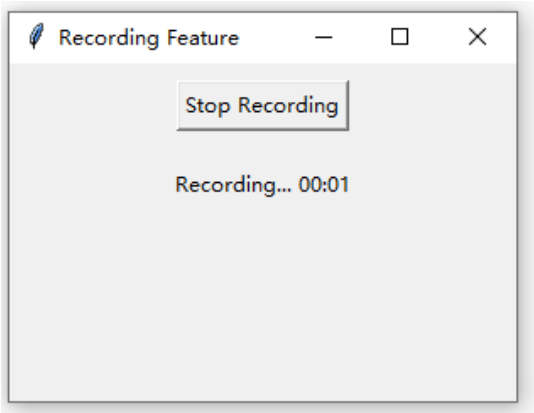
1. Main Interface

Upon launching the application, you will see the main interface with four primary features:

1. **Recording Feature**
2. **Audio Preprocessing (UVR5)**
3. **Voice Cloning**
4. **Video Processing**

2. Recording Feature

1. Click on the **Recording Feature** button to open the recording window.
2. In the recording window, click **Start Recording** to begin recording audio.
    - The interface will change to show the recording time.



3. Click **Stop Recording** to stop the recording.

4. You will be prompted to select a folder to save the recorded audio file.

3. Audio Preprocessing (UVR5)

1. Click on the **Audio Preprocessing (UVR5)** button to launch the UVR5 tool.

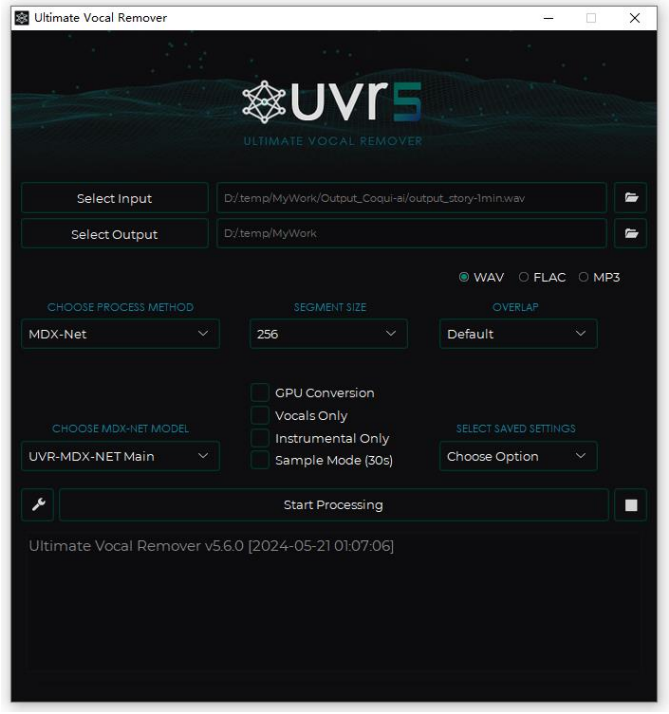
2. Follow the UVR5 user manual for detailed instructions on using this tool.

   Short UVR5 tutorial:

   https://www.youtube.com/watch?v=pzUQ232PU44

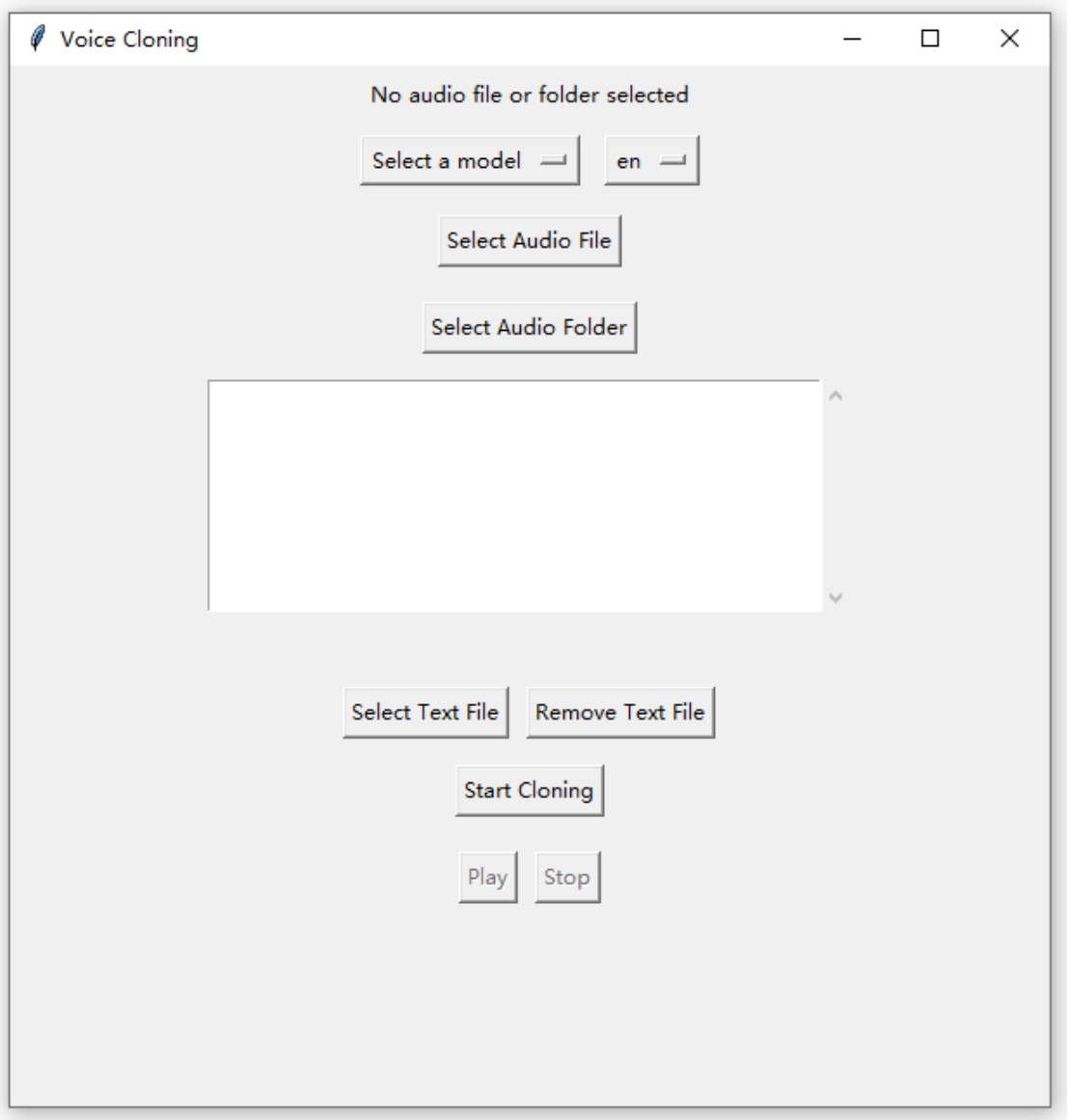
   Detailed UVR5 tutoral:

   https://www.youtube.com/watch?v=p4YSfh-FbAs



4. Voice Cloning

1. Click on the **Voice Cloning** button to open the voice cloning interface.



2. Select a voice cloning model from the dropdown menu.

    - Models available:

        - **`tts_models/multilingual/multi-dataset/xtts_v2`**
        - **`tts_models/multilingual/multi-dataset/your_tts`**

3. Choose the language corresponding to the selected model:

    - **xtts_v2** supports:

        - English (en), Spanish (es), French (fr), German (de), Italian (it), Portuguese (pt), Polish (pl), Turkish (tr), Russian (ru),

Dutch (nl), Czech (cs), Arabic (ar), Chinese (zh-cn), Japanese (ja), Hungarian (hu), Korean (ko), Hindi (hi)

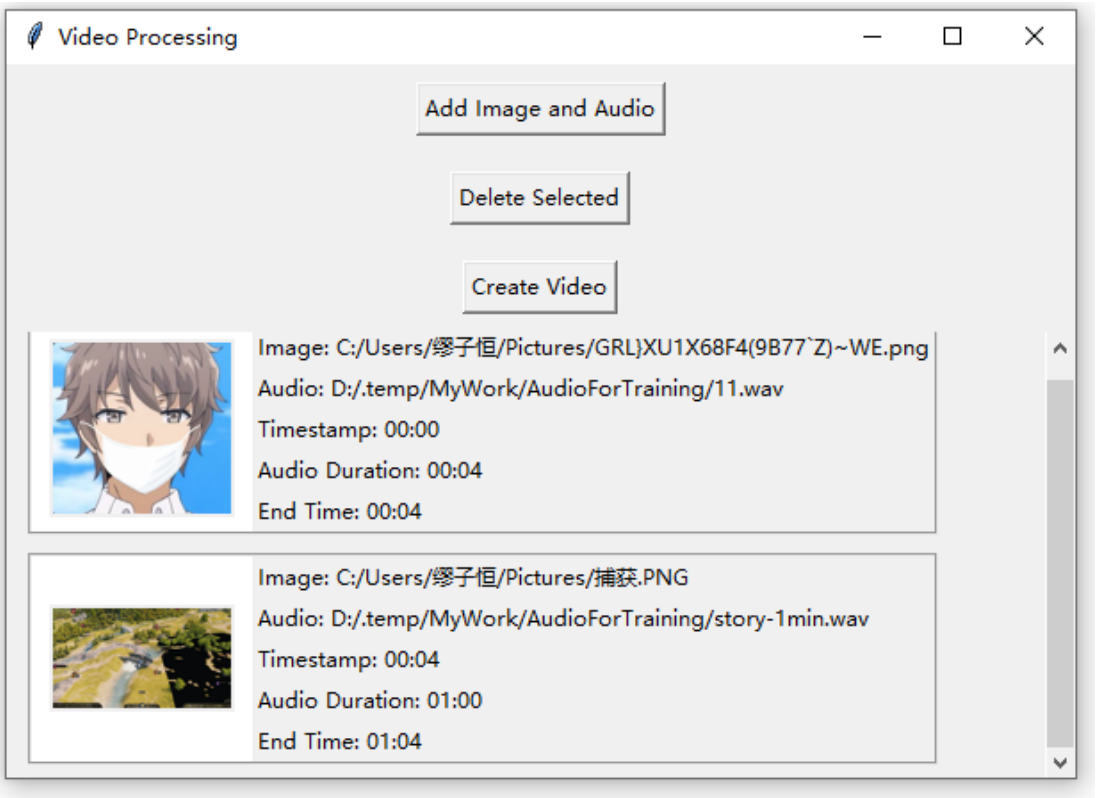
- **your_tts** supports:
    - English (en), Portuguese (pt-br), French (fr)

Note: The output quality of xtts_v2 model is much better than your_tts model but your_tts model is a little faster than xtts_v2 model. Xtts_v2 model is recommended.

4. Click **Select Audio File** or **Select Audio Folder** to choose the audio file for voice cloning. (**Select Audio Folder** means using all the audio files in one folder to get **one** output audio.)
5. Enter the text needs to be converted to speech in the text box. You can choose to:
    - Manually input text.
    - Click **Select Text File** to choose a text file containing the text.
    - Click **Remove Text File** to remove the selected text file.

    Note: You can only use either the text box or a text file, not both simultaneously.
6. Click **Start Cloning** to generate the synthetic speech.

5. Video Processing



Opening the Video Processing Module

1. Launch the main application.
2. Click on the "Video Processing" button to open the module.

Adding Image and Audio Files

1. **Add Image and Audio**

    - Click the "Add Image and Audio" button.
    - A file dialog will appear prompting you to select an image file. Supported formats include PNG, JPG, JPEG, BMP, GIF, TIFF, WEBP, and ICO.
    - After selecting the image, another file dialog will prompt you to select an audio file. Supported formats include WAV, MP3, AAC, FLAC, OGG, M4A, and WMA.
    - Enter a timestamp (MM:SS) indicating when this image and audio pair should start in the video. The default value is "00:00".

2. **View Added Pairs**

    - Each added image and audio pair will be displayed in the main window.
    - The display includes:
        - A thumbnail of the image.
        - The path to the image and audio files.
        - The duration of the audio file.
        - The timestamp when this pair will start in the video.

Selecting, Deleting, and Reordering Pairs

1. **Selecting a Pair**

    - Click on any part of the pair's display block (image, audio path, or timestamp) to select it.
    - The selected block will be highlighted in blue.

2. **Deleting a Pair**

    - Select the desired pair.
    - Press the "Delete" button to remove it from the list.

3. **Reordering Pairs**

    - Pairs are automatically sorted by their start timestamp. Ensure your timestamps are set correctly to maintain the desired order.

Creating the Video

1. **Generating the Video**

    - After adding all desired pairs, click the "Create Video" button.
    - The application will process the image and audio files, combining them into a video.

- A file dialog will prompt you to select a save location and file name for the output video. The default format is MP4.

2. **Handling Overlaps and Gaps**

    - If there is a gap between the end of one image/audio pair and the start of the next, the video will show the previous image with no audio during the gap.
    - If pairs overlap, the audio from the later pair will start at its specified timestamp, and the image from the later pair will be shown.

Optimizing Performance

- Ensure your system has sufficient resources (CPU, RAM) to handle video processing.
- Larger videos with more pairs or high-resolution images and high-bitrate audio files may take longer to process.

Tips and Best Practices

- **Use High-Quality Images and Audio**: For the best results, use high-quality images and audio files. Low-resolution images and low-bitrate audio can reduce the overall quality of the final video.
- **Plan Your Timestamps**: Plan the start times for each pair carefully to ensure a smooth and coherent video. Avoid overlapping timestamps unless intentionally creating a specific effect.
- **Review and Edit**: After generating a video, review it to ensure everything is synchronized correctly. If adjustments are needed, make the necessary changes and regenerate the video.