VAgyojaka: ASR Post Editor Documentation

| Vagyojaka: ASR Post Editor | 1 |
|---|----|
| Documentation | 1 |
| Overview | 5 |
| Installation | 5 |
| Prerequisites | 5 |
| Windows | 5 |
| Other Operating Systems | 5 |
| 1. Getting Started | 6 |
| Vagyojaka and Setup | 6 |
| Features of Vagyojaka | 6 |
| Launching Vagyojaka | 6 |
| Loading an ASR Transcript | 6 |
| Navigating the Interface | 6 |
| 2. Features | 7 |
| Media Player | 7 |
| Dictionary-based Lookup | 7 |
| Annotation and Tagging | 7 |
| Interactive Visualization | 7 |
| Saving in Multiple Formats | 7 |
| Custom Dictionary | 8 |
| ASR Validation | 8 |
| Waveform-based Editing | 8 |
| Customizable UI | 8 |
| Hotkeys | 8 |
| 3. Editor Tools | 9 |
| • Split Line (Ctrl + ;) | 9 |
| Merge Up / Merge Down ((Alt + Up) / (Alt + Down)) | 9 |
| Toggle Word Editor (Ctrl + W) | 9 |
| Change Speaker (Ctrl + R) | 9 |
| Change Transcript Language | 9 |
| Edit Tags (Ctrl + ') | 9 |
| Add Custom Dictionary | 9 |
| Auto Save | 9 |
| 4. Translation Tab | 10 |
| Manage and Edit Multilingual Transcripts Side-by-Side | 10 |
| 5. Waveform Tab | 11 |
| Accessing the Waveform | 11 |
| Displaying Transcript End Lines | 11 |

| Editing Timestamps Visually | 11 |
|---------------------------------------|----|
| Interactive Block Navigation | 11 |
| 6. ASR Validation Tab | 12 |
| Opening Transcripts and Loading Audio | 12 |
| Columns and Usage | 12 |
| Audio Column | 12 |
| Transcript Column | 12 |
| Mispronounced Words Column | 12 |
| Tags Column | 13 |
| ASR Quality Column | 13 |
| 7. Usage | 14 |
| 7.1. Menubar | 15 |
| File Menu | 15 |
| Edit Menu | 16 |
| View Menu | 18 |
| Player Menu | 19 |
| Editor Menu | 20 |
| Waveform Menu | 22 |
| Translation Menu | 23 |
| Git Menu | 24 |
| ASR Validation Menu | 25 |
| Help Menu | 26 |
| About Menu | 28 |
| 7.2. Open Video/Audio Option | 29 |
| 7.3. Open Transcript Option | 31 |
| Loading Media and Transcript | 32 |
| 7.4. Tabs | 33 |
| 7.5. Media Player & controls | 33 |
| Media Player | 33 |
| Controls | 33 |
| 7.6. Status Bar | 33 |
| 7.7. Font Options | 34 |
| 7.8. Editor | 34 |
| o Edited Words | 34 |
| o Invalid Words | 34 |
| ○ Mark as Correct (Ctrl + M) | 34 |
| ○ Saving Transcript | 35 |
| 7.9. Taglist Viewer | |

| Development Section | . 36 |
|---------------------|------|
| Github Repo | |
| API Overview | 36 |

Overview

Vagyojaka is an open-source post-editing and annotation tool for automatic speech recognition (ASR), designed to reduce the human effort required to correct ASR results. It uses a dictionary-based lookup method to highlight incorrect words in ASR transcripts and offers suggestions by generating the closest valid words. To curate a speech corpus, it provides a comprehensive set of tags to capture various spoken audio features. In a user study, post-editing with Vagyojaka was shown to take one-third less time compared to editing without the tool. The ASR Post Editor is tailored for editing transcripts generated by AI models for Indian native languages, significantly reducing time spent interacting with media files by offering an interactive visualization.

Installation

Prerequisites

- FFMPEG (Optional)
- Codec pack (Required to play different media formats)
 - o K-lite is recommended codec pack

Windows

1. Download the Vagyojaka Installer

Visit the Vagyojaka Releases Page and download the latest Windows .exe installer.

2. Run the Installer

Double-click the downloaded .exe file and follow the installation prompts. The installer will automatically handle dependencies.

3. Launch Vagyojaka

After installation, you can launch Vagyojaka from your desktop or the Start menu.

Other Operating Systems

For Linux, macOS, and other environments, please refer to the **Development Section** for detailed instructions on setting up Vagyojaka from source.

1. Getting Started

Vagyojaka and Setup

This section will guide you through the initial steps for setting up and using Vagyojaka. Here's a brief overview of what it will cover

Features of Vagyojaka

Overview of Vagyojaka and its purpose as an ASR post-editing and annotation tool.

Launching Vagyojaka

Instructions on how to open and navigate Vagyojaka after installation, including any initial settings or configuration recommendations.

Loading an ASR Transcript

A guide to loading your first ASR-generated transcript file and any accompanying audio files, preparing for post-editing and validation.

Navigating the Interface

Briefly introduce key tabs and menu options (like **Media Player**, **Transcript Editing**, **ASR Validation**, **Waveform Tab**, **Git**, etc.), so users know where to begin.

2. Features

Media Player

- Interactive Control from Transcripts Timestamps
 Allows users to click on any word in the transcript to jump directly to the corresponding timestamp in the audio, making it easy to verify and correct specific parts.
- Speed Control
 Adjust playback speed to speed up or slow down the audio, helping users edit at a comfortable pace.
- Playback Controls
 Includes standard controls like play, pause, rewind, and fast-forward, enabling precise navigation through audio files.

Dictionary-based Lookup

 This feature highlights incorrect words in the ASR-generated transcript and suggests replacements from a dictionary, saving time by offering possible corrections without needing to search externally.

Annotation and Tagging

 Annotate the transcript with tags to mark features such as speaker identity, background noise, or tone changes. This helps in categorizing and organizing the transcript based on specific audio characteristics.

Interactive Visualization

• Displays the audio waveform alongside the transcript, allowing users to see visual cues of the audio for more precise navigation and editing. Users can also interact with the waveform to guickly move through the audio.

Saving in Multiple Formats

 Export edited transcripts in various formats like XML, TXT, or PDF to ensure compatibility with different tools and workflows.

Custom Dictionary

 Build and use a custom dictionary, particularly helpful for specialized terminology or proper nouns that aren't commonly found in general dictionaries, improving the accuracy of ASR corrections.

ASR Validation

• Provides tools for efficient verification of ASR outputs, allowing users to confirm and correct the ASR transcript quickly, especially useful for multiple transcripts.

Waveform-based Editing

• Use the waveform to directly interact with and edit the audio, helping users make precise adjustments to the audio sections aligned with the transcript.

Customizable UI

 Adjust the layout, theme, and display options within Vagyojaka to tailor the interface to personal preferences, providing a more comfortable editing environment.

Hotkeys

• Includes customizable hotkeys for quick access to essential tools and actions, allowing users to streamline their workflow and work more efficiently.

3. Editor Tools

These tools offer advanced options for organizing and refining the transcript:

• Split Line (Ctrl + ;)

Divide a line into two segments to make specific edits to individual parts of a sentence or thought.

Merge Up / Merge Down ((Alt + Up) / (Alt + Down))

Combine lines to unify related sections in the transcript, improving flow and structure.

• Toggle Word Editor (Ctrl + W)

Activate a word-by-word editor for precise editing of individual words without affecting the entire sentence.

Change Speaker (Ctrl + R)

Assign or modify the speaker label for sections in multi-speaker transcripts, making it easy to track who is speaking.

• Change Transcript Language

Adjust the language setting to tailor dictionary suggestions and tagging options for a specific language.

• Edit Tags (Ctrl + ')

Add or modify tags within the transcript to annotate features like noisy, multiple speaker, or different language.

• Add Custom Dictionary

Expand the dictionary with specific terms, including names or technical language, to improve accuracy in ASR correction.

Auto Save

Enables automatic saving at regular intervals, helping prevent data loss and ensuring the work is always up-to-date.

4. Translation Tab

• Manage and Edit Multilingual Transcripts Side-by-Side

Enables users to open multiple transcripts in different languages and edit them side-by-side. This feature supports seamless translation, comparison, and editing of transcripts, making it easier to manage multilingual content and ensure consistency across translations.

5. Waveform Tab

Visual Audio and Transcript Alignment for Precise Timestamp Editing

The Waveform Tab allows users to view and interact with the audio waveform of a media file, facilitating precise timestamp alignment for transcript blocks.

Accessing the Waveform

After opening a media file, select the waveform option from the menu bar and click "Show Waveform" to process the audio and display its waveform.

Displaying Transcript End Lines

Once a transcript is loaded, enable "Show Endlines" from the waveform menu to display individual lines representing transcript blocks.

Editing Timestamps Visually

Right-click and drag end lines to adjust transcript block timings directly on the waveform. This helps in aligning timestamps efficiently, allowing users to remove silent audio sections and refine timestamps visually.

Interactive Block Navigation

Click on any line in the waveform to highlight corresponding transcript lines, simplifying navigation through the transcript and enhancing editing accuracy.

6. ASR Validation Tab

Review and Validate ASR Transcript Quality and Audio Content

The **ASR Validation Tab** in Vagyojaka enables users to review ASR-generated transcripts alongside audio files, assess transcript accuracy, and annotate various quality metrics for improved ASR performance analysis.

Opening Transcripts and Loading Audio

From the **ASR Validation menu bar**, select **Open** to load transcripts and automatically fetch audio files from the current directory. If the opened transcript file includes audio file names in each row and matching audio files are present in the directory, they will appear in the first column. This setup allows users to play and review each audio file along with its transcript for efficient validation.

Columns and Usage

Audio Column

Lists the loaded audio files corresponding to the transcript. Users can play each audio file to listen while reviewing and validating the transcript.

• Transcript Column

Displays the ASR-generated transcripted words, allowing easy follow-along with audio playback.

• Mispronounced Words Column

Allows users to mark words identified as mispronounced during playback. Mispronounced words can be listed here, with multiple words separated by spaces.

• Tags Column

Users can add content and speaker tags to categorize the audio more effectively. Common tags include:

- Content Tags: For context, such as educational for learning-related content.
- Speaker Tags: Specify male, female, or multi if there are single or multiple speakers. Separate multiple tags using semicolons.

• Sound Quality Column

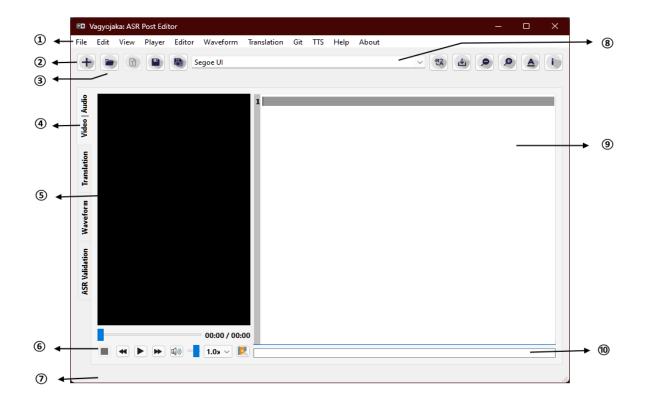
Rate audio clarity from **1** (unclear) to **5** (very clear), helping to gauge how well the audio quality supports accurate transcription.

ASR Quality Column

Rate ASR transcription accuracy based on the number of mispronunciations:

- 1 indicates high ASR quality with minimal errors.
- **0** indicates significant ASR errors, requiring more corrections.

7. Usage

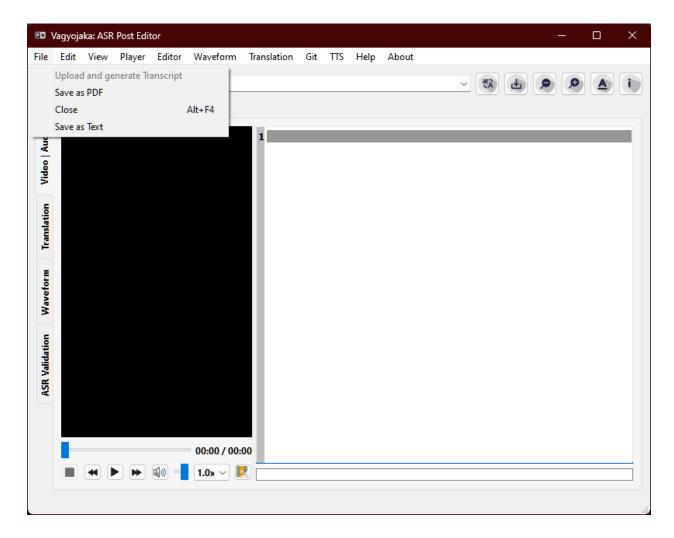


- 7.1. Menubar
- 7.2. Open Video/Audio Option
- 7.3. Open Transcript Option
- 7.4. Tabs
- 7.5. Mediaplayer & controls
- 7.6. Statusbar
- 7.7. Font options
- 7.8. Editor
- 7.9. Taglist Viewer
- 7.10. Waveform
- 7.11. ASR Validation

7.12. Git

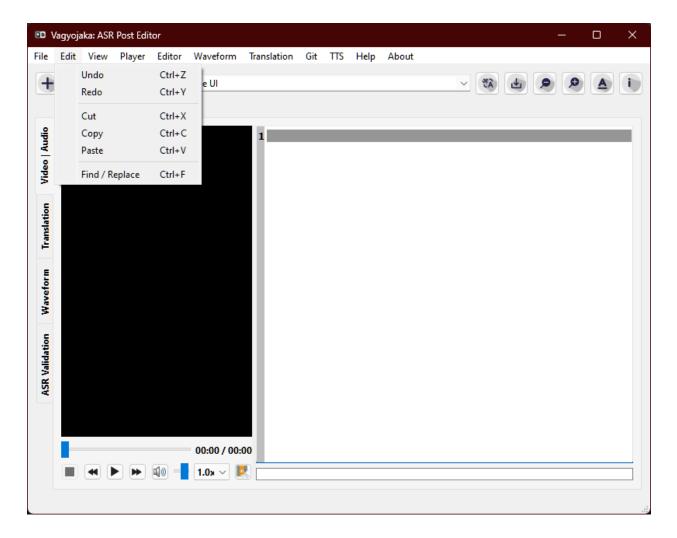
7.1. Menubar

File Menu

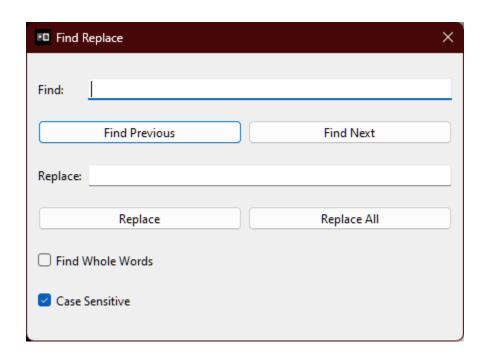


- Save as PDF: Save the current transcript as a PDF file for easy sharing and printing.
- Close: Exit the application, closing all open files and media.
- Save as Text: Save the transcript as a plain text (.txt) file, ideal for basic text storage and compatibility.

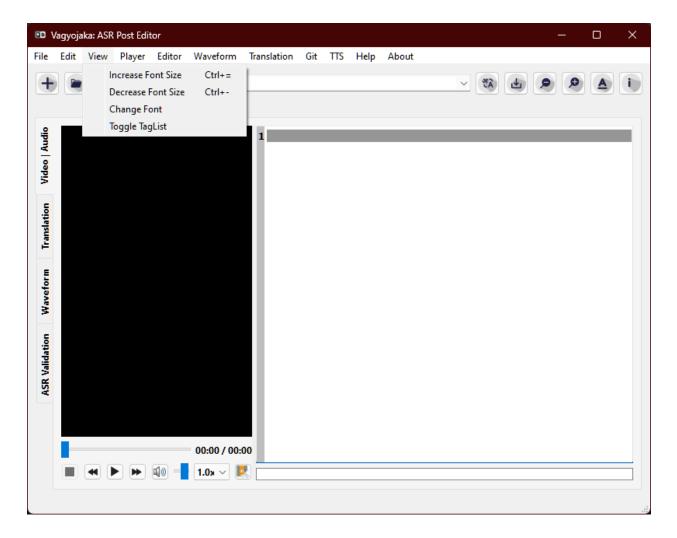
Edit Menu



- Undo: Revert to the previous state in the editor, undoing the last change.
- Redo: Return to a state that was undone, reapplying the last undone action.
- Cut: Remove the selected text from the editor and place it on the clipboard.
- **Copy**: Copy the selected text to the clipboard without removing it from the editor.
- Paste: Insert text from the clipboard into the editor at the cursor's location.
- **Find / Replace**: Search for specific words or phrases within the editor and replace them as needed.

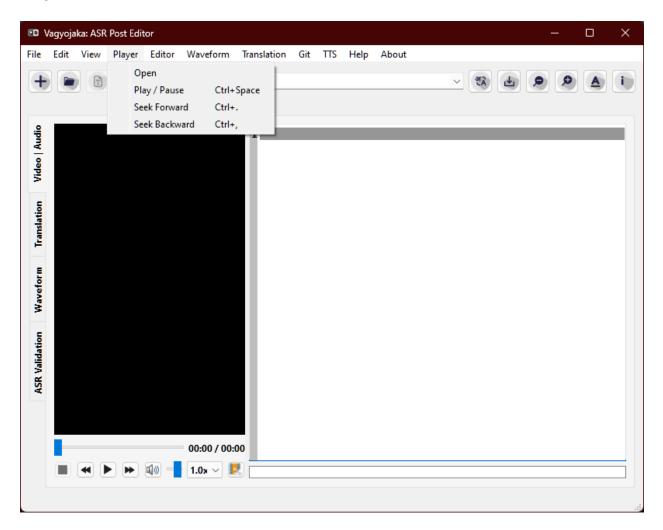


View Menu



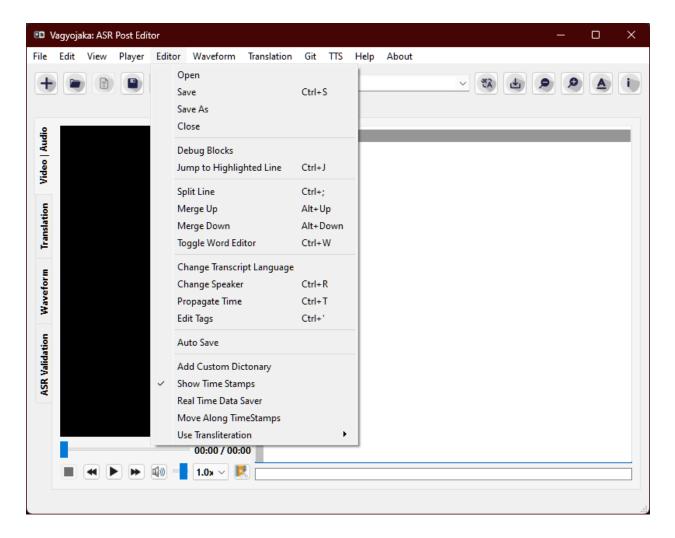
- Increase Font Size: Increase the font size in the editor for better readability.
- **Decrease Font Size**: Decrease the font size in the editor to fit more content on the screen.
- **Change Font**: Select a different font to use in the editor for a customized appearance.
- **Toggle Taglist**: Toggle the visibility of the taglist from the menu bar (10) from the fig.)

Player Menu



- **Open**: Load a video or audio file into the media player for playback and editing.
- **Play/Pause**: Toggle between playing and pausing the currently loaded media.
- **Seek Forward**: Advance the media playback by 3 seconds for quick navigation.
- Seek Backward: Move back 3 seconds in the media to review previous sections.

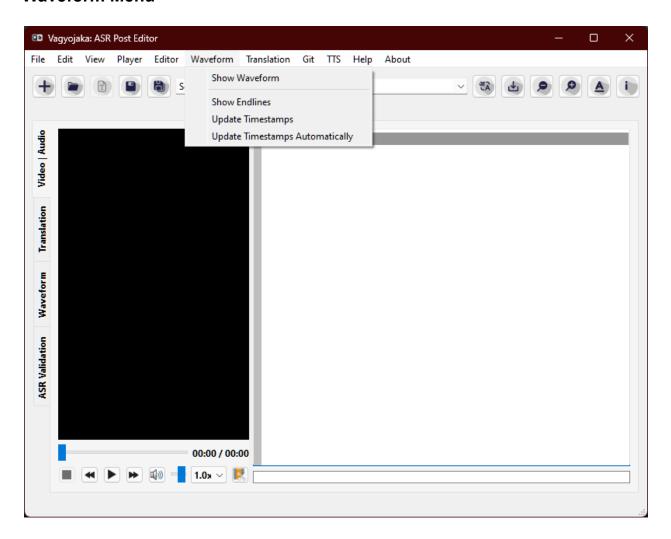
Editor Menu



- Open: Load a new transcript file and corresponding audio for editing.
- Save: Save changes to the current transcript.
- Save As: Save the current transcript as a new file.
- Close: Close the current file and return to the main interface.
- Jump to Highlighted Line: Quickly navigate to highlighted sections of the transcript.
- Split Line: Divide a transcript line at the cursor position to create a new line.
- Merge Up: Merge the current line with the line above it.
- Merge Down: Merge the current line with the line below it.
- **Toggle Word Editor**: Enable editing for individual words within a transcript block.

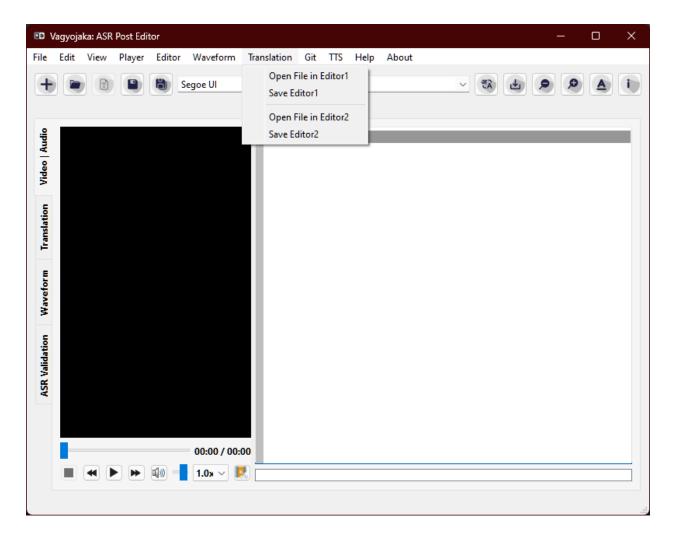
- **Change Transcript Language**: Switch the language setting for the current transcript.
- Change Speaker: Assign or change the speaker label for a specific line.
- Edit Tags: Add or edit tags that categorize content within the transcript.
- **Auto Save**: Enable automatic saving of transcript changes at regular intervals.
- Add Custom Dictionary: Include specialized words or terms in a custom dictionary to improve ASR accuracy.
- **Show Timestamps**: Display timestamps for each line to aid in audio alignment.

Waveform Menu



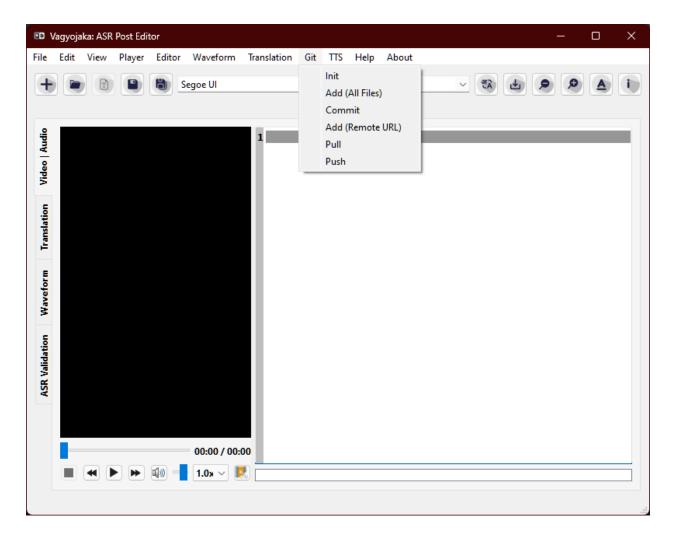
- **Show Waveform**: Display the audio waveform of the loaded media file, allowing for precise visual alignment with the transcript.
- **Show Endlines**: Reveal end-of-line markers on the waveform, representing transcript blocks, to assist with line-by-line editing and timing.
- **Update Timestamps**: Manually adjust timestamps for each transcript block to improve synchronization with the media.
- Update Timestamps Automatically: Enable automatic timestamp updates to maintain alignment between the audio and transcript as you edit.

Translation Menu



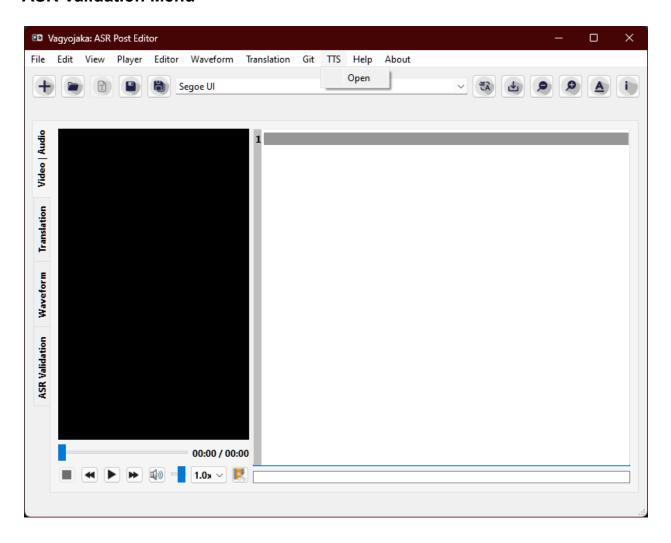
- Open File in Editor 1: Load a file into Editor 1 for editing, allowing for side-by-side or split-screen comparisons.
- Save Editor 1: Save changes made in Editor 1 to the current file.
- Open File in Editor 2: Load a file into Editor 2 for parallel editing, useful for comparing different versions or translations.
- Save Editor 2: Save changes made in Editor 2 to the current file.

Git Menu



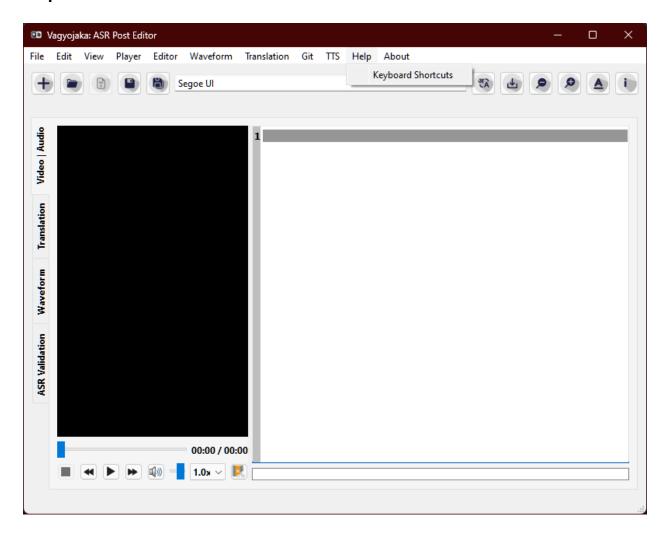
- Init: Initialize a new version control repository for tracking changes to your files.
- Add (All Files): Stage all files in the current directory for the next commit.
- **Commit**: Save a snapshot of the current staged changes with a descriptive message.
- Add (Remote URL): Link a remote repository URL to your local project, enabling collaboration and backup.
- Pull: Fetch and merge updates from the remote repository into your local project.
- Push: Upload local commits to the remote repository, sharing changes with collaborators.

ASR Validation Menu

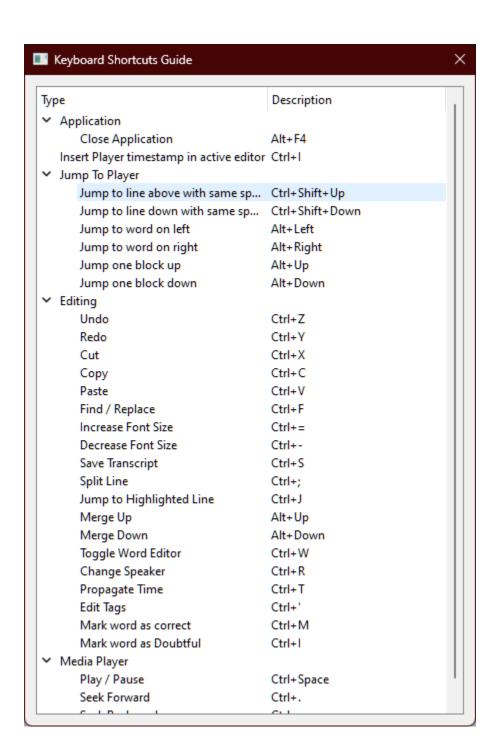


 Open: Opens a file in the ASR Validation tool, allowing users to review and validate the transcript accuracy against the audio.

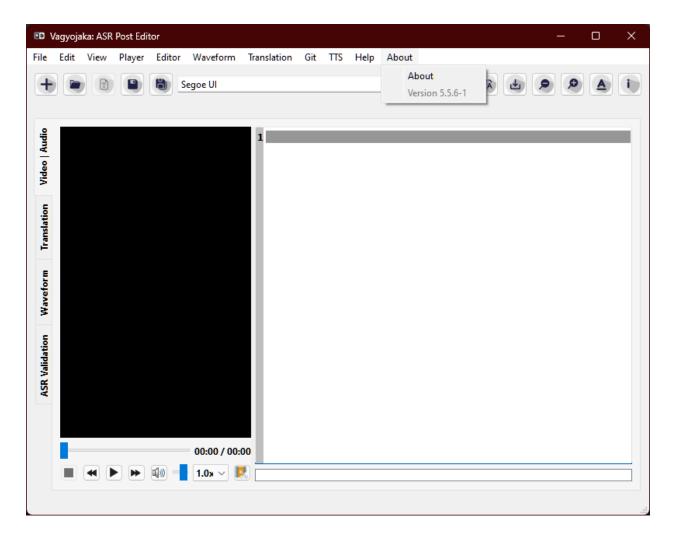
Help Menu



• **Keyboard Shortcuts**: Opens the Keyboard Shortcuts GUI, allowing users to view keyboard shortcuts for various functions within the application.



About Menu



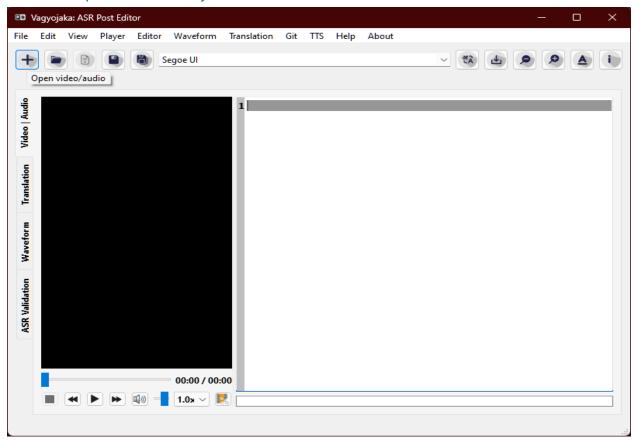
- **About**: Displays information about the Vagyojaka application, including its purpose, features, and development background.
- **Version**: Shows the current version of the application, helping users to verify if they are using the latest release.

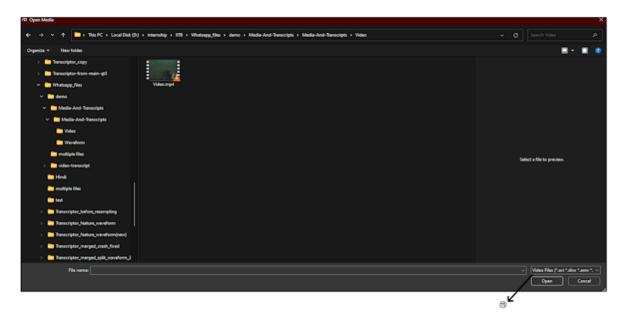
7.2. Open Video/Audio Option

To load a media file into the media player

1. Click on the + icon.

2. Or choose Open from the Player menu bar.



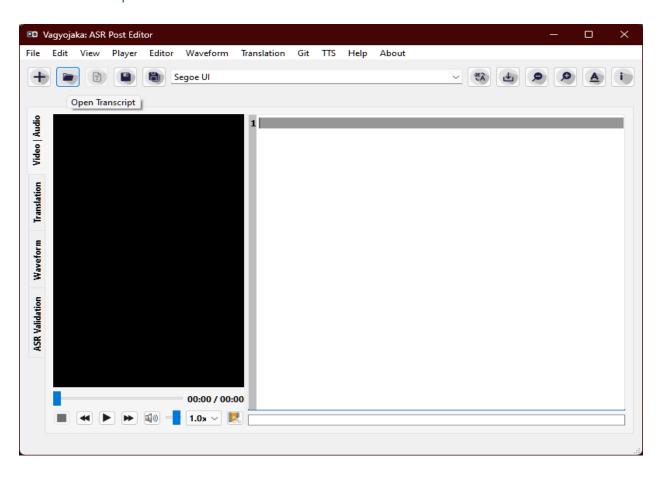


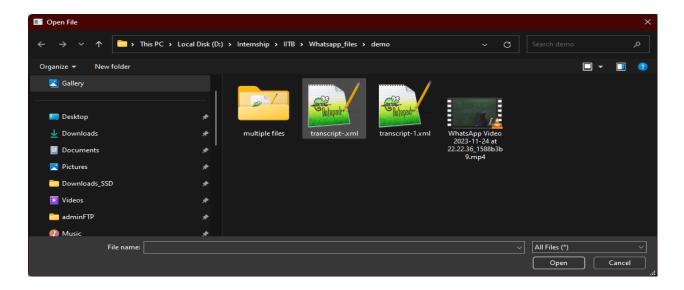
① → Option to choose different file formats

7.3. Open Transcript Option

To load a transcript file into the editor

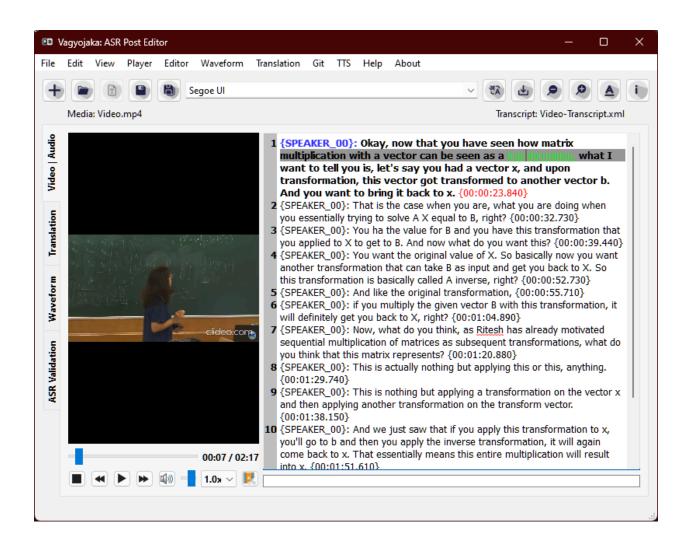
- 1. Click on the folder-like icon.
- 2. Or choose Open from the Editor menu bar.





Loading Media and Transcript

After opening both the media file and the transcript, the tool will process the data and highlight spoken words within the transcript. This enables easy identification of words as they are spoken, improving the editing and validation experience.



7.4. Tabs

Switching Between Tabs:

The tool provides a set of vertical tabs, allowing users to navigate between different functionalities. Users can switch between the tabs

- Video/Audio Player & Transcript Editor Tab
- Translation Tab
- Waveform Tab
- ASR Validation Tab

7.5. Media Player & controls

Media Player

The media player can play any media files opened within the tool, as long as they are supported by the operating system. If a selected media file fails to play, installing a codec package, such as **K-Lite**, is recommended to improve compatibility and enable seamless playback.

Controls

- **Seek Forward / Backward**: Jump forward or backward by 3 seconds for precise navigation.
- Play / Pause: Start or pause media playback using Ctrl + Space for quick control.
- Volume and Mute: Adjust the volume or mute the audio as needed.
- **Speed Control**: Change the playback speed to suit your reviewing needs.

7.6. Status Bar

Status Bar Functionality: The status bar provides real-time information about
the tool's operations. It displays messages regarding actions such as opening a
file, saving progress, or other brief status updates. Additionally, it may include
light debug information to assist users in monitoring tool activity.

7.7. Font Options

- **Font Customization:** Users can adjust the appearance of the text by selecting a different font from the dropdown menu in the tool.
- Additionally, font size can be modified to improve readability and comfort during editing. Increase font size: (Ctrl + =) & Decrease font size: (Ctrl + -)

7.8. Editor

- **Editing Transcripts**: Users can edit both opened and newly created transcripts within the Editor tab. Key features include:
 - Edited Words

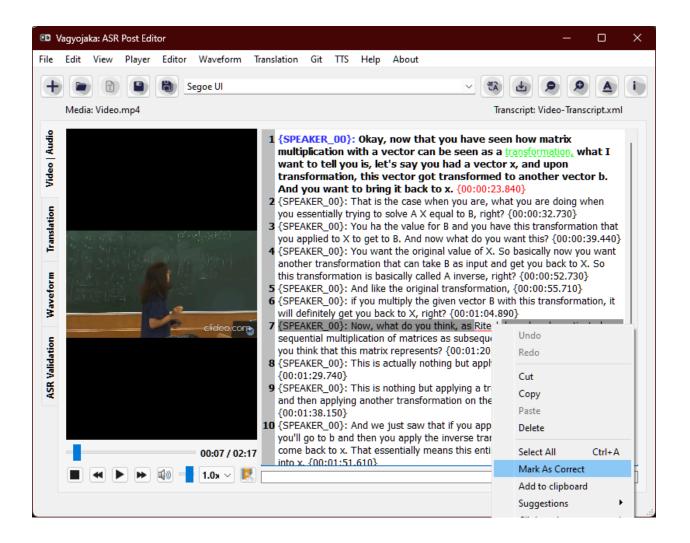
Highlighted in **yellow** for easy identification.

Invalid Words

Underlined in **red** to indicate possible errors.

Mark as Correct (Ctrl + M)

Adds the selected word to the custom dictionary, ensuring it is recognized in future sessions.



Auto-Completion

Provides word suggestions from existing dictionaries to enhance editing speed and accuracy.

Saving Transcript

Transcripts can be saved in **XML**, **PDF**, and **TXT** formats. Use **Ctrl + S** for quick saving or access the save options through the Editor menu bar.

7.9. Taglist Viewer

- Tagging Lines: Adding tags to a line block changes its color to blue, making it
 easy to identify tagged lines at a glance. Clicking on a tagged line displays the
 associated tags for quick review.
- **Toggle Taglist:** The Toggle Taglist option in the View menu bar allows users to show or hide the Taglist Viewer. To edit tags, use the shortcut Ctrl + ' (apostrophe).

7.10 Waveform

7.11 ASR Validation

ASR Validation

- **ASR Validation Overview**: This tab is dedicated to validating ASR-generated transcripts by cross-checking audio files and identifying potential errors.
- Loading Files: From the TTS menu bar, select Open to load transcripts and audio files from the current directory. Audio files matching transcript names are automatically displayed in the first column, allowing users to play and review them.

Validation Options:

- Transcript Column: Shows the transcribed words for review.
- Mispronounced Words: Users can mark mispronounced words in the transcript, with each word separated by a space.
- Tags Column: Tags can be added to indicate media characteristics (e.g., "educational," "male/female speaker," "multi-speaker"). Multiple tags are separated by a semicolon.
- Sound Quality: Rated from 1 to 5 to indicate audio clarity, where 1 represents poor clarity and 5 represents excellent clarity.
- ASR Quality: Mark 1 for high ASR quality with minimal errors, and 0 if errors are present.

Validation Guide

0. Boundary Words Check

Ensure that the first and last words of the transcript are clearly audible in the audio. If not, correct the transcript.

1. Transcript Accuracy

Verify if the transcript matches the audio.

- 1. A: If the transcript is correct, proceed as usual.
- 2. B: If the transcript contains errors, correct it.
- 3. C: If you cannot correct the transcript due to unclear speech, mark the asr-quality as 0.

2. Sound Quality Evaluation

Assess the overall sound quality based on the following scale:

- A: 1 Very poor, highly distorted or unclear audio.
- B: 2 Noticeable background noise but understandable.
- C: 3 Acceptable quality with some minor noise.
- D: 4 Good quality, clear and easy to understand.
- E: 5 Excellent, studio-like recording quality.

3. Speaker Tagging

Identify the speaker's gender:

- A: Use Male or Female as appropriate.
- B: Default is Male, but update to Female if the speaker is a female.

4. Special Tagging

In addition to gender, tag the transcript for the presence of the following:

- A: Number If there are numbers mentioned in the transcript.
- B: English Word If the transcript contains any English words.

Example tags: Male; Number, Male; English Word, Male; Number; English Word, etc.

5. Multiple Speaker Tagging

If multiple speakers are present in the audio, add Multi to the tag.

• Combine the tags as needed

Example tags: Male; Number or Male; Number; Multi or Male; Number; Multi; English Word etc.

Development Section

Github Repo

• https://github.com/ljn7/Vagyojaka

API Overview

• https://vagyojaka.vercel.app/class_tool.html