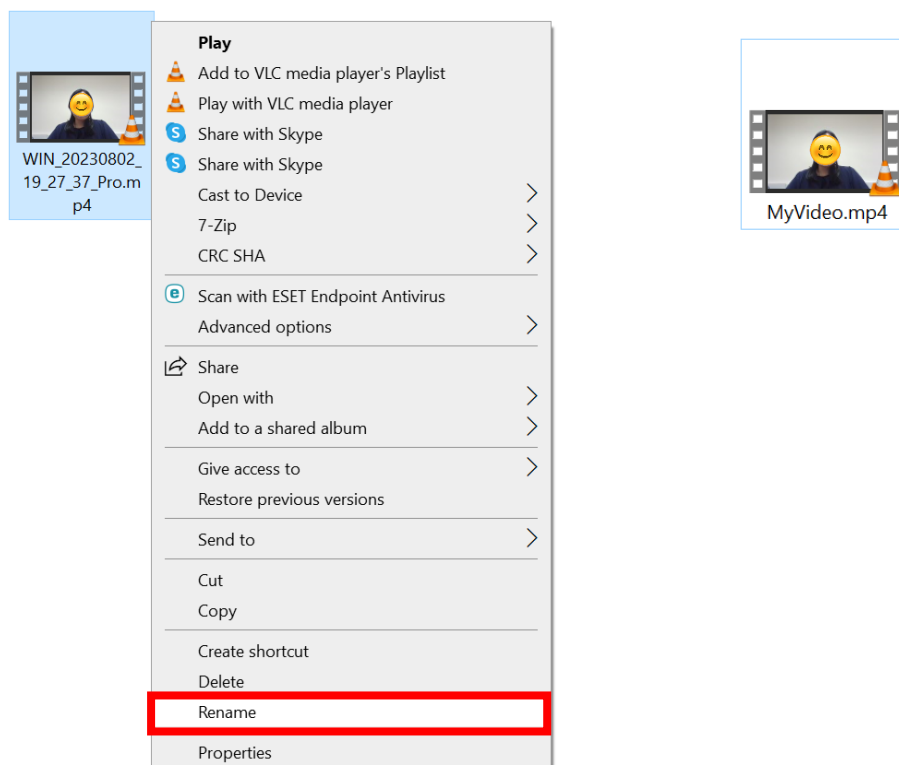
4. To end the video recording after about 30 seconds, click the stop video taking button.
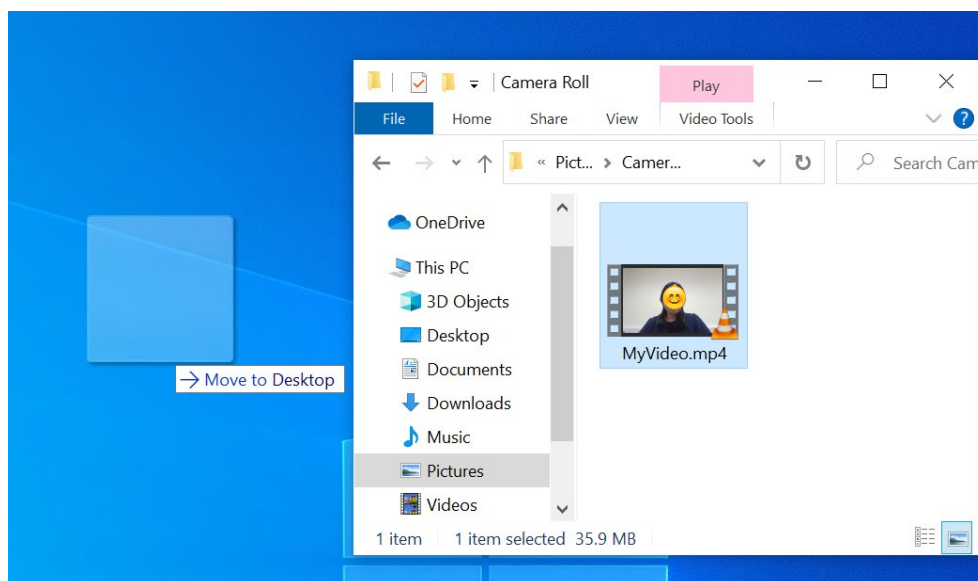

5. Press the Windows logo key ⊞ + E on your keyboard to open File Explorer and you can find the video in C:\Users\User\Pictures\Camera Roll.
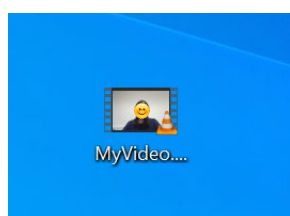

6. Right-clicking the file and choose the Rename option to rename the video to "MyVideo.mp4".

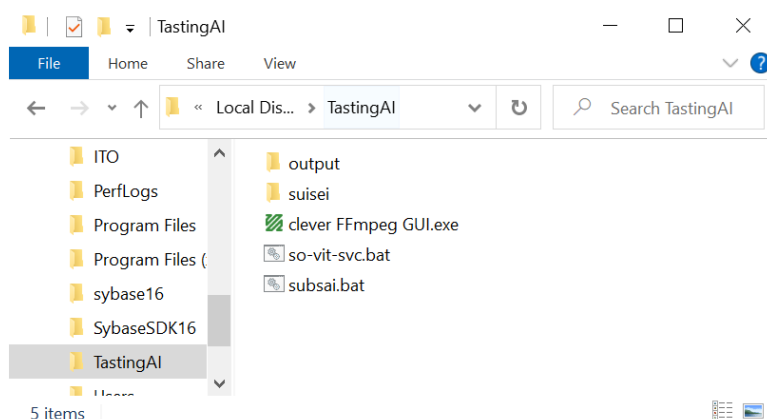7. Left-click and hold the left mouse to move the MyVideo.mp4 to Desktop.

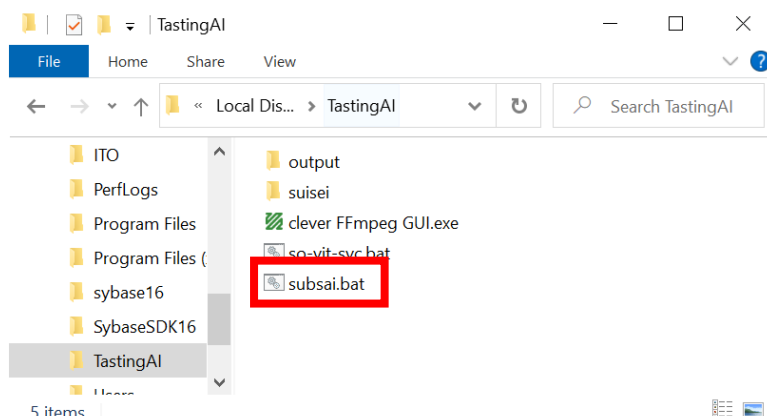8. Now, the MyVideo.mp4 is on Desktop.

Part 2: Creating Subtitles using Whisper with Subs AI

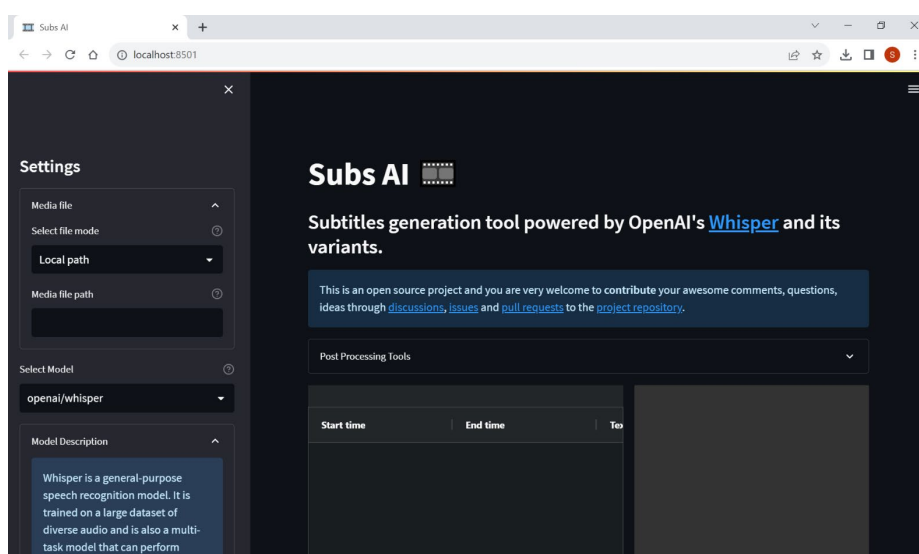Before going on, check that MyVideo.mp4 is on the Desktop.

1. Press the Windows logo key ■■ + E on your keyboard to open File Explorer and go to C:\TastingAI folder.
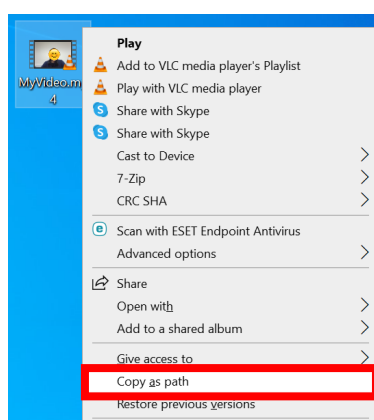
2. Double-click subsai.bat to start Whisper with Subs AI.

3. After starting Subs AI, a browser will open automatically. If not, please open the browser and go to http://localhost:8501. Then you will see the following page:


4. On the desktop, press and hold the Shift key on keyboard, and right-click the MyVideo.mp4, then select "Copy as path" to copy the file path.

5. Press Ctrl + V on your keyboard to paste the copied path into Subs AI's "Media file path" area, which is located on the left side of the webpage. **Delete the quotation marks at the start and end of the path** and it should now be C:\Users\User\Desktop\MyVideo.mp4.
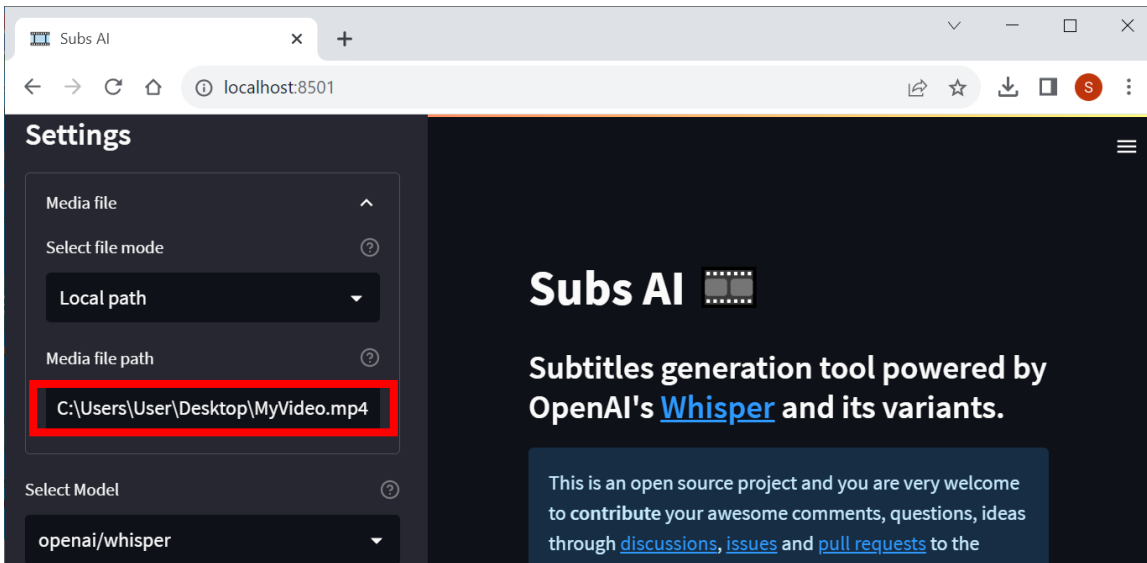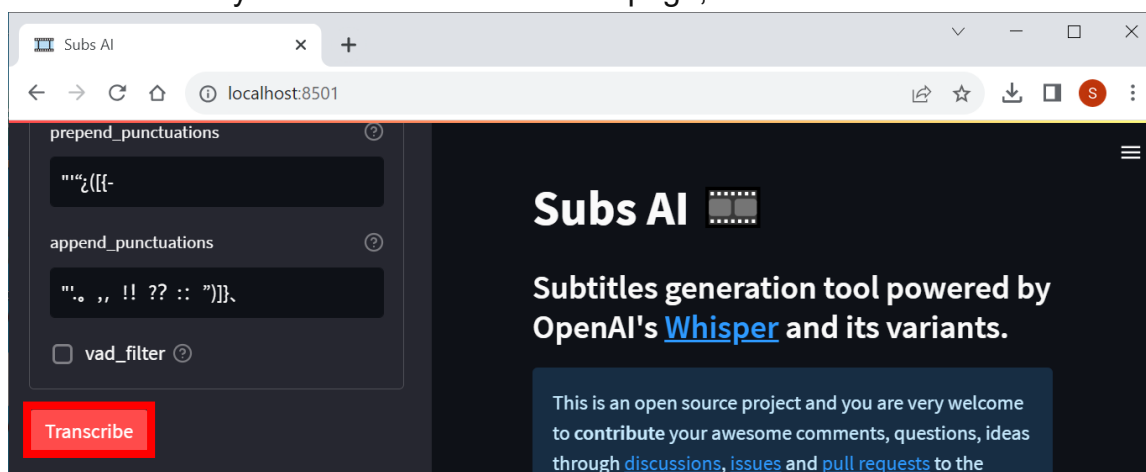

6. Choose "guillaumekln/faster-whisper" under Select Model.
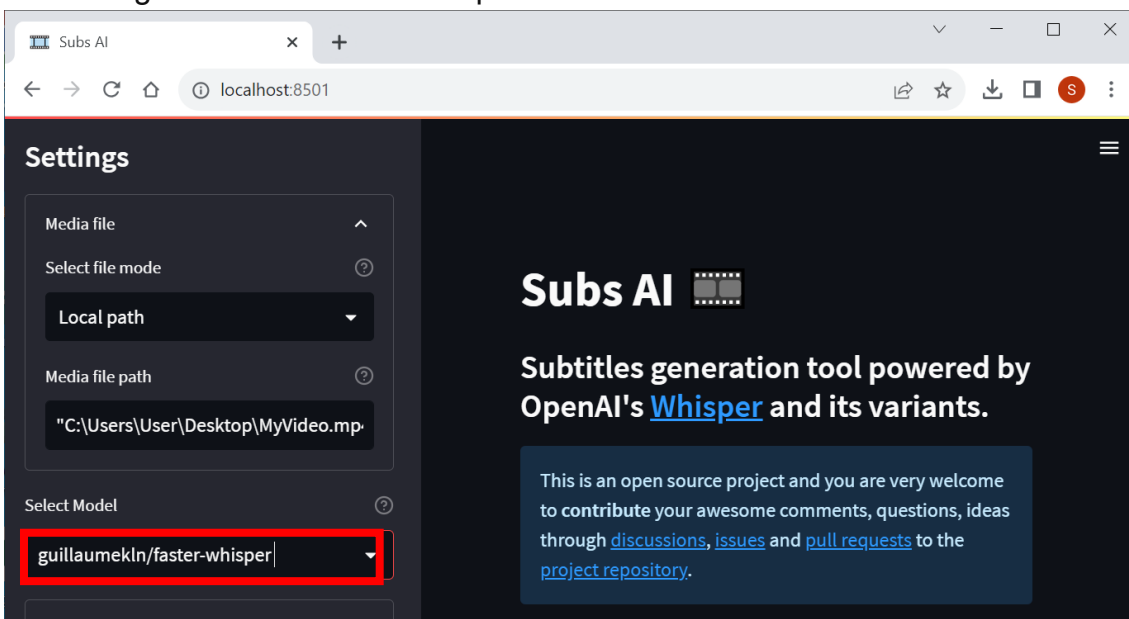
7. Scroll down and click on "Model Configs" to expand the section. Once expanded, you will see an option named "model_size_or_path", click on the dropdown menu and select "medium_en".
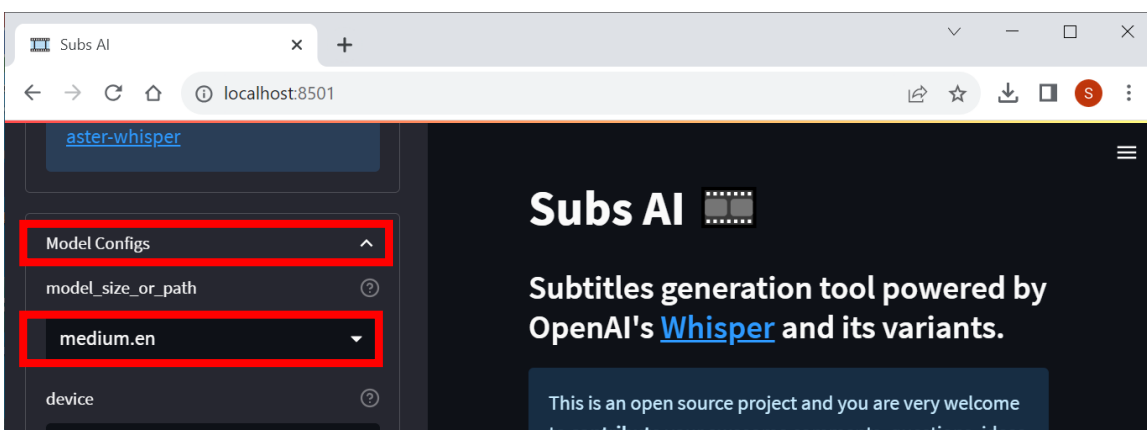
8. Scroll all the way down to the bottom of the page, and click on "Transcribe".


9. Wait for few minutes, and you will see the subtitles appear on the right-hand side of the page. You can play the video with the subtitles to check if the subtitles are correct.
Note: We can modify the subtitles in subsequent steps.
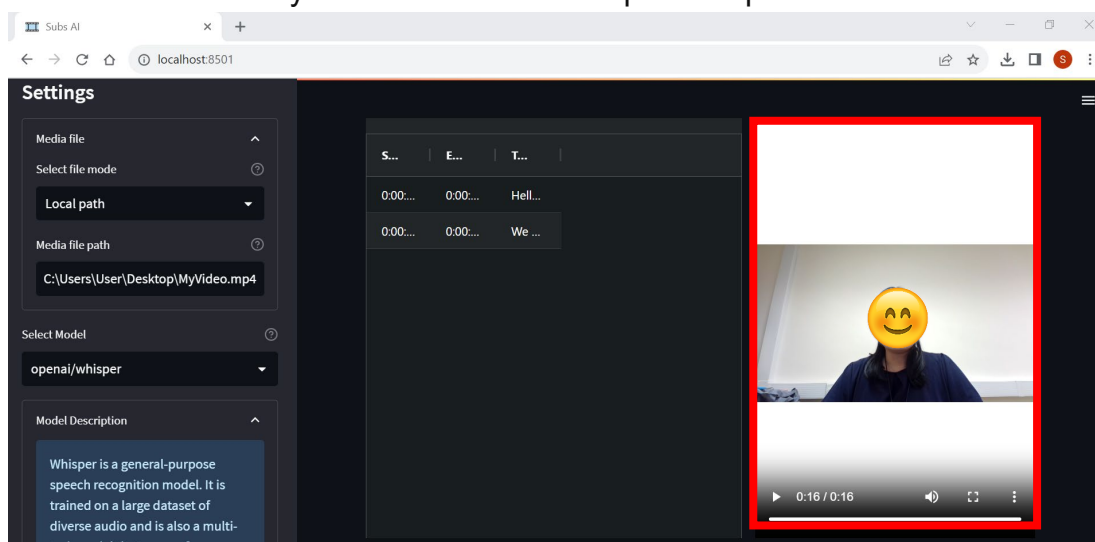
10. Click "Export" at the bottom of the page to expand the section, select "srt" as the export format, and then click the "Export" button.
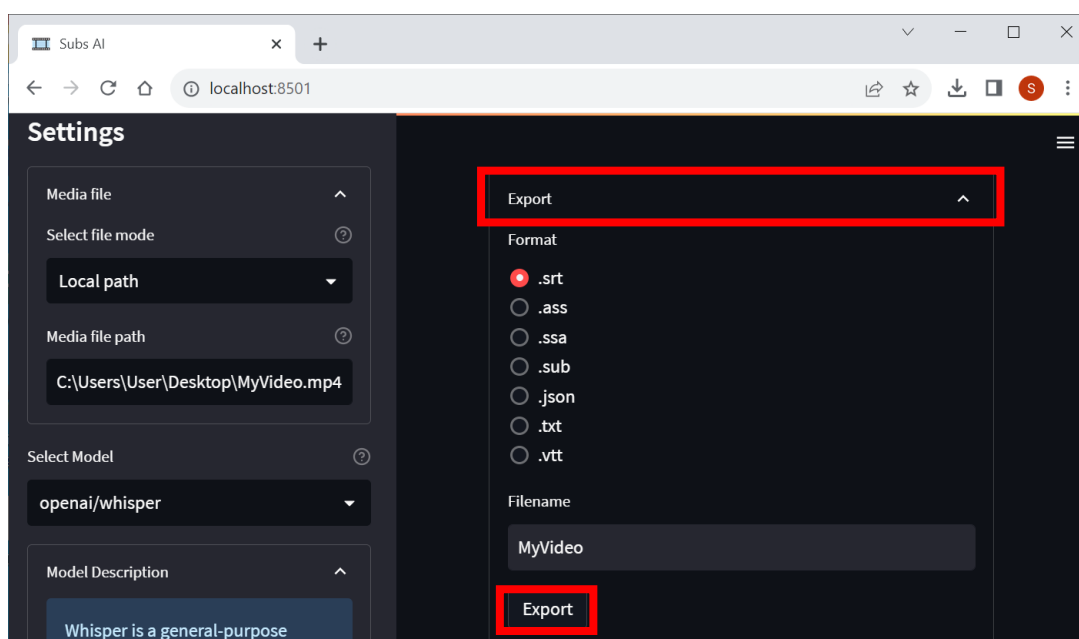
11. On the desktop, a new file called MyVideo.srt has been created. You can now play the video with subtitles. Please don't change the file name as the file name has to be the same as the video file name in order to play the video with the subtitles in most of the video players!
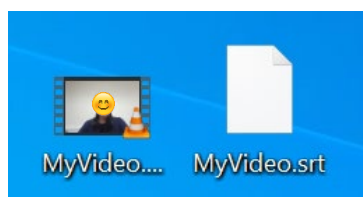
12. If the subtitles are not correct, you can edit the MyVideo.srt file with any text editor such as Notepad. Right-click the MyVideo.srt file, choose "Open with" and then select Notepad.
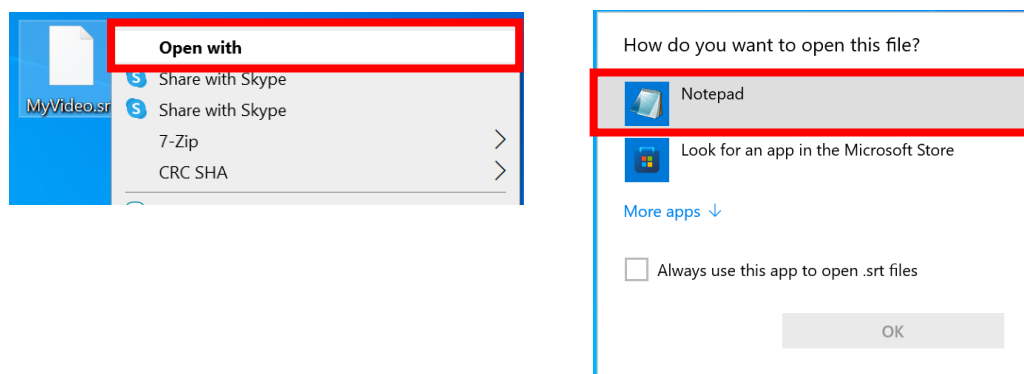
13. The format of the subtitles file is very simple, you can try to understand it yourself and edit it manually. Save the file after that.
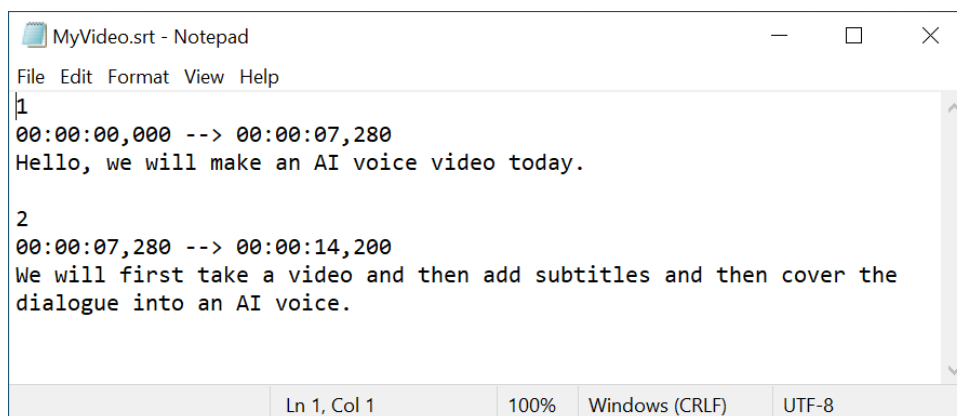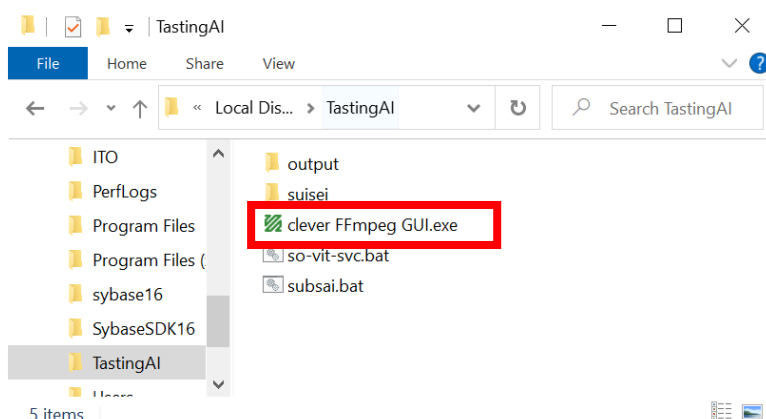
## Part 3: Changing a person's voice into another person's voice

A. Separating audio from video

We will use a tool called "FFmpeg" to separate the audio from the video.
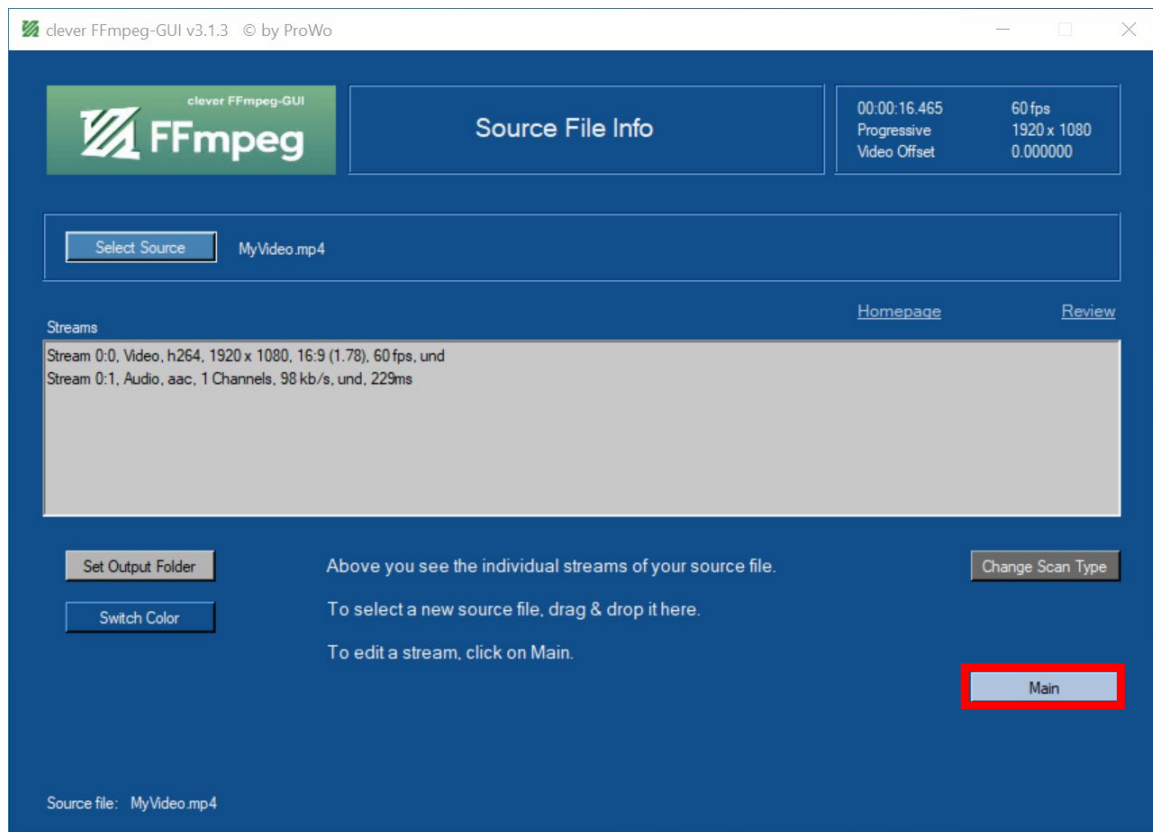
1. Press the Windows logo key ⊞ + E on your keyboard to open File Explorer and go to C:\TastingAI folder. Double-click "clever FFmpeg GUI.exe" to start FFmpeg.
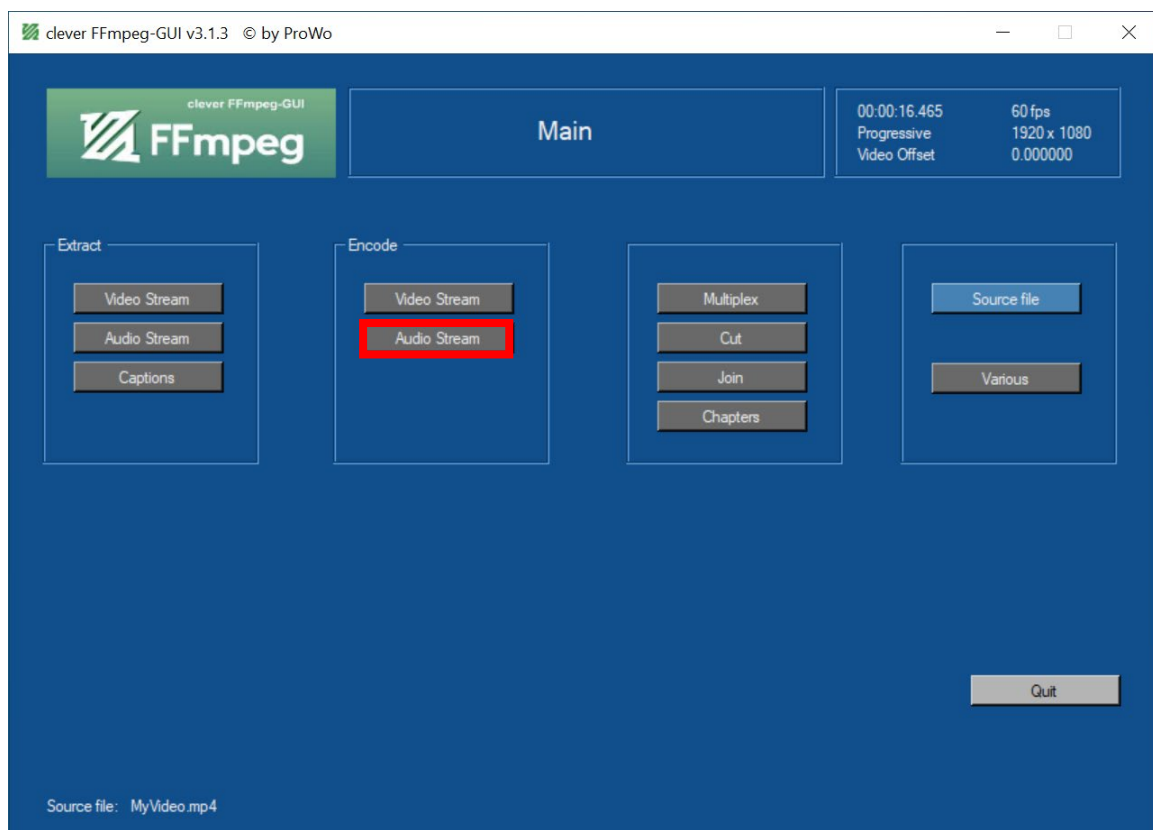

2. Click the "Select Source" button, choose the MyVideo.mp4 on Desktop and click "Open".
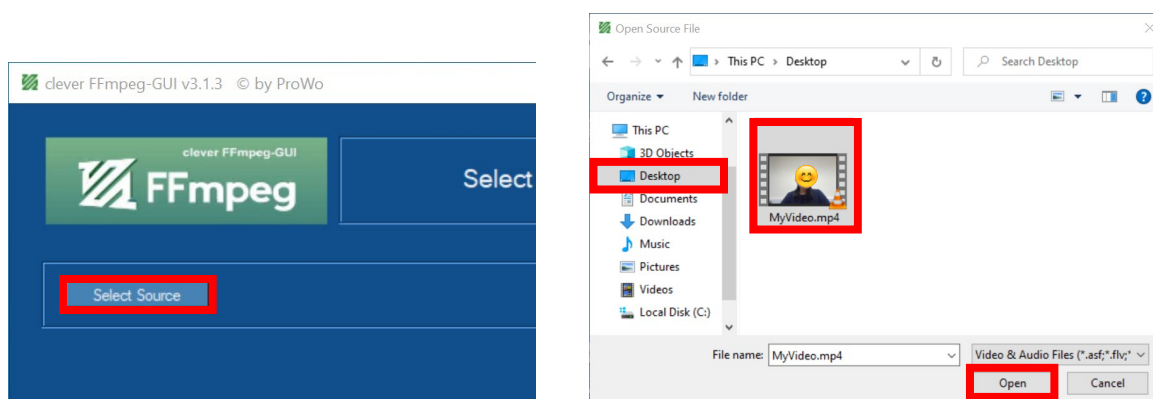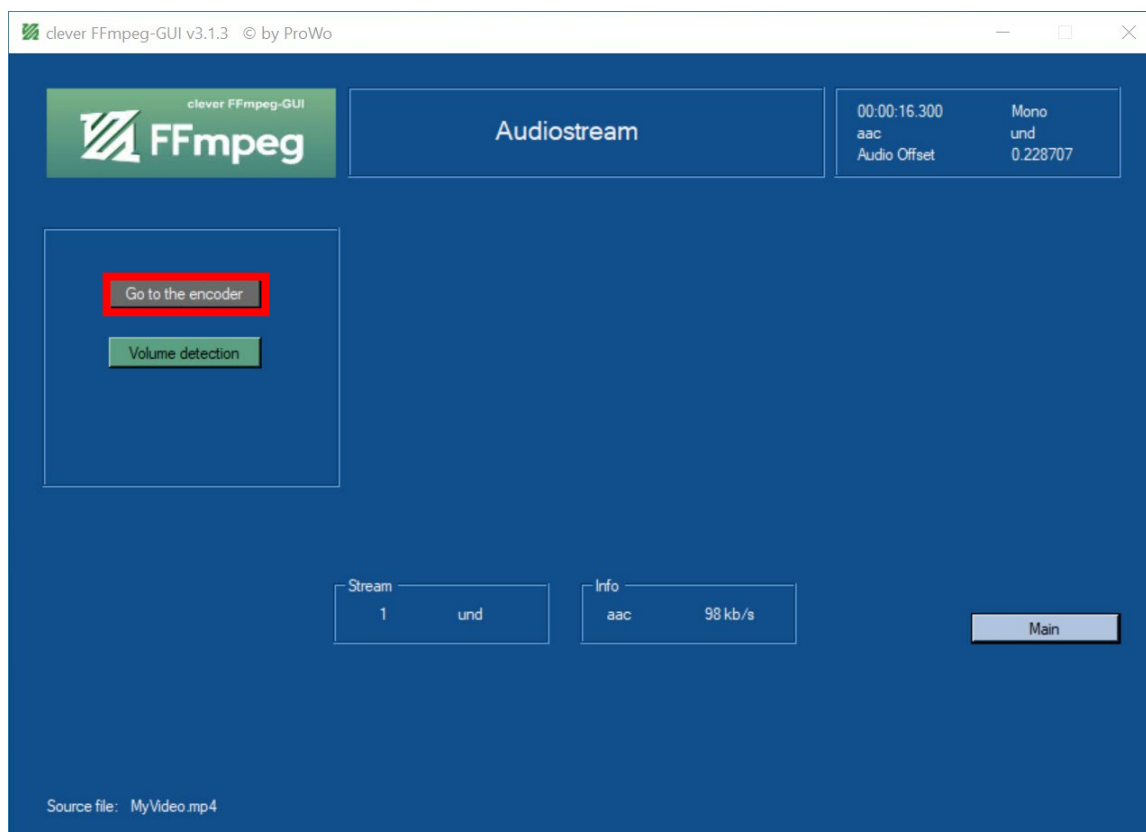
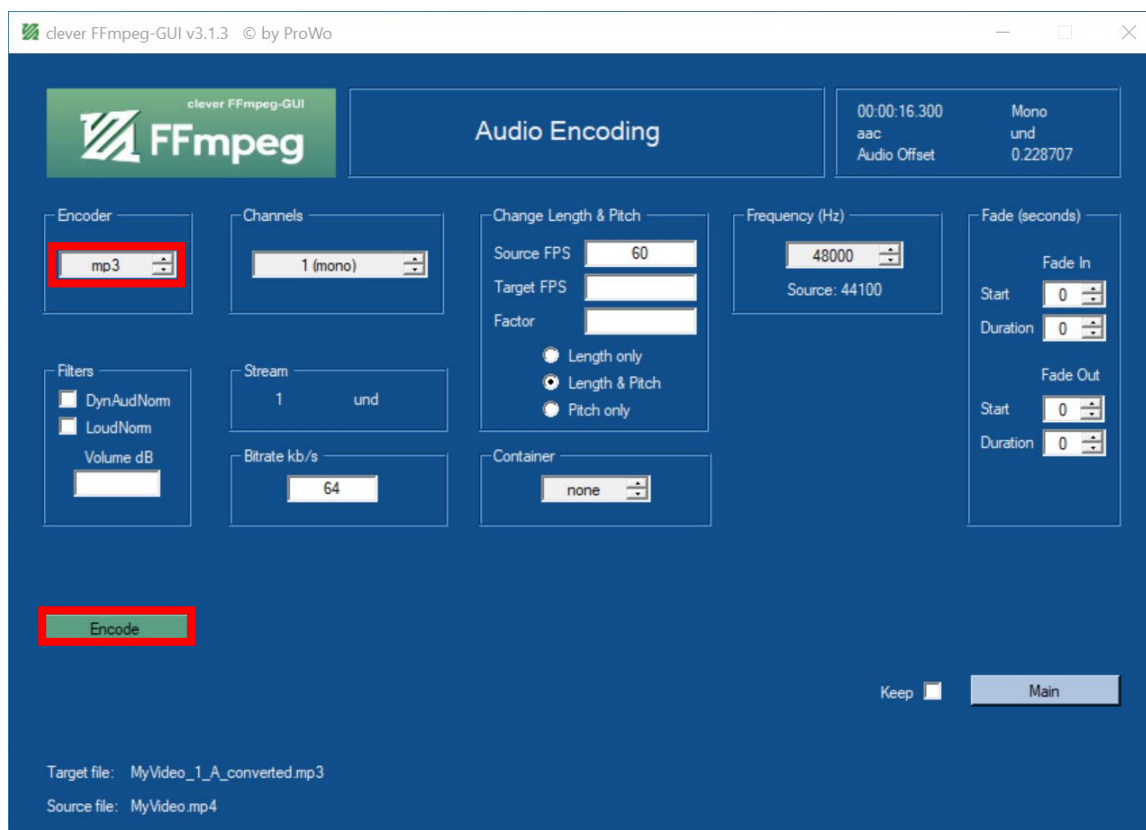3. Click the "Main" button at the bottom right to go to the main menu.

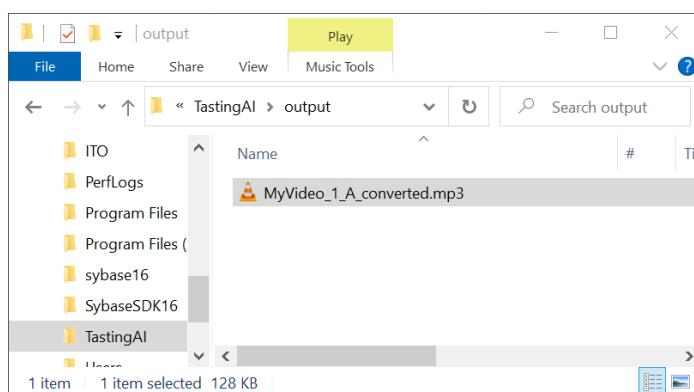4. Click the "Audio Stream" button under the Encode section.

5. Click the "Go to the encoder" button.

6. Choose "mp3" in the Encoder section, and then click "Encode".

7. The extracted audio file is now in C:\TastingAI\output.

B. Changing the voice by so-vits-svc

Now we have the audio file, we can use so-vits-svc to convert your voice to sound like a famous singer's voice.

The voice model, that we are going to use, was trained with the voice of a famous singer, called Hoshimachi Suisei. The model is trained with 2.8 hours of her singing voice for a week, with an RTX 3090 GPU, and it can convert any voice to sound like her voice.
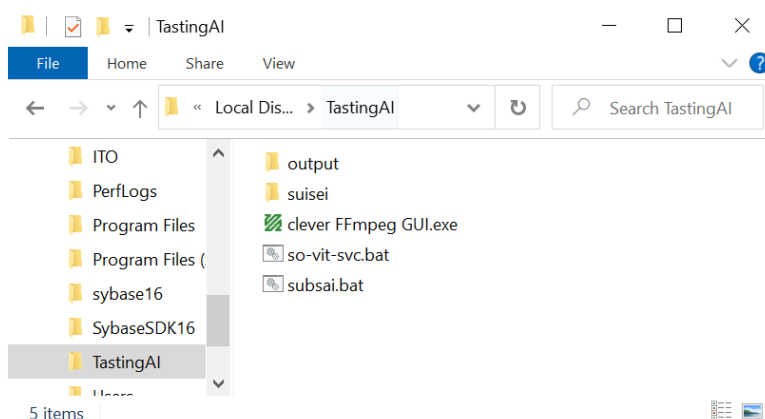

Stellar Stellar / 星街すいせい(official)
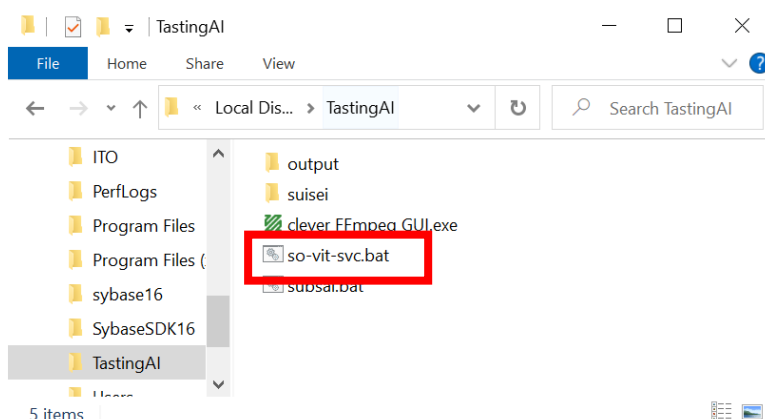
Bluerose / 星街すいせい(official)

The voice model is for educational purpose only, please do not distribute the model or any files generated from the model, please do not use the model for any other purpose. You should use it for this lab only.
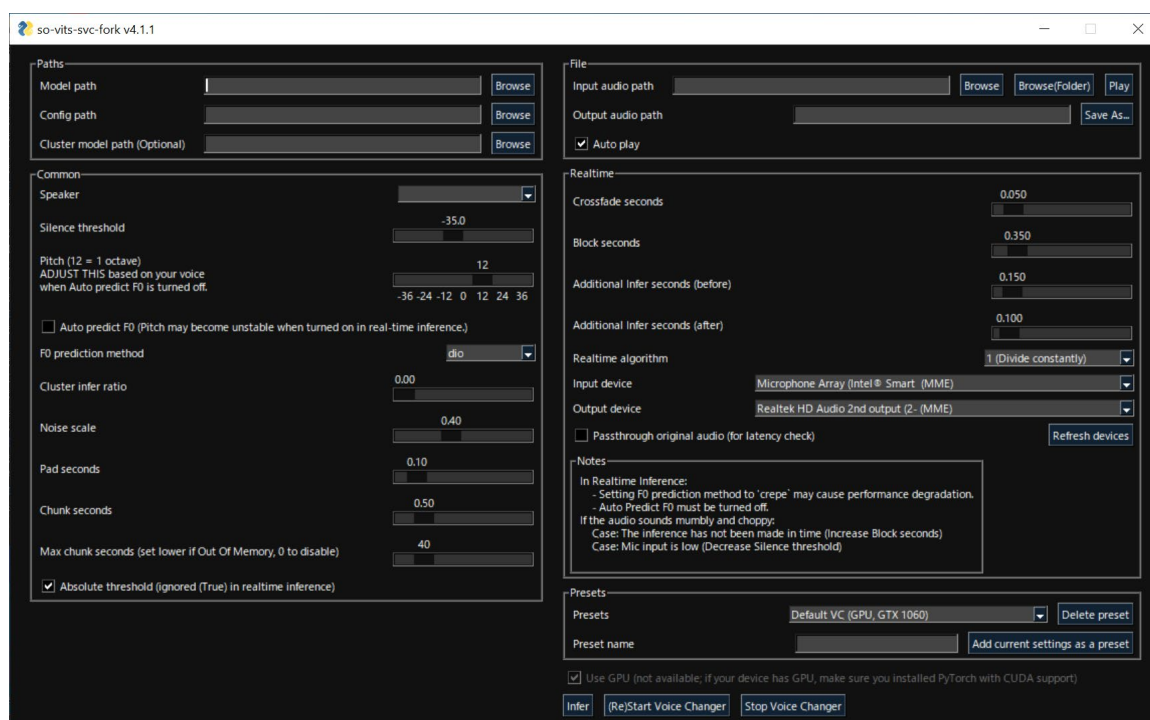
1. Press the Windows logo key ⊞ + E on your keyboard to open File Explorer and go to C:\TastingAI folder.
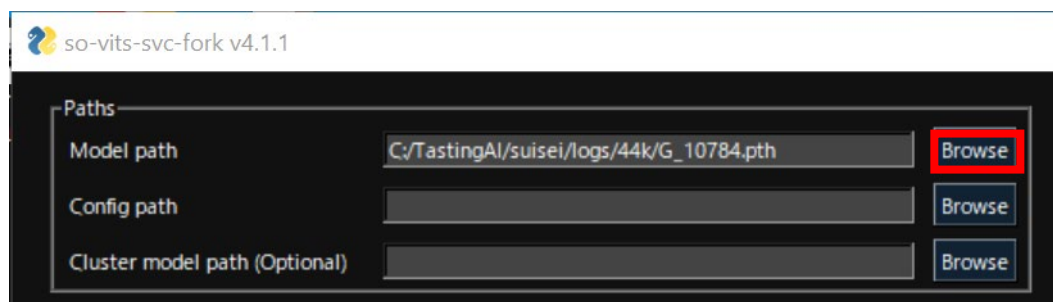

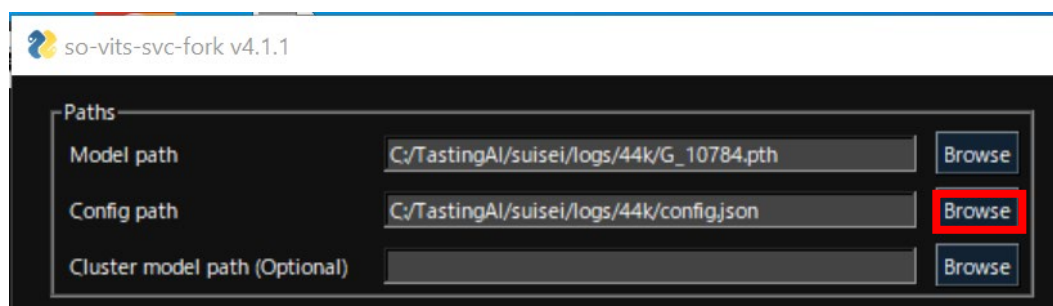2. Double-click so-vit-svc.bat.

3. Wait for few seconds, the software will open automatically:

4. Click "Browse" on Model path at the top left of the window to choose C:/TastingAI/suisei/logs/44k/G_10784.pth.

5. Click "Browse" on Config path also at the top left of the window to choose C:/TastingAI/suisei/logs/44k/config.json.

6. Under the File section at the top right, and uncheck "Auto play" and click "Browse" to choose the extracted audio file, such as C:/TastingAI/output/MyVideo_1_A_converted.mp3.
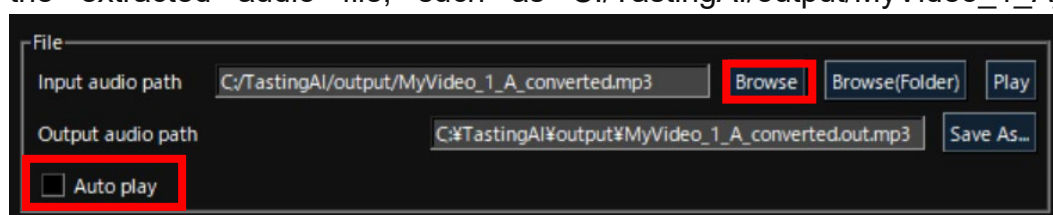
7. Under the Common section on the left, uncheck "Audio predict F0". It will sound off pitch if you are singing in your video.

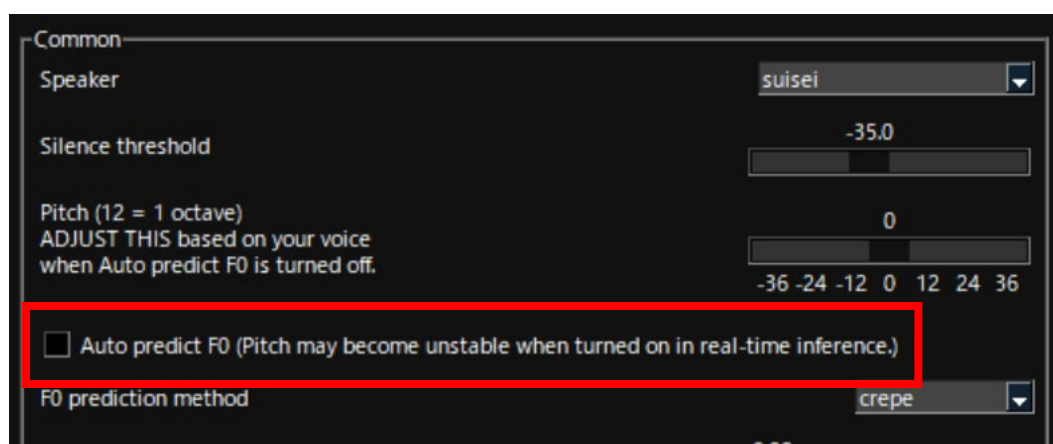8. Optional: Under the Common section, please don't change any configurations other than the following:

Config name	Meaning	Recommended value
Silence threshold	If the volume of the input audio is lower than this value, it will be considered as silence.	-35.0
Pitch	The pitch of the output audio.	Depends on the input audio, if you have a middle to high pitch voice, you can remain this value as 0. And if you have a deep voice, you can try to increase this value to +12.

9. Click the "Infer" button at the bottom to start processing the audio. It will take few minutes to process the audio, please be patient.
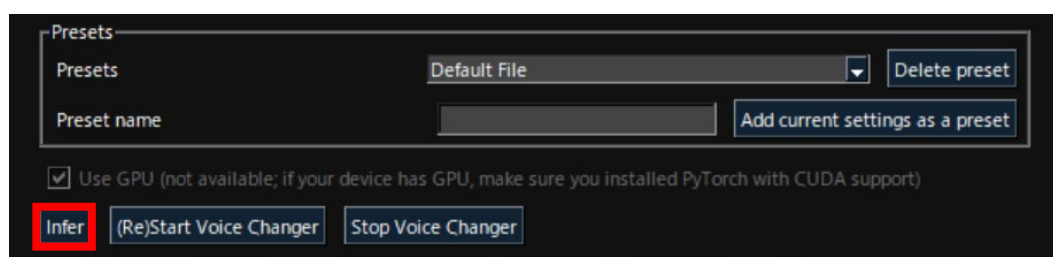
10. The output audio file will be saved to the same folder as the input audio file. The name will be the same as the input file name with ".out" appended to the end. You can find the file in C:\TastingAI\output.
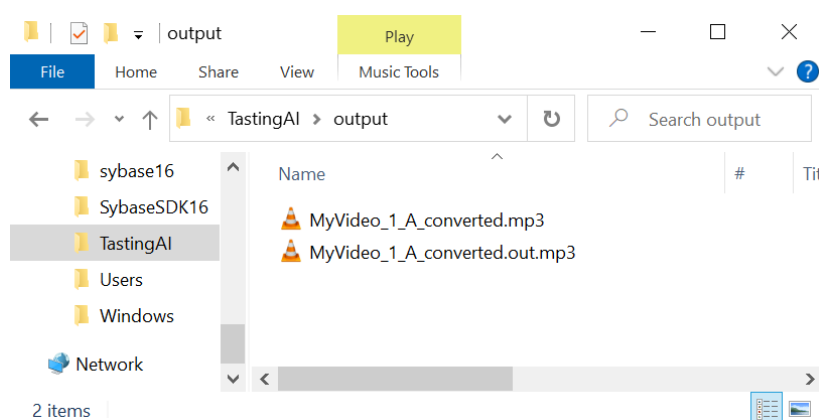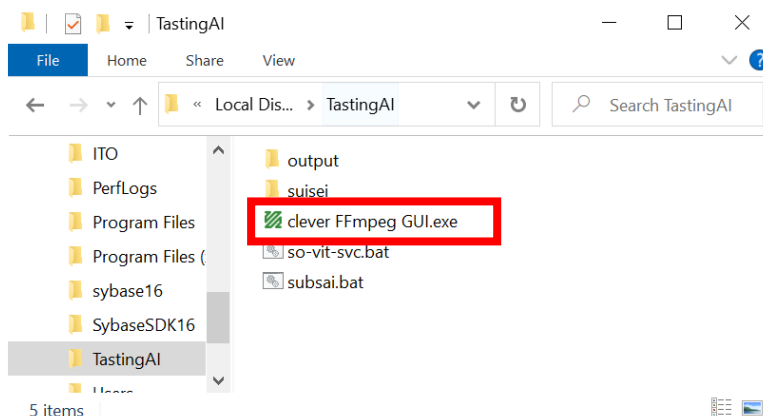

11. Open the output audio file with an audio player to check the result. (e.g. VLC media player)
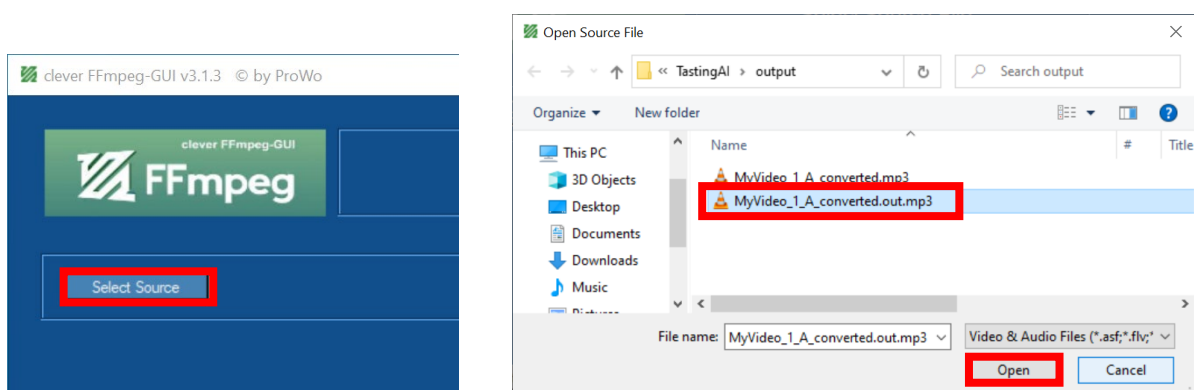
C. Merge the new voice with the video

Again, we will use FFmpeg to merge the output audio with the video.

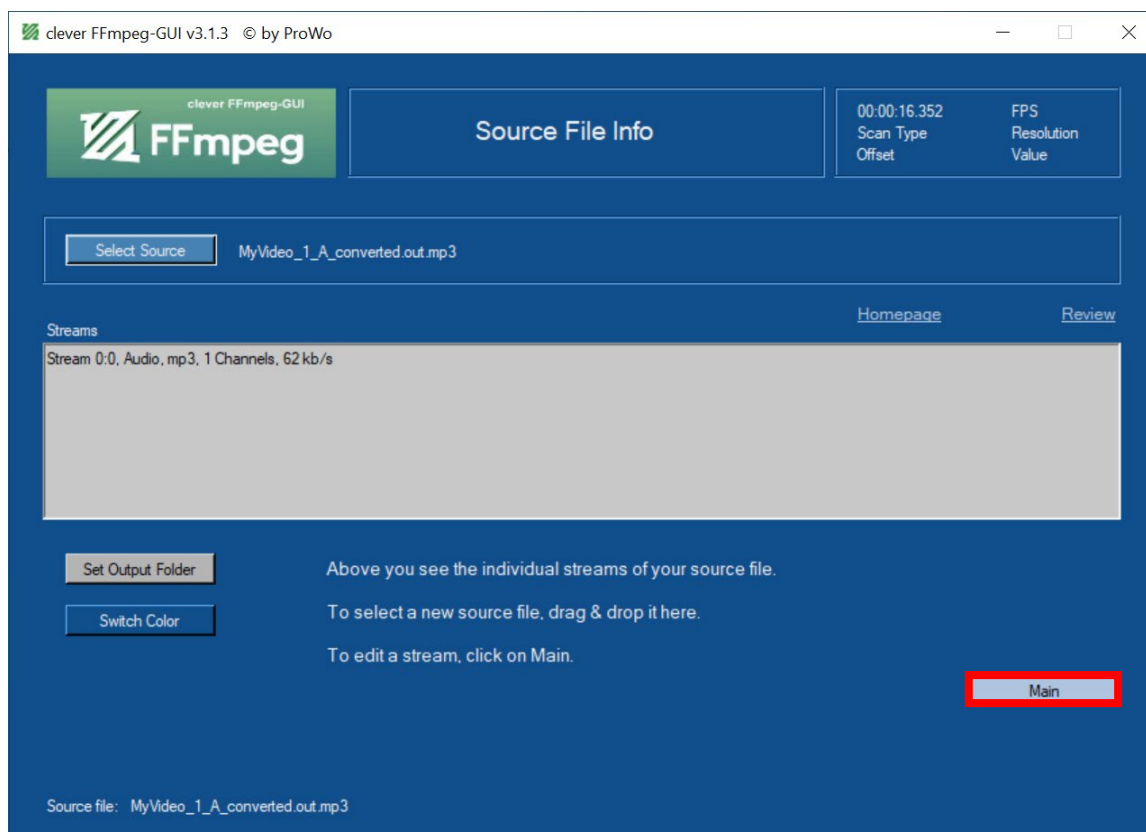1. Press the Windows logo key ▦ + E on your keyboard to open File Explorer and go to C:\TastingAI folder. Double-click "clever FFmpeg GUI.exe" to start FFmpeg.
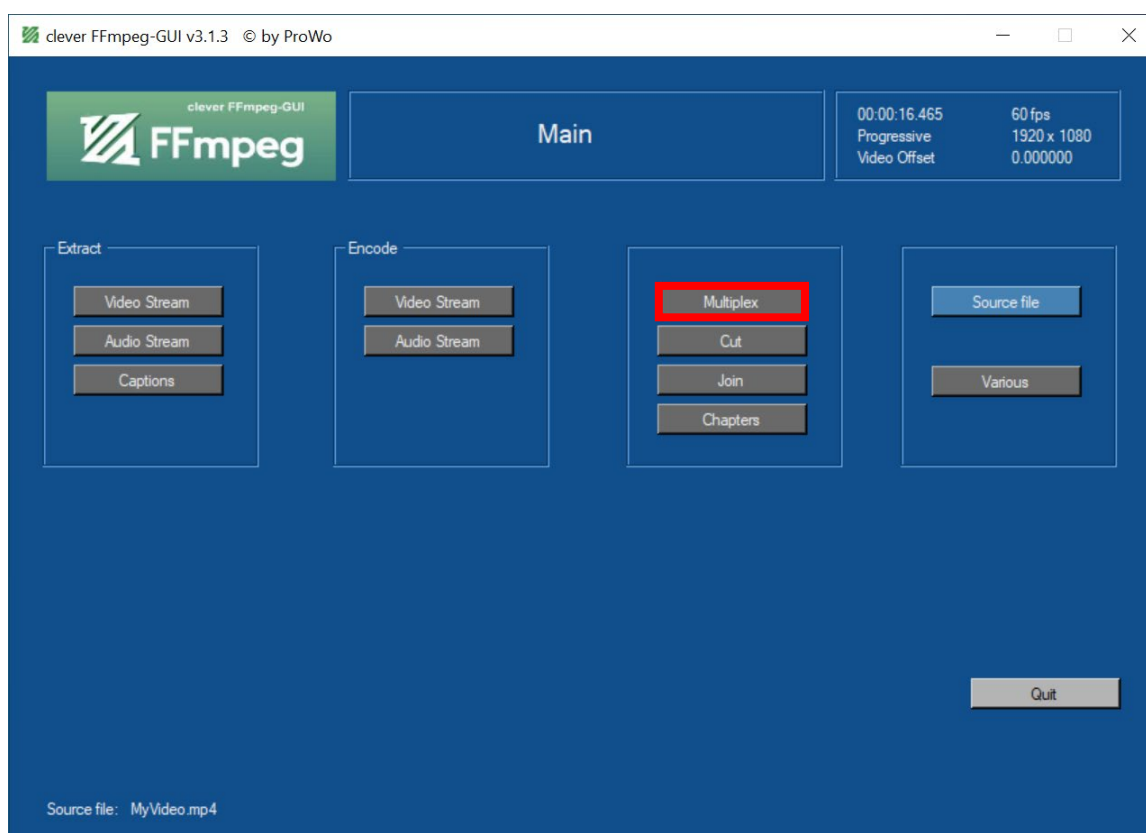

2. Click the "Select Source" button, choose the output audio file such as C:\TastingAI\output\MyVideo_1_A_converted.out.mp3 and click "Open".
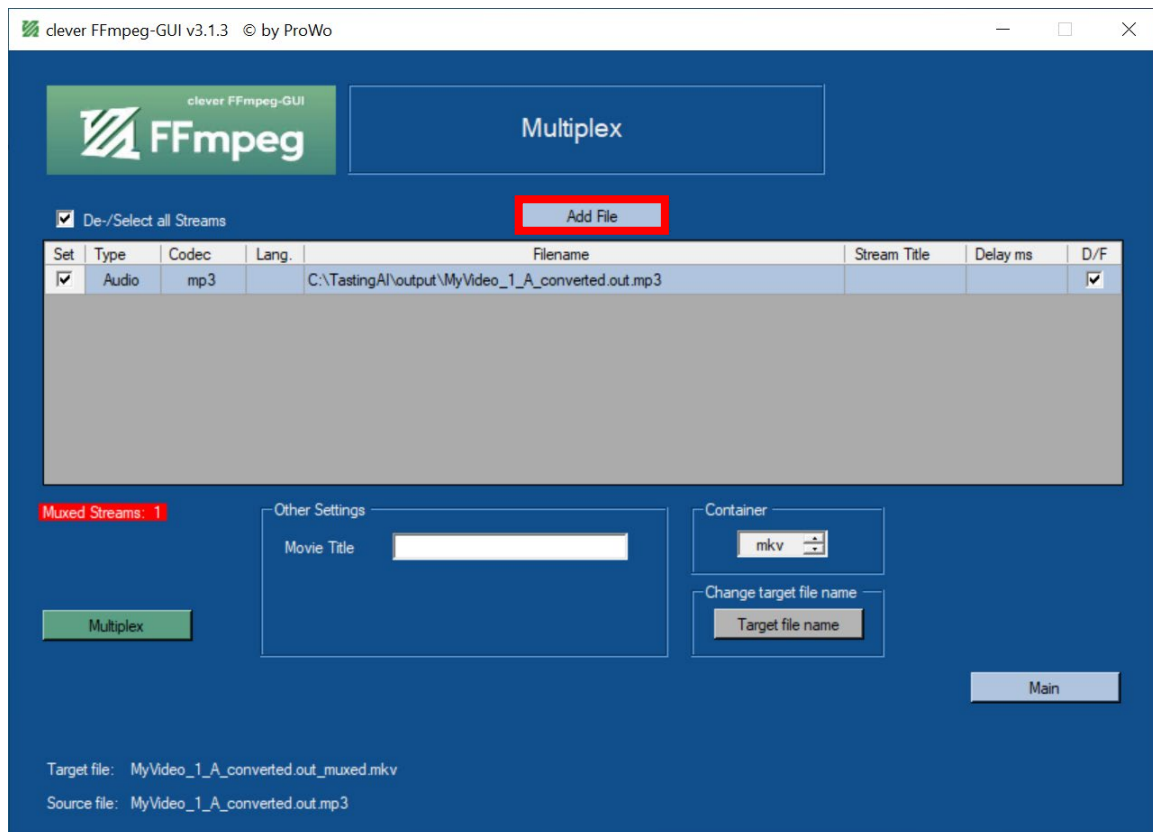
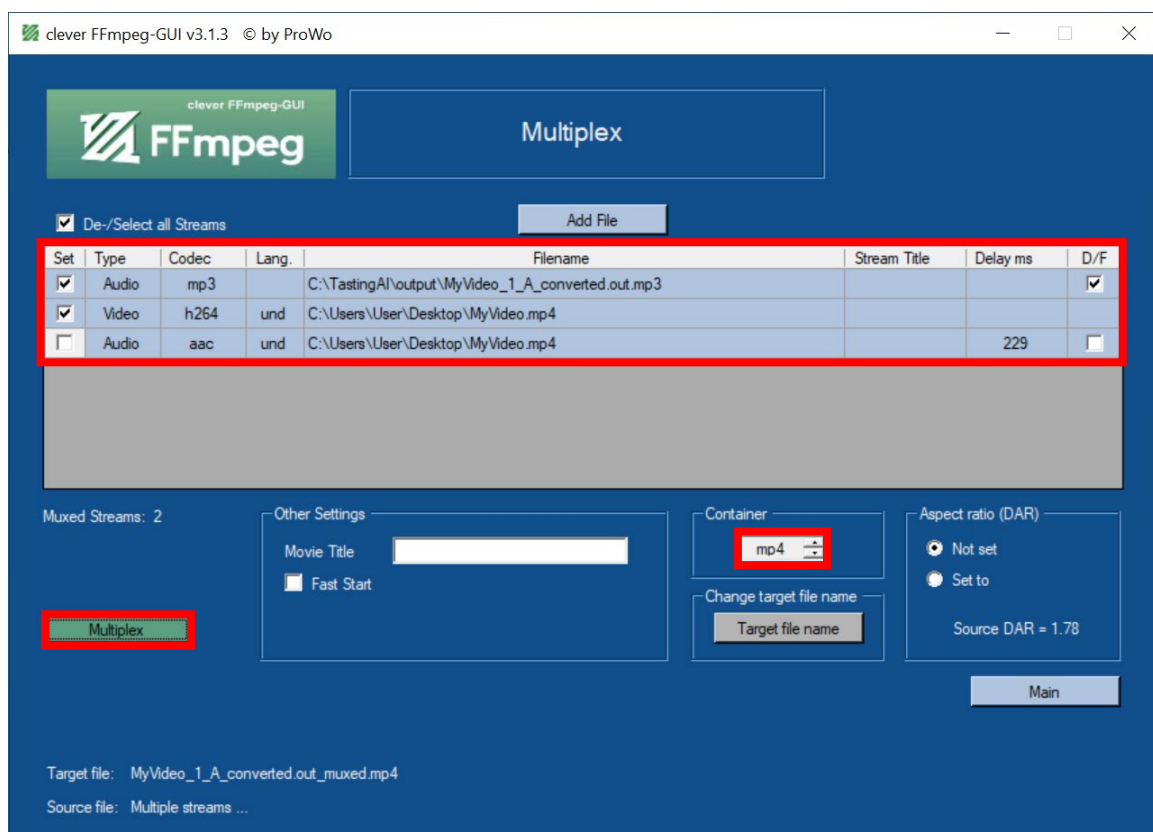3. Click the "Main" button at the bottom right to go to the main menu.

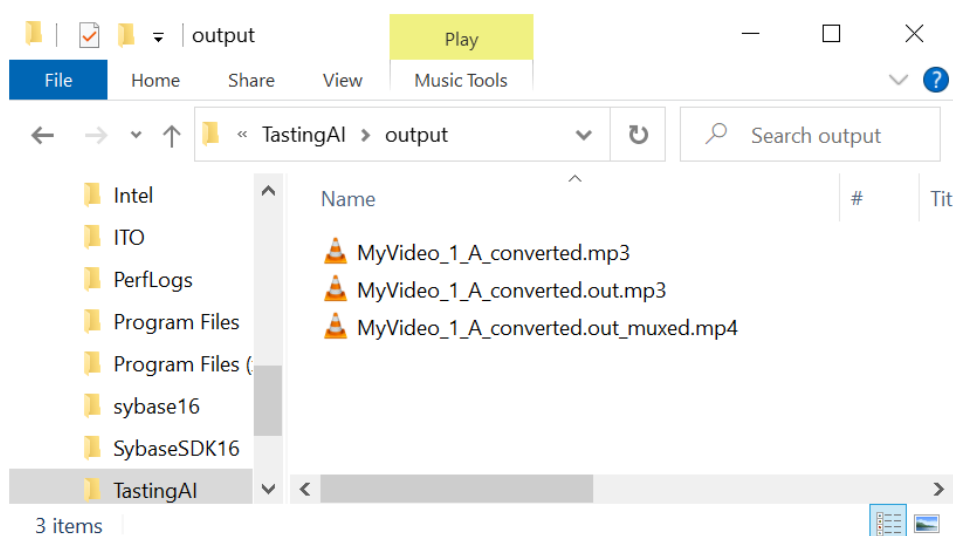4. Click the "Multiplex" button.

5. Click "Add File" to choose the "MyVideo.mp4" on Desktop.

6. Only select the output audio and the video in the table as follows, choose "mp4" for container and then click the "Multiplex" button at the bottom left.

7. You can find the video file in C:\TastingAI\output. The name will be the same as the input file name with "_muxed" appended to the end.

8. Open the video file with a video player to check the result. (e.g. VLC media player)

Discussion

Question: Is Artificial Intelligence (AI) Good or Bad? Justify your answer.
- Answer the discussion question in the Moodle submission box.

Submission

Submit the following items to the submission box on Moodle:
1. 30 seconds video (e.g. MyVideo.mp4)
2. Subtitle of the video in .srt format (e.g. MyVideo.srt)
3. Extracted Audio (e.g. MyVideo_1_A_converted.mp3)
4. Converted Audio (e.g. MyVideo_1_A_converted.out.mp3)
5. Final Merged Video (e.g. MyVideo_1_A_converted.out_muxed.mp4)
6. Discussion question

REFERENCES

1. Abdeladim-S/subsai: 🎞 subtitles generation tool (Web-UI + CLI + Python package) powered by OpenAI's whisper and its variants 🎞 . (n.d.). GitHub. https://github.com/abdeladim-s/subsai

2. Clever FFmpeg-GUI 3.1.3 free download. (n.d.). VideoHelp - Forum and Software downloads. https://www.videohelp.com/software/clever-FFmpeg-GUI

3. (n.d.). FFmpeg. https://www.ffmpeg.org/

4. Introducing whisper. (n.d.). OpenAI. https://openai.com/research/whisper

5. Suisei Channel. (n.d.). YouTube. https://www.youtube.com/@HoshimachiSuisei