BTP-2 Final Report

SPEECH INTERFACE APPLICATION

Under the guidance of Dr. Anil Kumar Vuppala, Speech Processing Lab

Abhinav Reddy Boddu - 2021101034

Sasidhar Chavali - 2021101111

Saikiran Survi - 2021101130

LSA Mourya - 2021101121

Introduction to the existing application

1. This is a MATLAB based speech-transformation visualization tool that helps analyze speech data by providing a detailed visual representation of various speech properties. It's designed to make it easier for users to understand and explore different aspects of speech transformations.

2. This platform has important sections for plotting audio signals, examining zero-time windowing spectrogram (ZTWS), spectral flatness feature (SFF), S-transform, Constant-Q analysis, spectrogram display, identification of formant peaks, pitch contour visualization, Gammatone analysis, and detection of voice activity.

3. The application has a GUI built using MATLAB, allowing users to easily analyze different transformations and representations of audio signals in various domains.

4. This tool allows users to compare multiple audio files at the same time, making the visualization process easy. Users can adjust plot sections according to their preferences while comparing. Additionally, the tool now supports various audio file formats, not just .wav files.

5. Users can create a standalone version of the application that others can use on their own computers. They just need to install the free MATLAB Runtime to run the application locally with their own file

Issues with the previous application

1. Performance:
   The app sometimes slows up because of the heavy computations involved in processing the audio signals. This may affect the user experience and the accuracy of the visualizations. The potential cause for this might be because MATLAB might be inefficient in generating/rending GUI, and the MATLAB App Builder might be generating inefficient code.

2. Licensing:
   MATLAB is a commercial software that requires a license to use. This limits the accessibility and affordability of the application code for some users who may want to modify or extend the application's functionality. Although the App itself can be installed and run with just MATLAB Runtime (free), users cannot explore or modify the app functionality without MATLAB Licence.

3. Optimization:
   MATLAB may not exhibit the same level of performance as lower-level languages for some tasks. This may require additional effort to optimize the code and ensure the efficiency of the app.

4. User Interface:
   The user interface of the app is not very intuitive or convenient for some users. Using the MATLAB app developer to create this application makes the GUI look and feel not very good & aesthetic.

Objectives Completed

1. Cross-Platform Compatibility:
   Ensure the Python version is platform-independent, allowing users on different operating systems (Windows, Linux) to run the application seamlessly.

2. User Interface Enhancement:
   Create an intuitive and user-friendly GUI for the Python application. Gather user feedback during development to improve the design and functionality of the interface.

3. Documentation and Support:
   Provide comprehensive documentation for users, including installation instructions and usage guidelines.

4. Additional Functionality:
   Support additional useful functionality:
   a. About page with useful links to support docs, repository code, copyright marks.
   b. Ability to save the current analysis file, and be able to resume later from that point.
   c. Multigraph support: introducing the concept of Pane, each Pane corresponds to a graph type. Users can simultaneously view multiple graph types to better analyse the files. Dynamic 'Add' and 'Delete' functionality of each Pane.
   d. Simple drag and select feature to specify a subsection of the audio to analyse. Users can simply use the Red color slider to select the segment of interest, all the Panes will be shown only corresponding to that segment.
   e. Directly open all '.wav' files in the application from native file browser itself. Users need not open the application and then select file location, instead they can directly open files from windows file explorer itself.
   f. Export functionality to directly export the analysis into formats such as `PDF`, `EPS`, `PNG`, `JPG`, `JPEG`, `SVG`.
   g. Ability to set custom title, row title, column title, common axes, autolabelling features while exporting the analysis.
   h. Support for multi file analysis to compare and contrast multiple audio files simultaneously.
   i. `New Window` option to start new analysis in a quick manner.
   j. Visual depictions at multiple places to inform the users while the application is loading something.

k. Support both Menu Bar and Tool Bar for easy and quick access to all the functionality.

Methodology Used

1. Matlab Function Analysis:
   Identify the functions used in Matlab for implementing different graphs. Collect the code for all these functions.

2. Map to inbuilt functions:
   Map the matlab functions to equivalent python functions from `scipy`, `numpy`, `librosa` libraries if available

3. Rewrite/translate functions:
   If the matlab function doesnt have any equivalent python function, then translate (rewrite) the matlab function in python.

4. Basic GUI using PyQT5:
   Develop a wrapper GUI using `PyQT5` framework for the Python environment. Integrate the Python-based GUI with the developed functions.

5. Feature Enhancement:
   Implement the target features using `PyQT5` components `Matplotlib` connector .

6. Testing and Optimization:
   After each feature development, rigorous manual testing is done to ensure the functionality, performance, and stability is maintained.

7. Packaging and Bundling:
   Leverage `Github actions` for CI/CD of the application. Use `PPG` (`PyInstaller`) tool to bundle the application and generate `exe` files for windows installation, `deb` files for ubuntu installation.

Installation:

1. Download the installation files from the GitHub Repository:
   Get the files from the Releases Page of the Repo and run it.

   For ubuntu 20.04 - Waveform-Wizard_ubuntu-20_04.deb

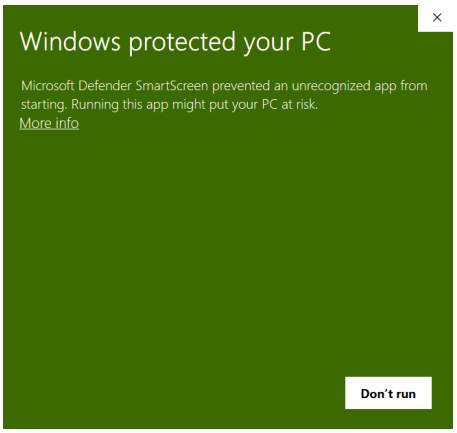
   For ubuntu 22.04 - Waveform-Wizard_ubuntu-22_04.deb

   For Windows 10 - Waveform-Wizard-Setup_windows_10.exe

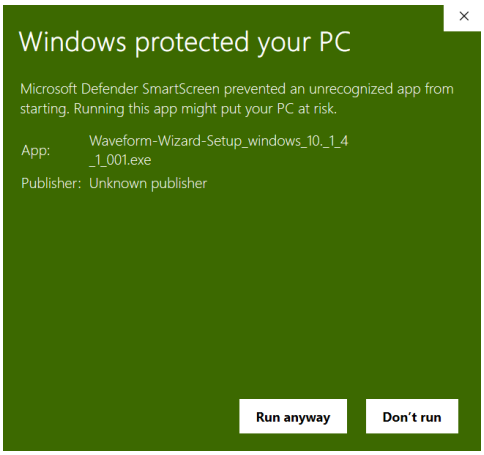
2. **Ubuntu:**
   Open the .deb with ubuntu's software installer (or) in the terminal, go to the directory containing the downloaded file and run:

   $ sudo apt install ./Waveform-Wizard_ubuntu-<version>.deb

3. **Windows:**
    a. Download the latest installation file from Releases Page.
    b. Run the .exe file
    c. Since we are not licensed developer with Microsoft, Windows doesn't detect the application as safe as shown below:

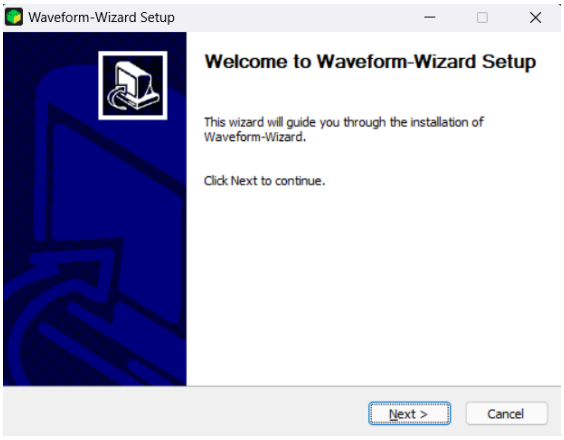


    d. Click on More info:

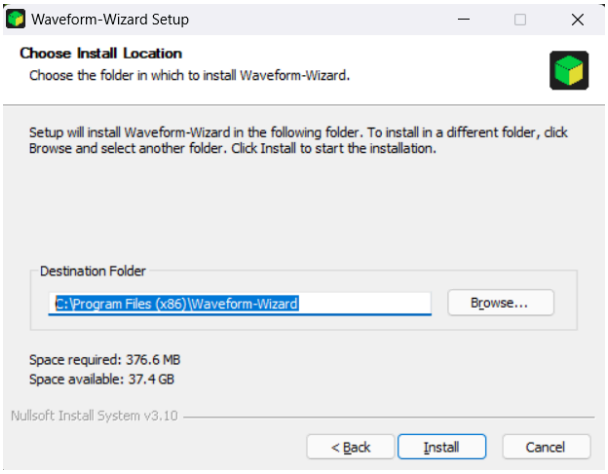


    e. Click on Run anyway.
    f. It will ask for permission to make changes to computer, Click on yes.

g. An installer will open, click next:



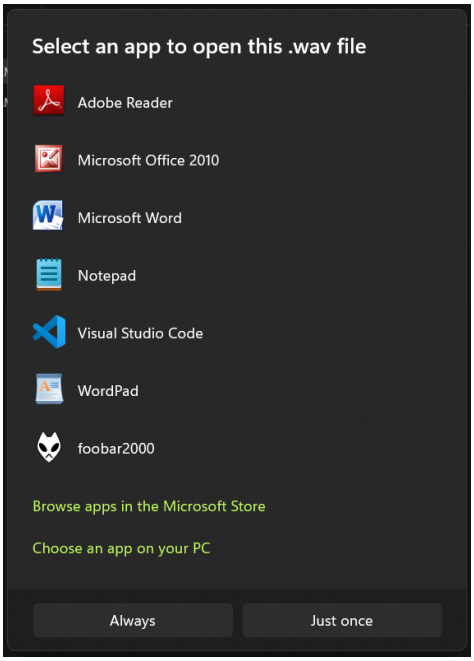
h. It will prompt for an Installation location. Change the location by clicking on Browse.. and choose the desired location. It defaults to C:\Program Files (x86)\Waveform-Wizard as shown. Then click install:



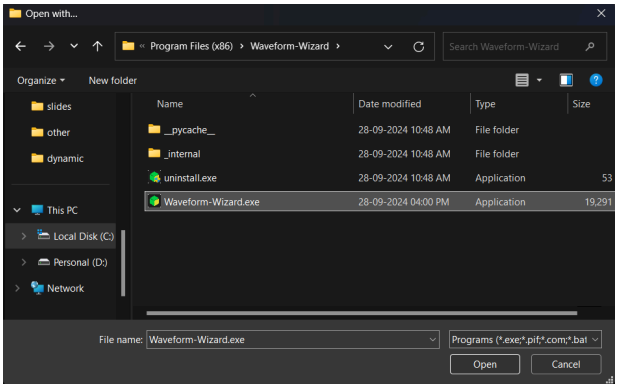
i. Wait for the installation to complete and click next.

j. click Finish.

4. **(optional) (windows)** File browser setup:
    a. Right-click on any .wav file, then click on Open With > Choose another App. A prompt appears as follows:

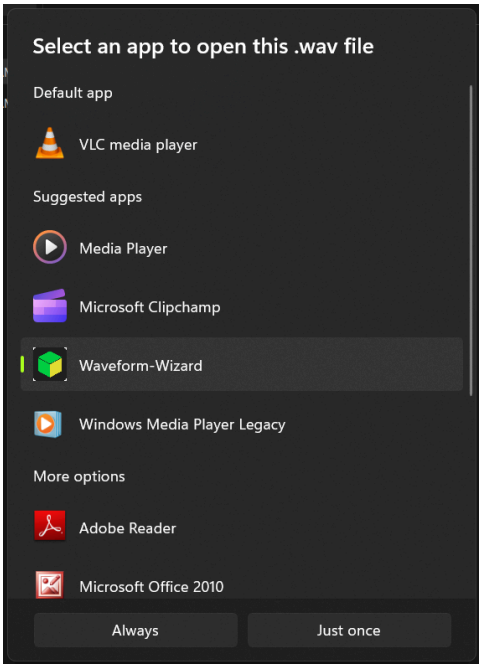


    b. Scroll down and click on `Choose an app on your PC`
    c. A file browser appears; now select <installation path>/Waveform-Wizard.exe and click Open.
    (Note: The default installation path is C:\Program Files (x86)\Waveform-Wizard, so the path to choose becomes C:\Program Files (x86)\Waveform-Wizard\Waveform-Wizard.exe.)



    d. Now you will be back to the application choosing screen. You can now see Waveform-Wizard in the list of applications; select it, and click on Just Once (or choose Always if you want to open all .wav files

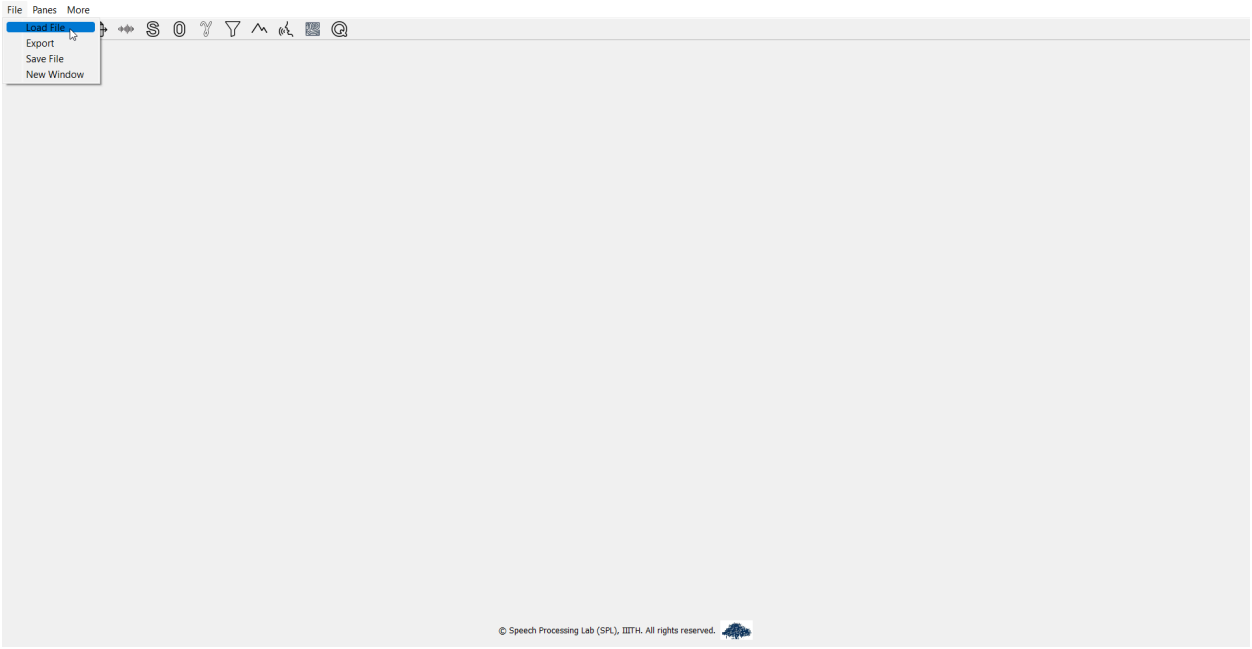
using this application).

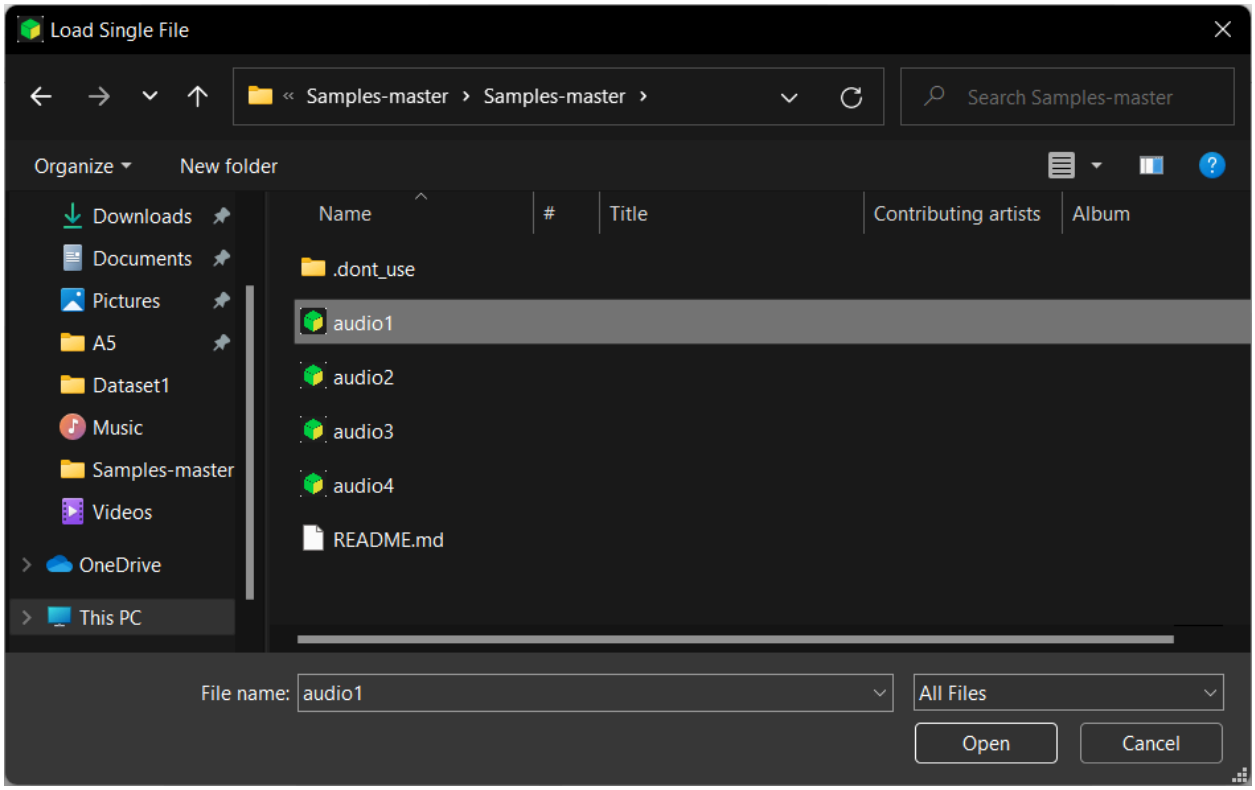


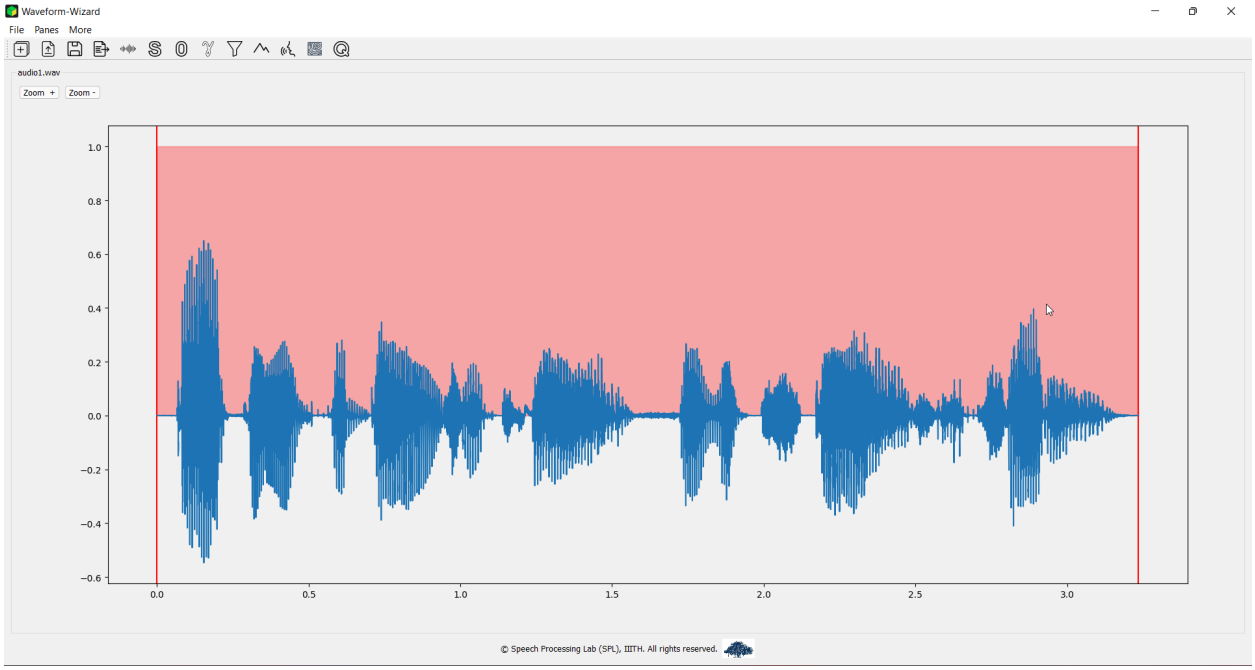
Using the Application:

1. Select an Audio File:

File -> Load File -> File picker

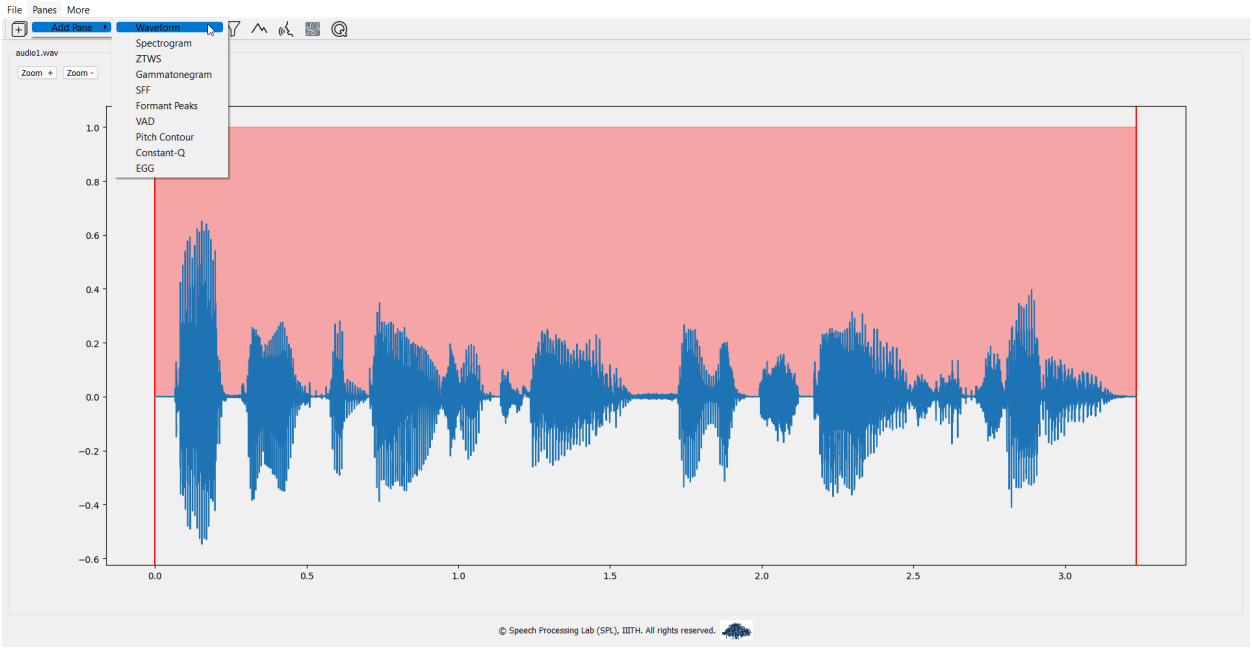


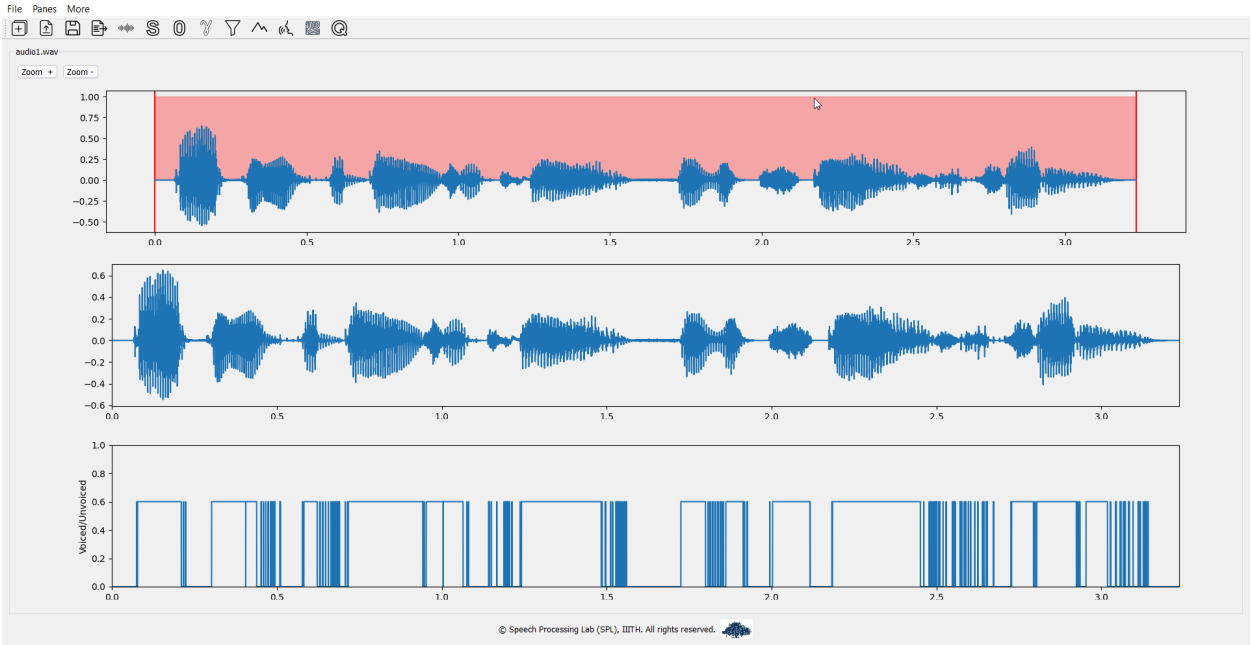

Load Single File
Samples-master › Samples-master ›
Search Samples-master
Organize
New folder
Downloads
Documents
Pictures
A5
Dataset1
Music
Samples-master
Videos
OneDrive
This PC
Name
#
Title
Contributing artists
Album
.dont_use
audio1
audio2
audio3
audio4
README.md
File name: audio1
All Files
Open
Cancel


Waveform-Wizard
File
Panes
More
audio1.wav
Zoom +
Zoom -
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
0.0
0.5
1.0
1.5
2.0
2.5
3.0
© Speech Processing Lab (SPL), IIITH. All rights reserved.

2. Add Panes:

Panes -> Add Pane -> Pane type





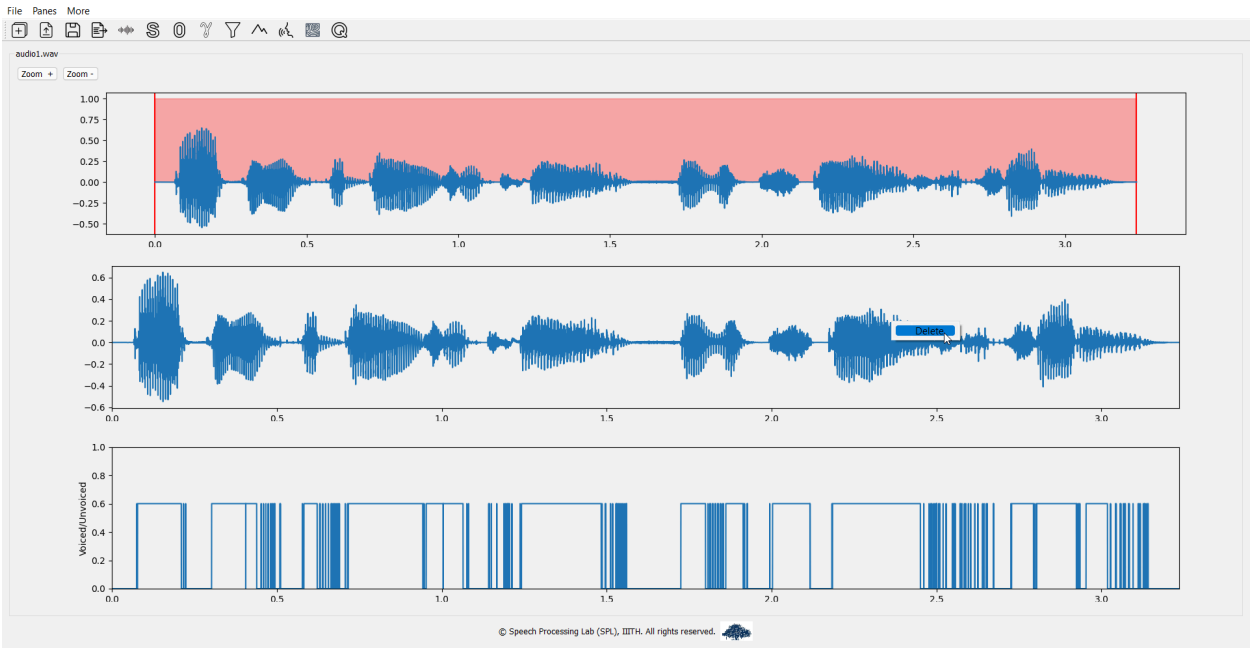
Available options

a. Waveform
b. Spectrogram
c. Zero Time Wind Spectrum
d. Gammatonegram

e. Spectral Flatness Feature (SFF)
f. Formant Peaks
g. Voice Activity Detection (VAD)
h. Pitch Contours
i. Constant-Q

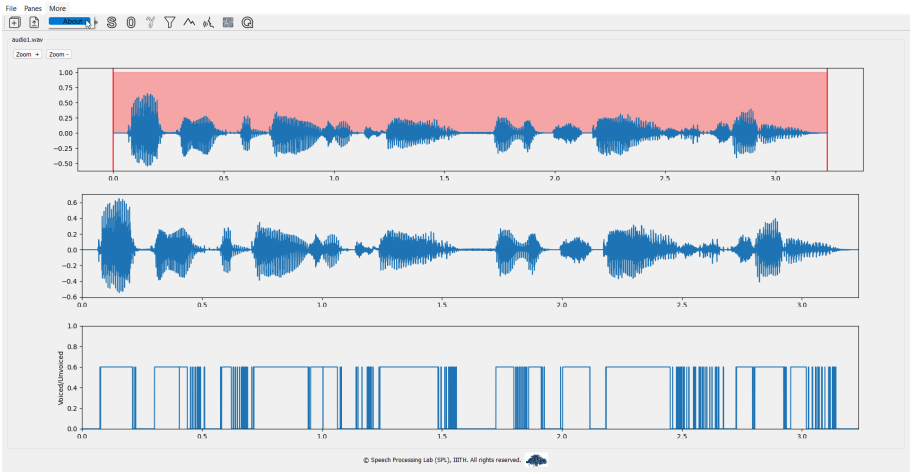
3. Delete Pane:

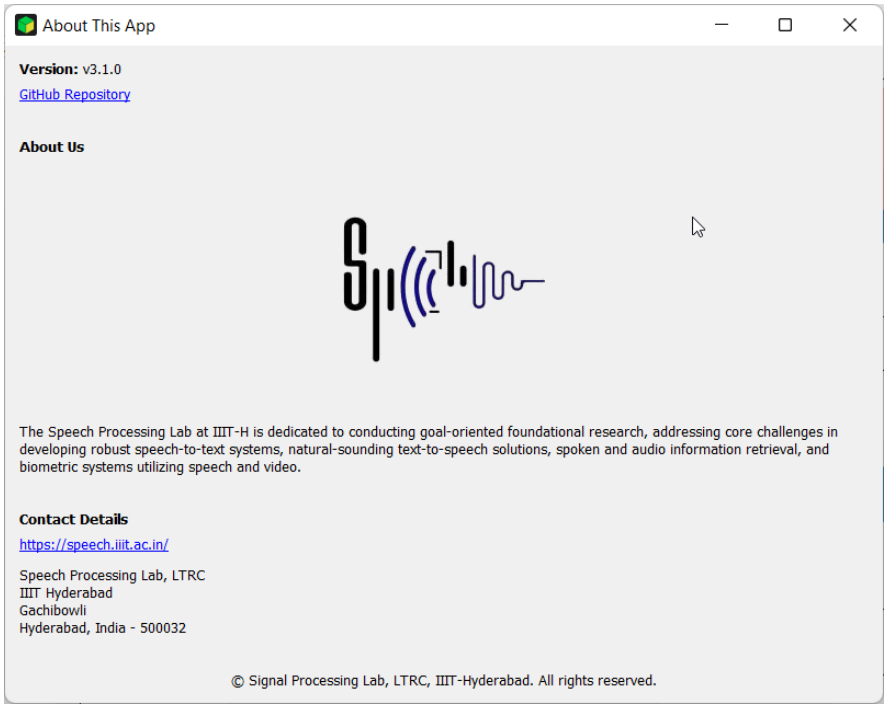
Right click on Pane -> Delete



4. About Info window:

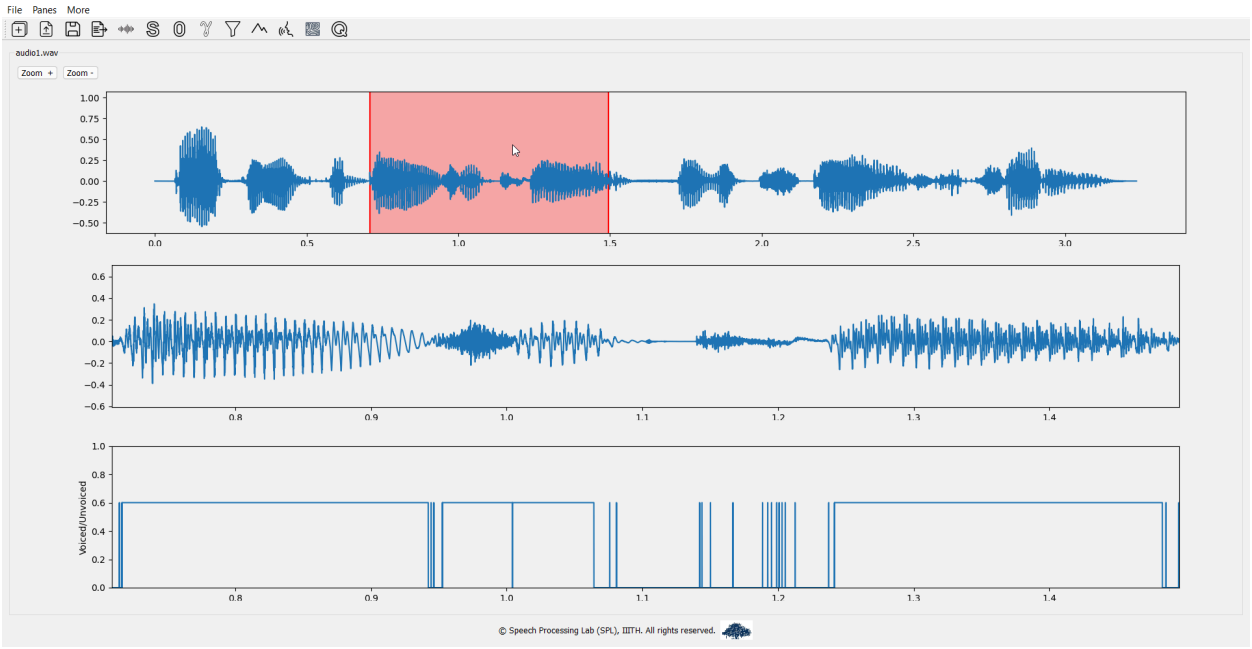
More -> About





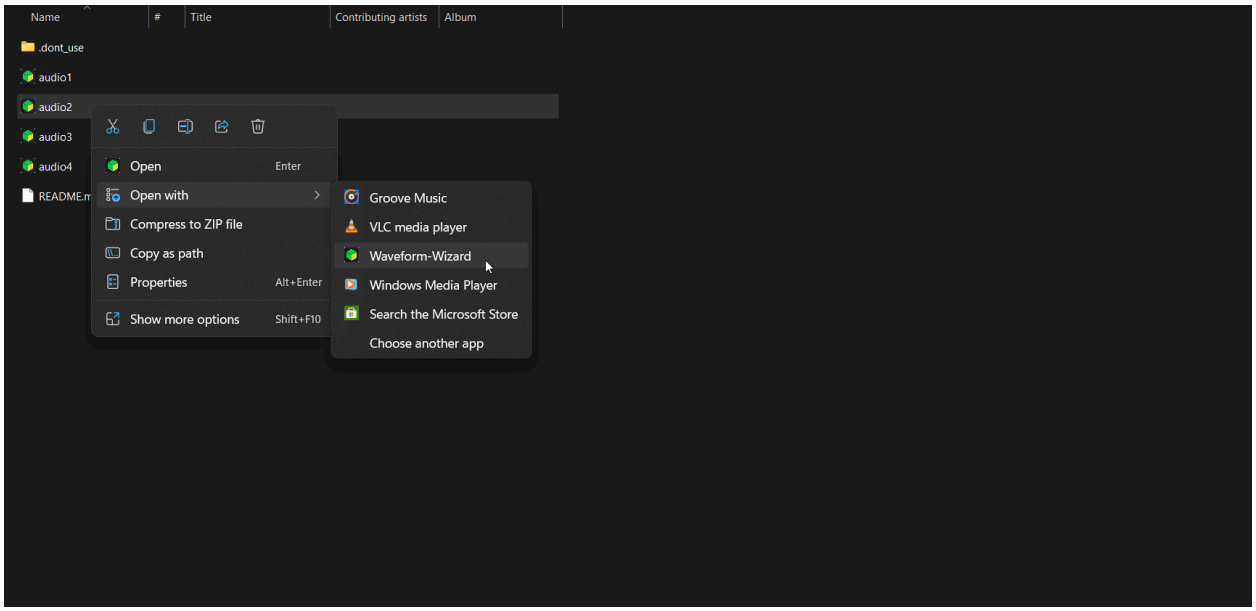
5. Drag & Drop section selector:

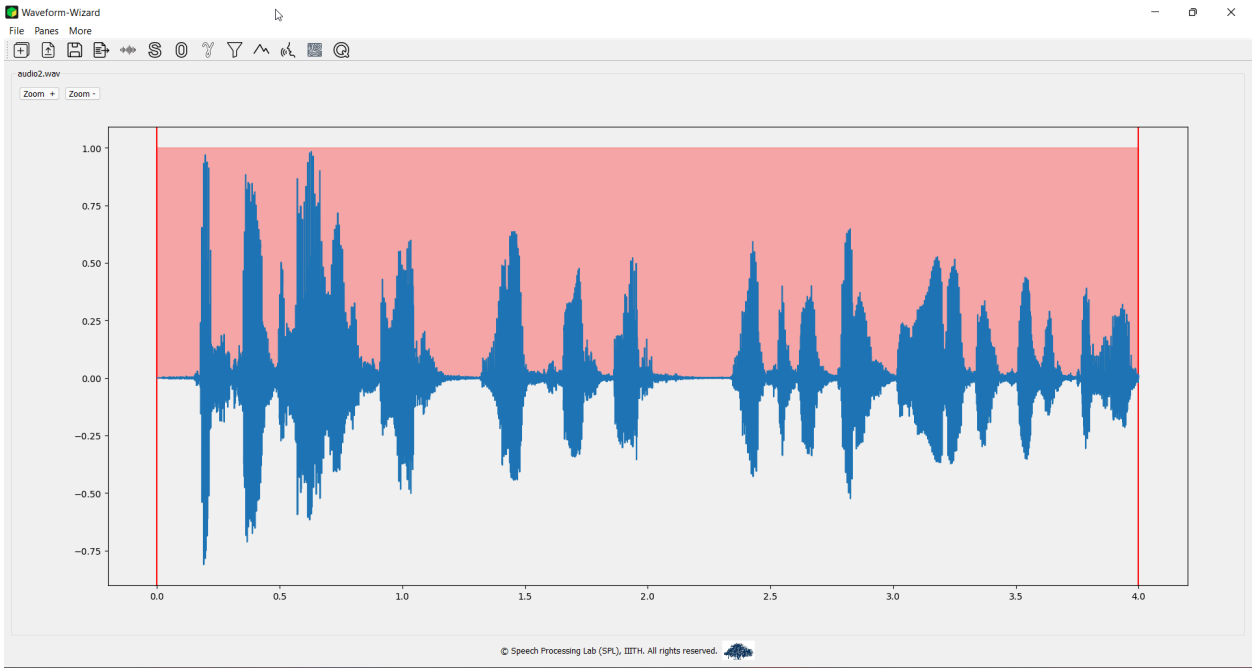
Simply select the red section, and drag to the section of interest. Use ‘Zoom +’ and ‘Zoom -’ options to increase and decrease size of the section.



6. Directly open wav files from windows file browser:

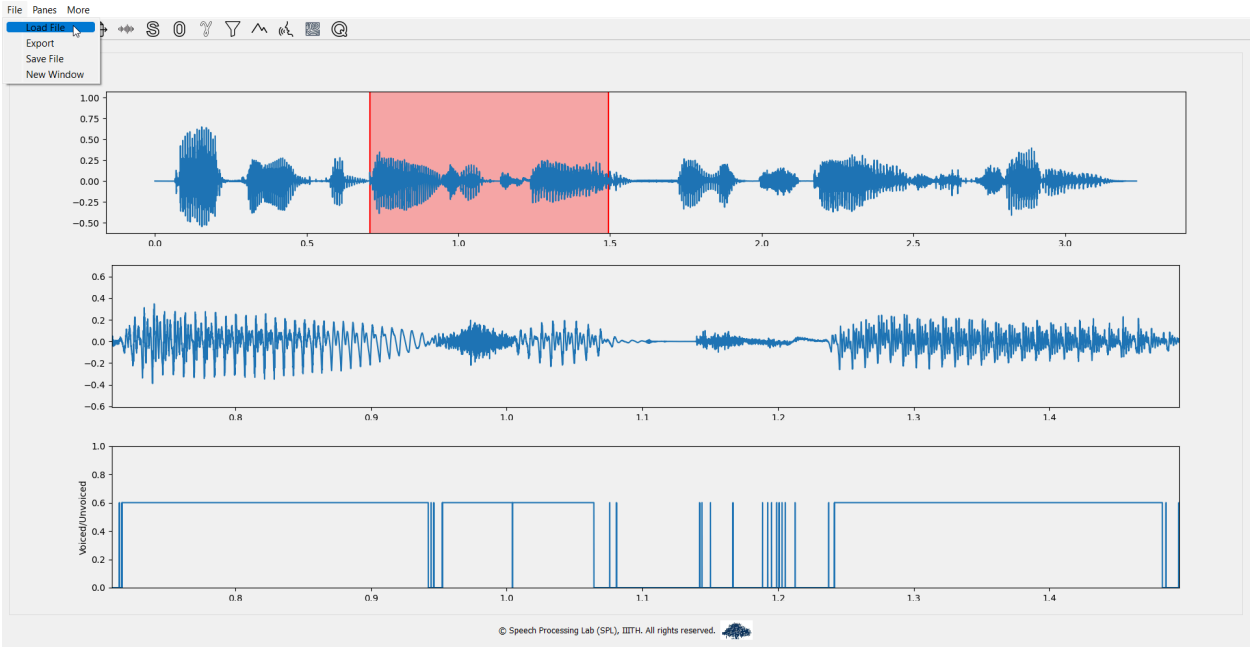
Select file -> Right click -> Open With -> Waveform-Wizard

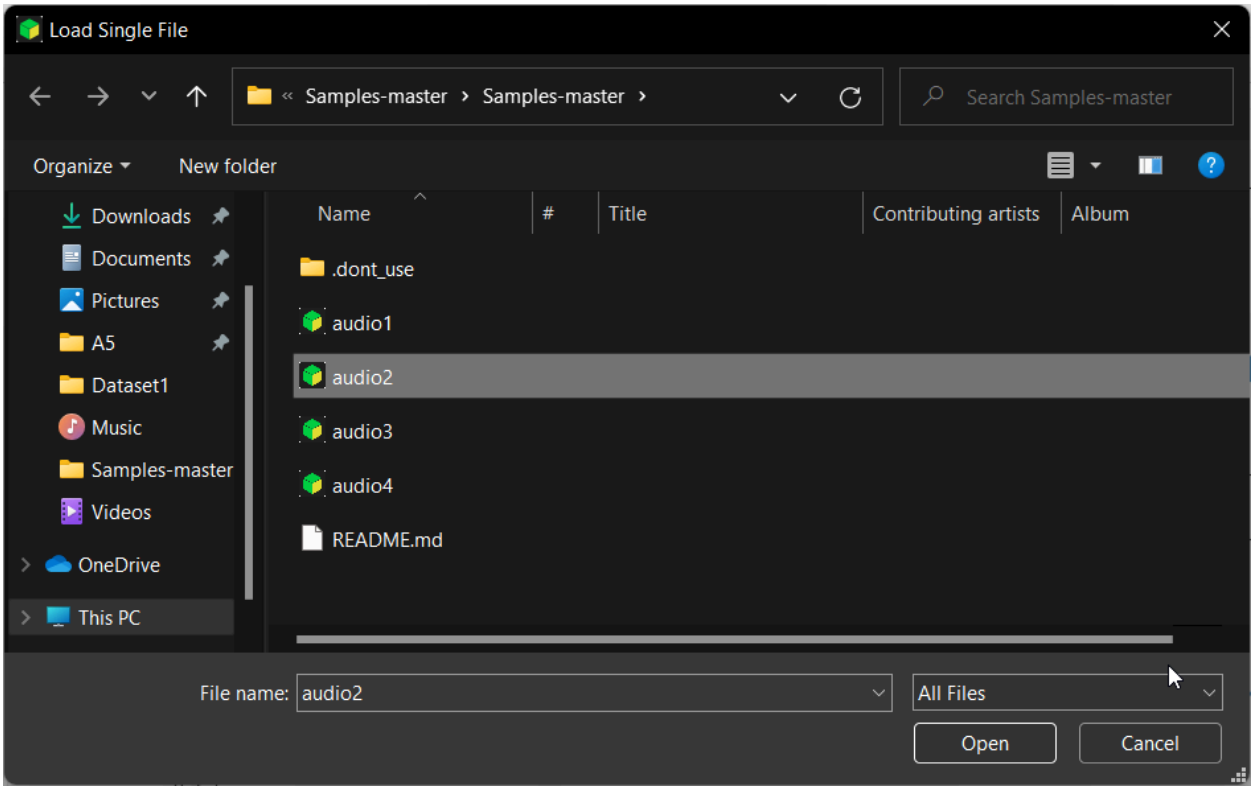


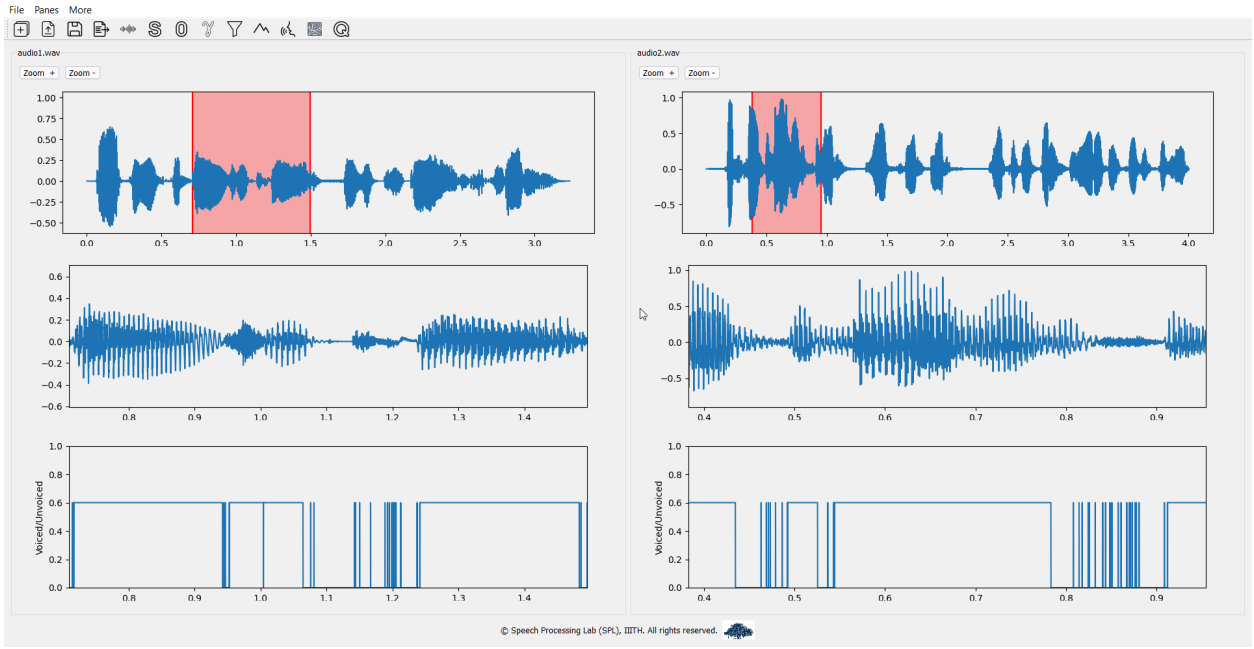


7. Multi File:

In an already opened analysis, just load another file



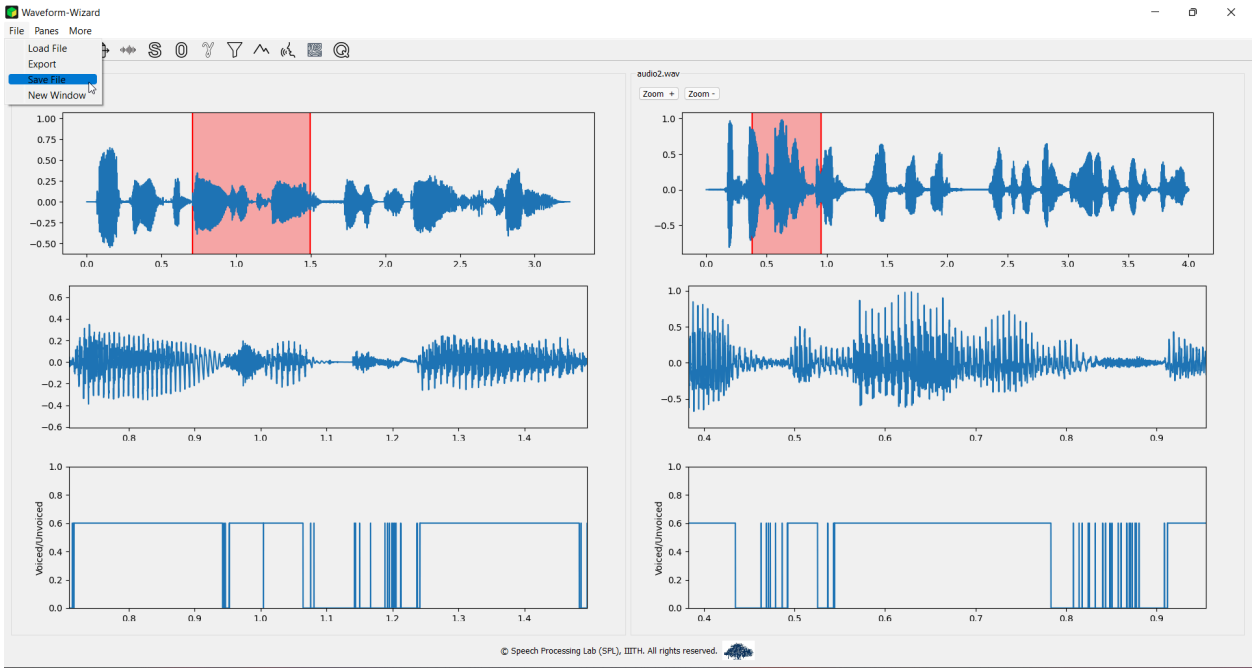


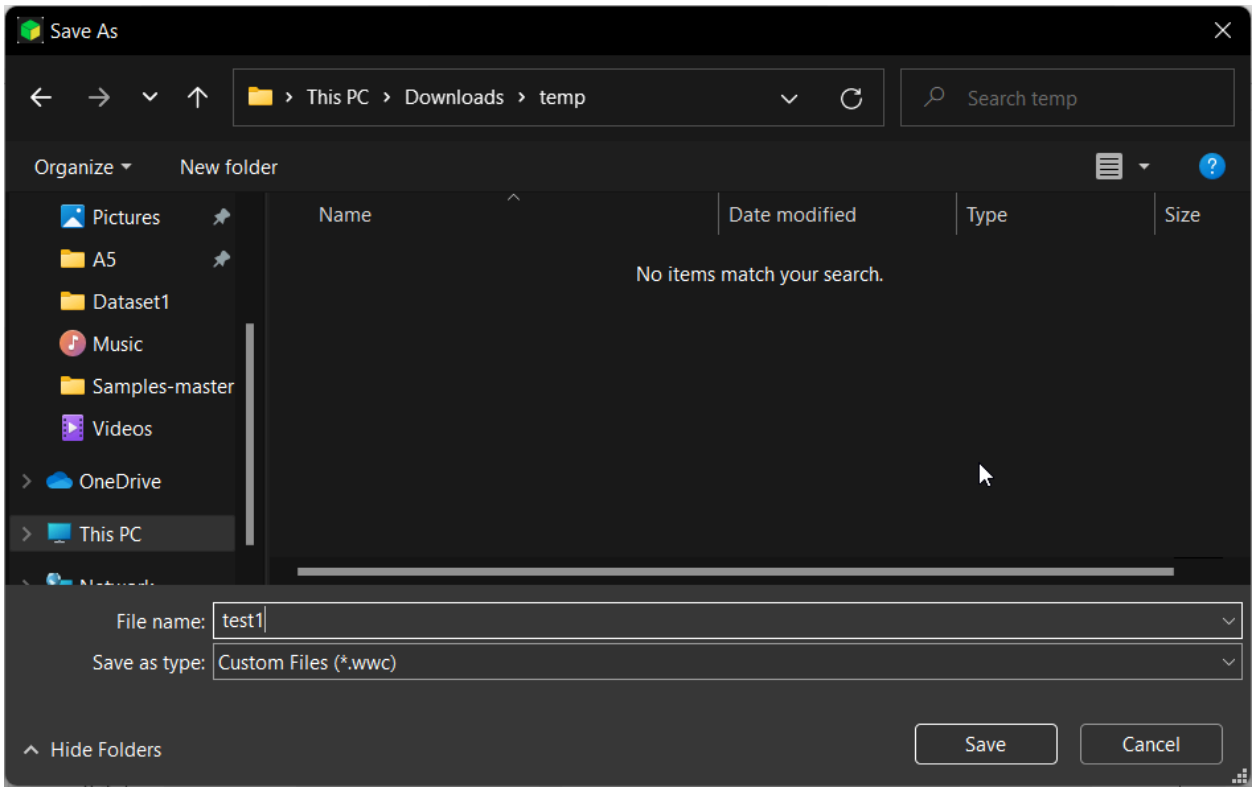

File
Panes
More
audio1.wav
audio2.wav
Zoom +
Zoom -
Voiced/Unvoiced
© Speech Processing Lab (SPL), IIITH. All rights reserved.

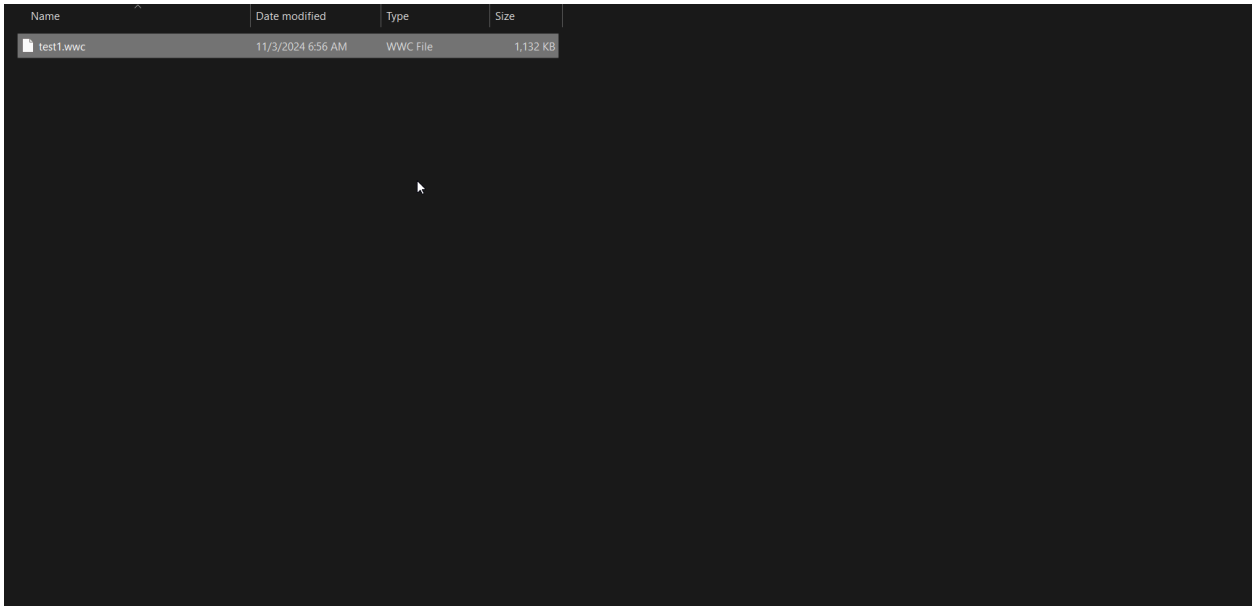
8. Save/Load File:

File -> Save File -> Select location -> Save

File will be saved in a custom file format `.wwc`, it is a simple pickle file, but specific to this application.



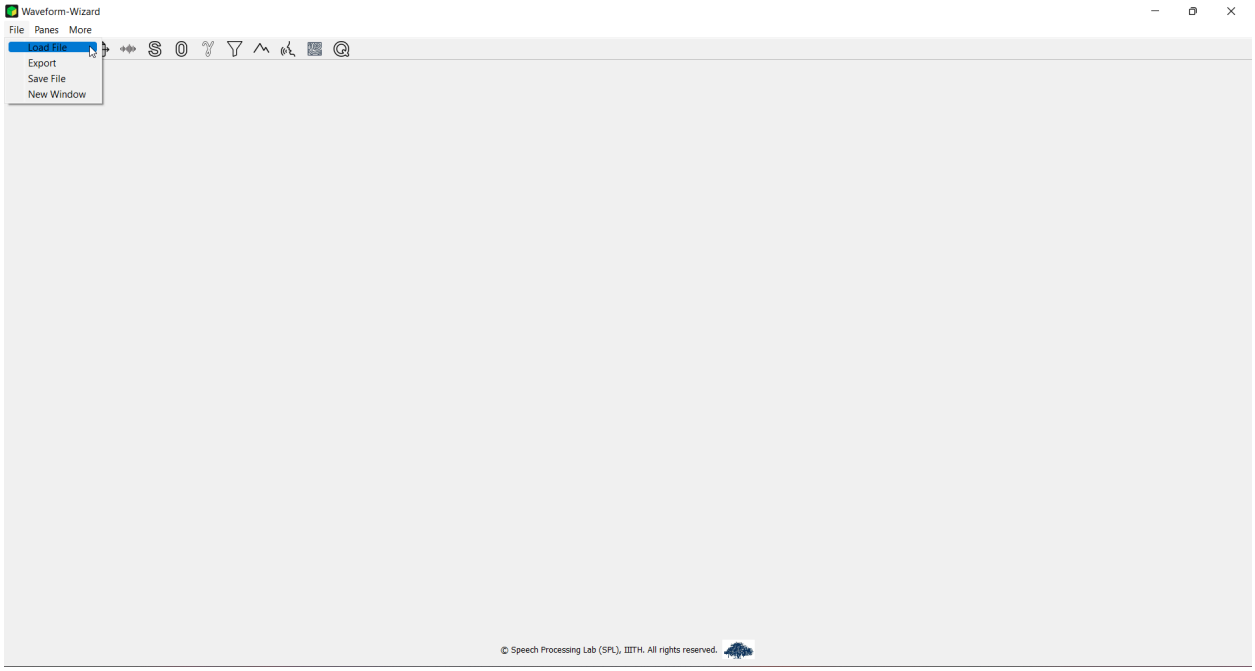


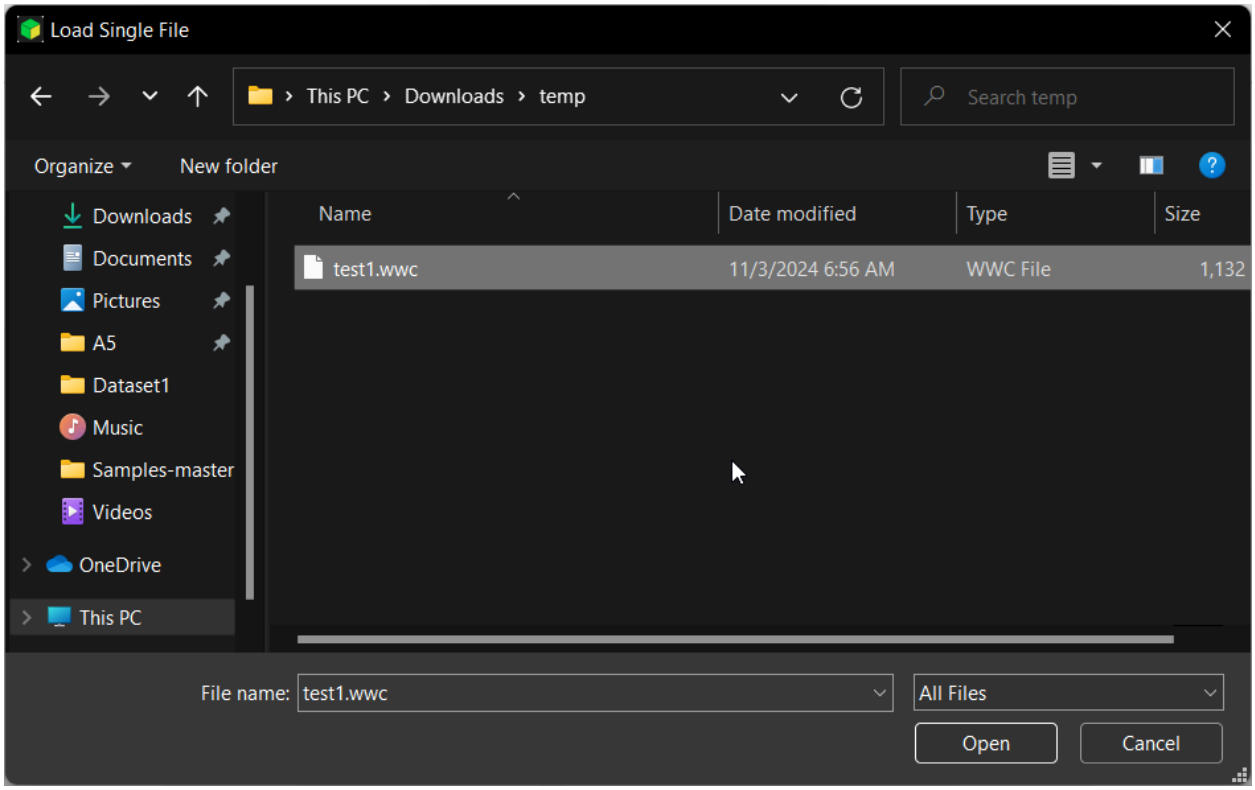


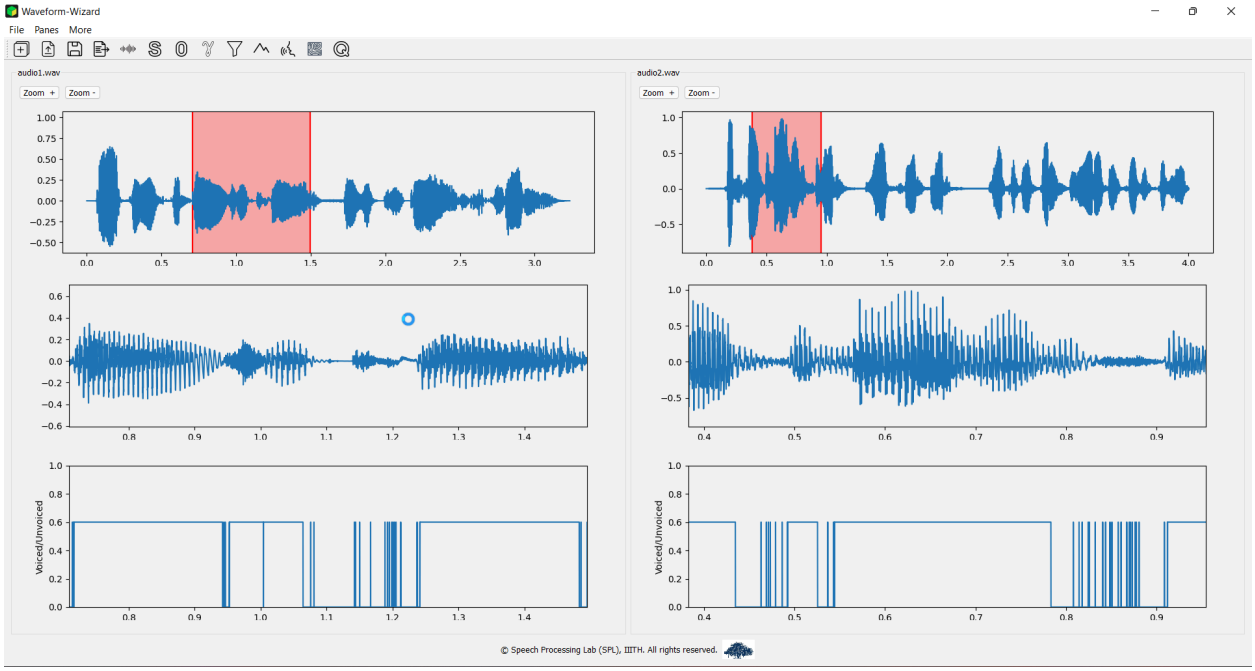
To load this saved wwc file:

File -> Load File -> Select the WWC file

The analysis is restored from that point!

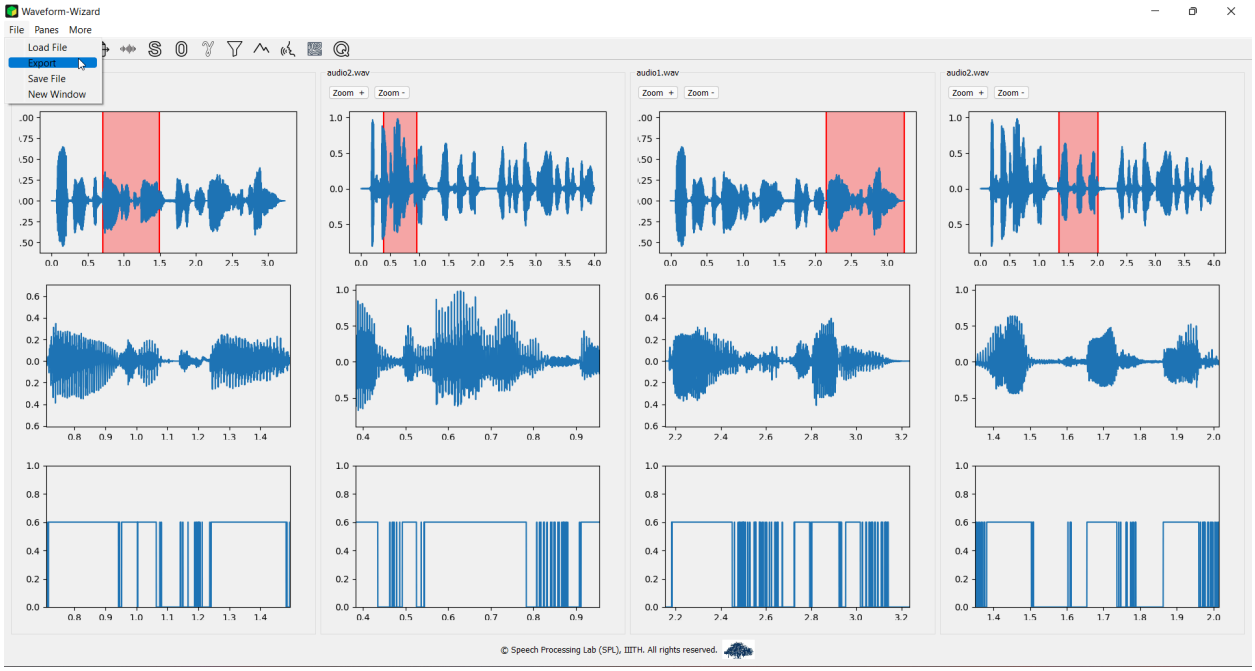

Waveform-Wizard
File
Panes
More
Load File
Export
Save File
New Window
© Speech Processing Lab (SPL), IIITH. All rights reserved.

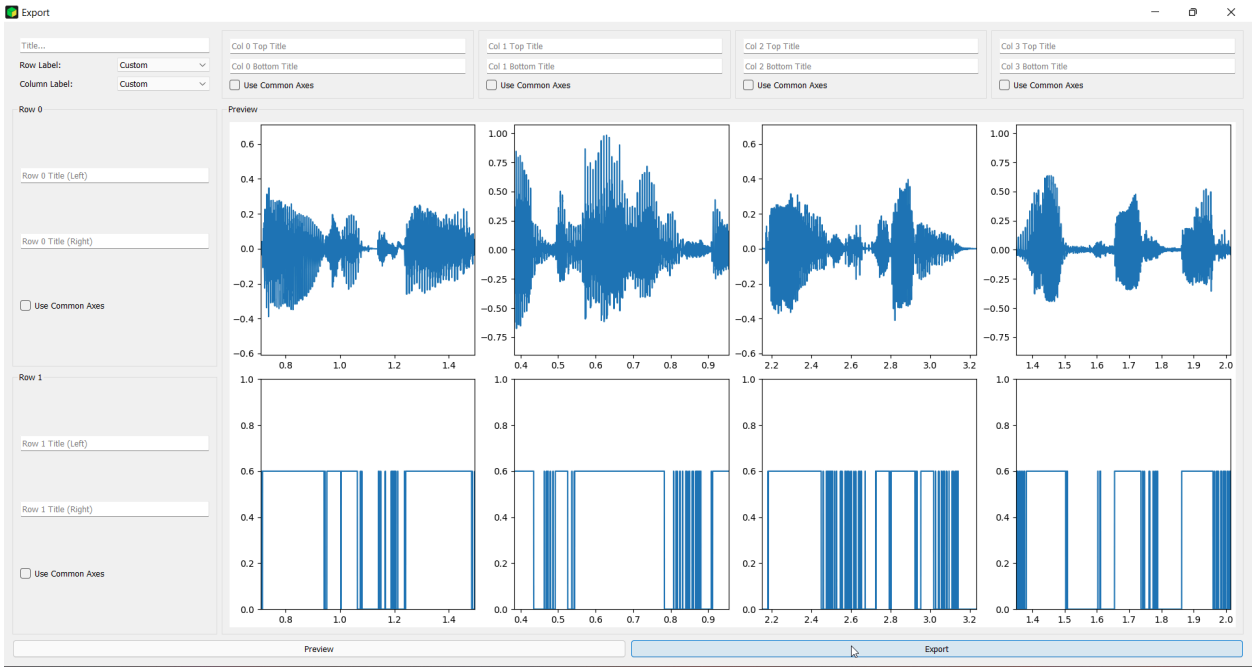

Load Single File
This PC
Downloads
temp
Search temp
Organize
New folder
Downloads
Documents
Pictures
A5
Dataset1
Music
Samples-master
Videos
OneDrive
This PC
Name
Date modified
Type
Size
test1.wwc
11/3/2024 6:56 AM
WWC File
1,132
File name:
test1.wwc
All Files
Open
Cancel

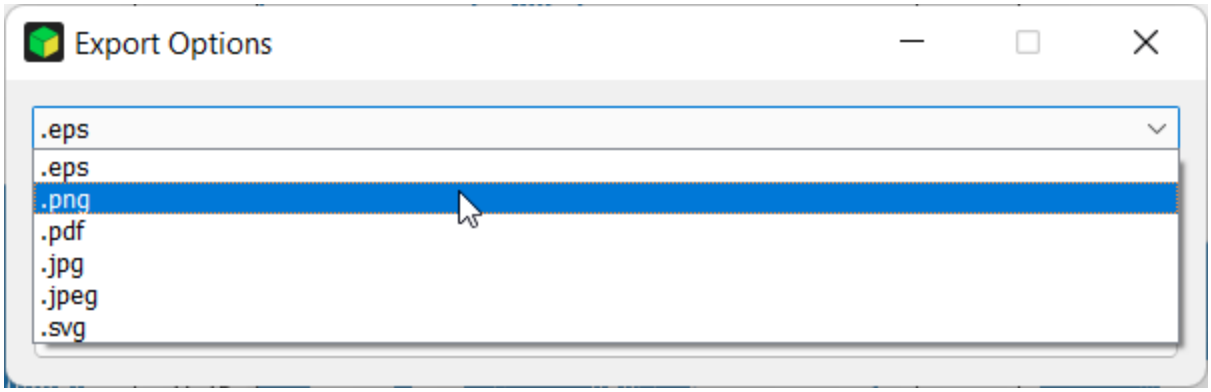

Waveform-Wizard
File
Panes
More
audio1.wav
Zoom +
Zoom -
audio2.wav
Zoom +
Zoom -
Voiced/Unvoiced
Voiced/Unvoiced
© Speech Processing Lab (SPL), IIITH. All rights reserved.

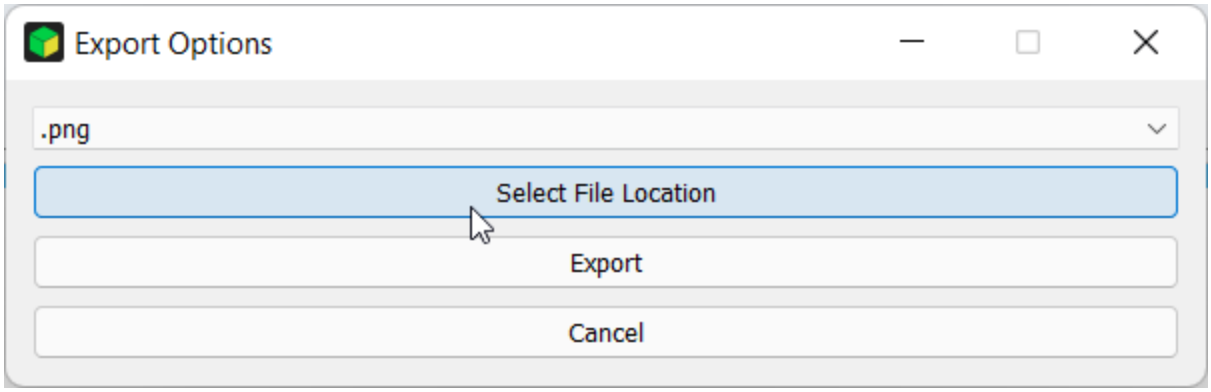
9. Export functionality:

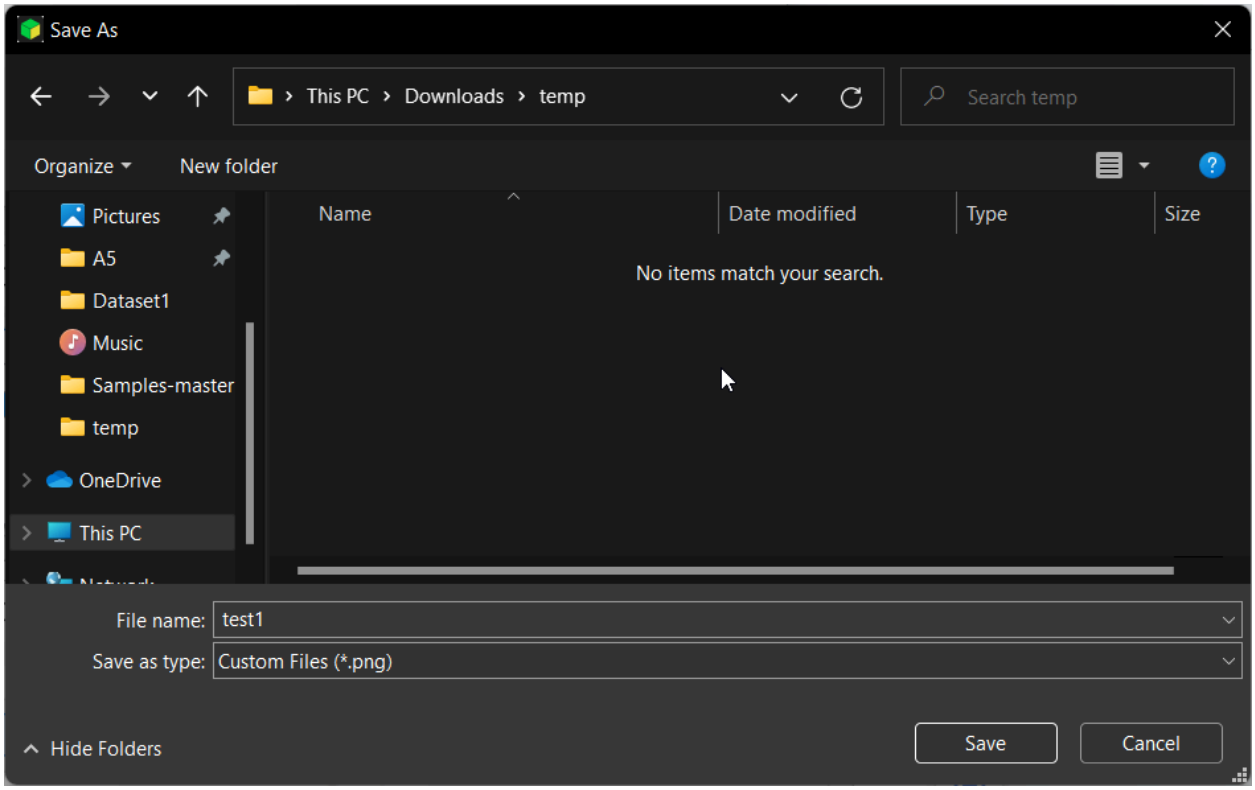
File -> Export -> Configure -> Export -> Select FIle format, Select Location -> Export.

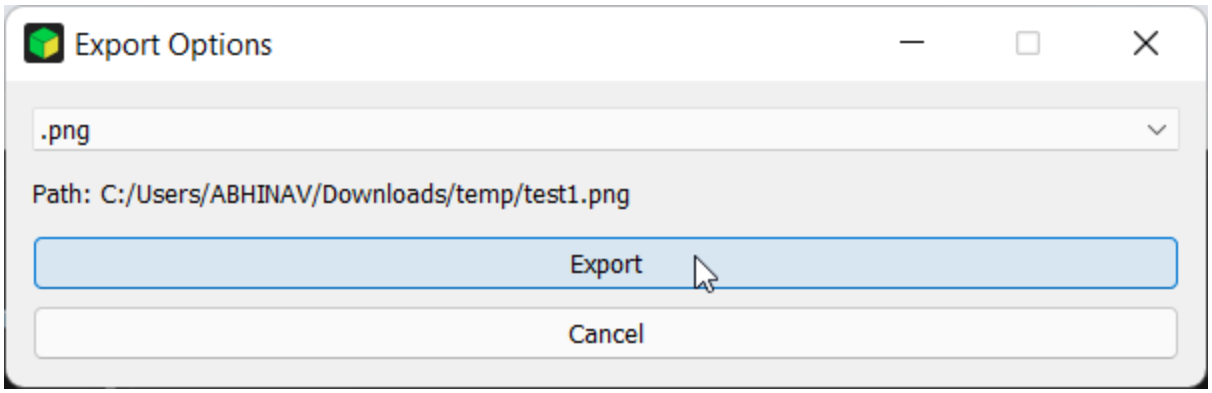


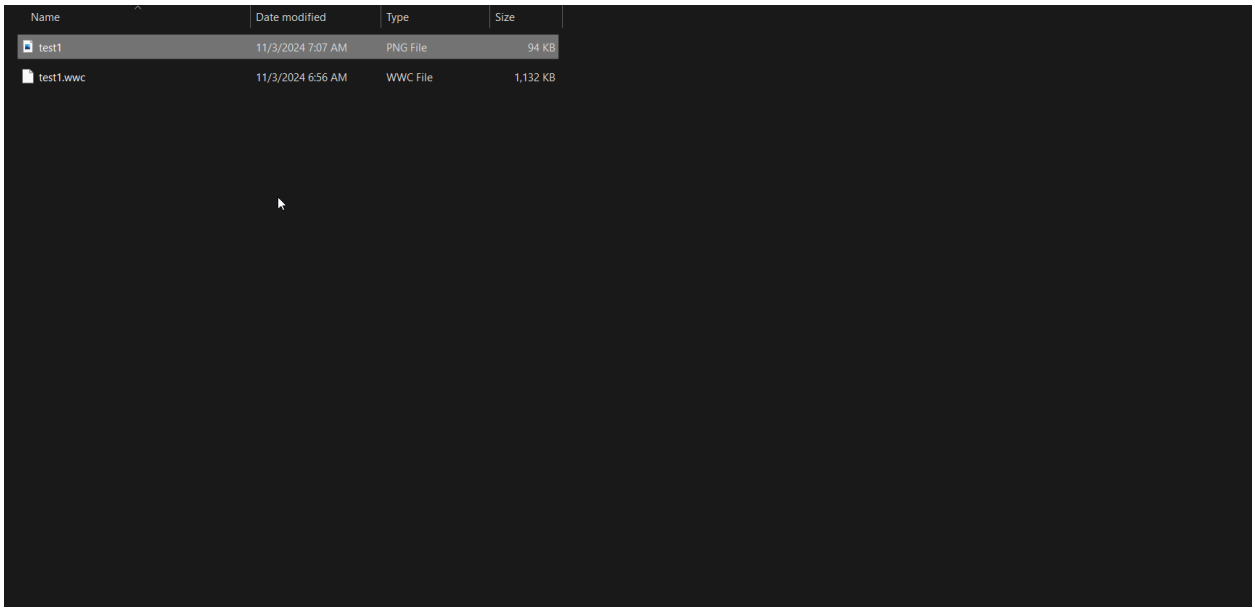


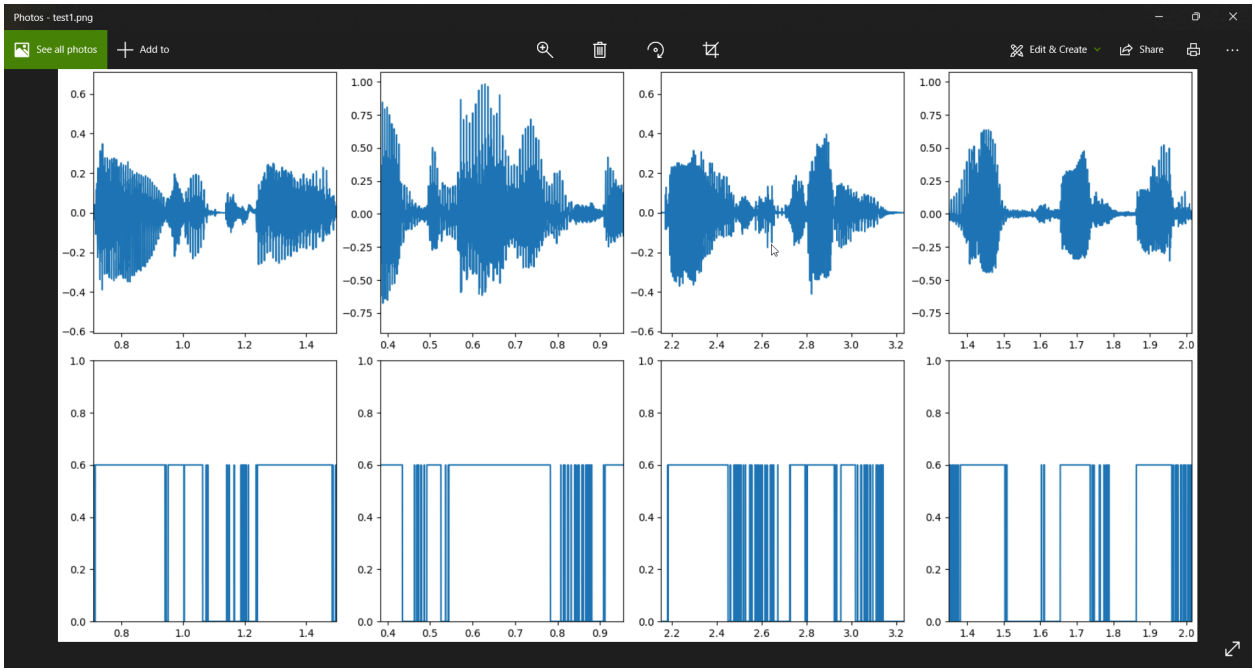

Export Options
.eps
.eps
.png
.pdf
.jpg
.jpeg
.svg


Export Options
.png
Select File Location
Export
Cancel


Save As
This PC › Downloads › temp
Search temp
Organize
New folder
Pictures
A5
Dataset1
Music
Samples-master
temp
OneDrive
This PC
Name
Date modified
Type
Size
No items match your search.
File name: test1
Save as type: Custom Files (*.png)
Hide Folders
Save
Cancel


Export Options
.png
Path: C:/Users/ABHINAV/Downloads/temp/test1.png
Export
Cancel


Name
Date modified
Type
Size
test1
11/3/2024 7:07 AM
PNG File
94 KB
test1.wwc
11/3/2024 6:56 AM
WWC File
1,132 KB


Photos - test1.png
See all photos
Add to
Edit & Create
Share

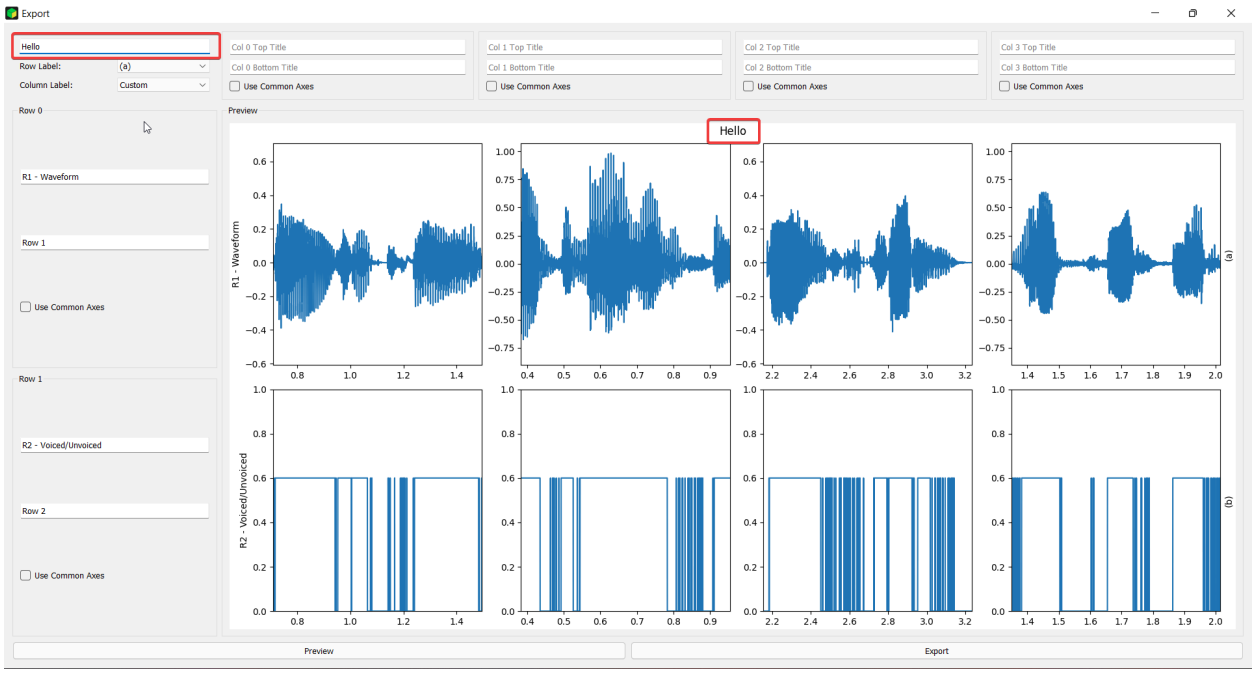
10. Custom titles and options in Export:

Supported Functionality:

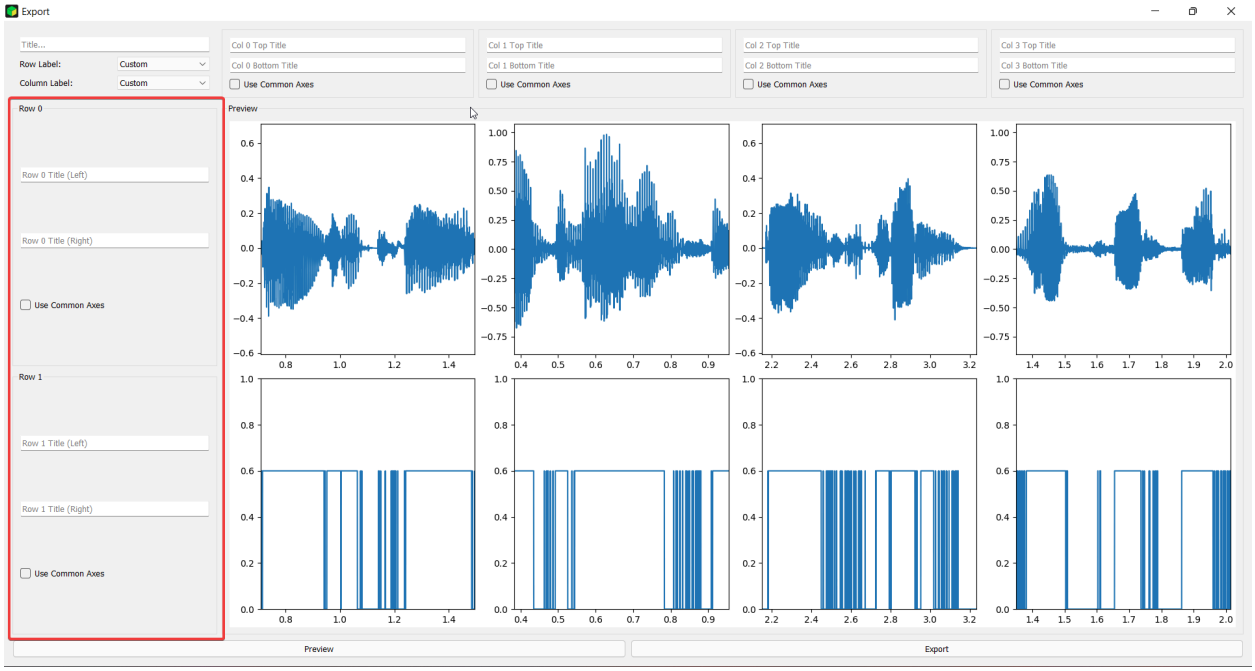
a. Main title
b. Row Titles
c. Column titles
d. Custom labels for Rows
e. Custom labels for Columns
f. Autolabels for Rows and Columns (formats ‘a’ , ‘i’ , ‘1’)

After changing configurations, ensure to click on ‘Preview’ to update the preview

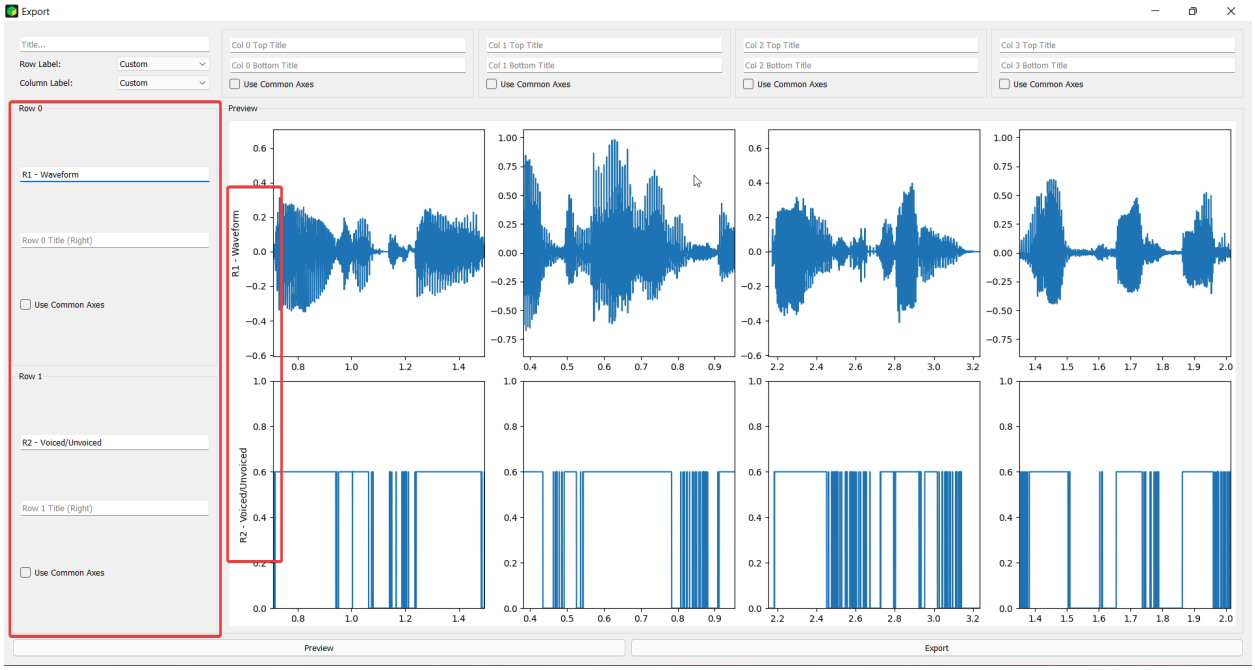
Title:



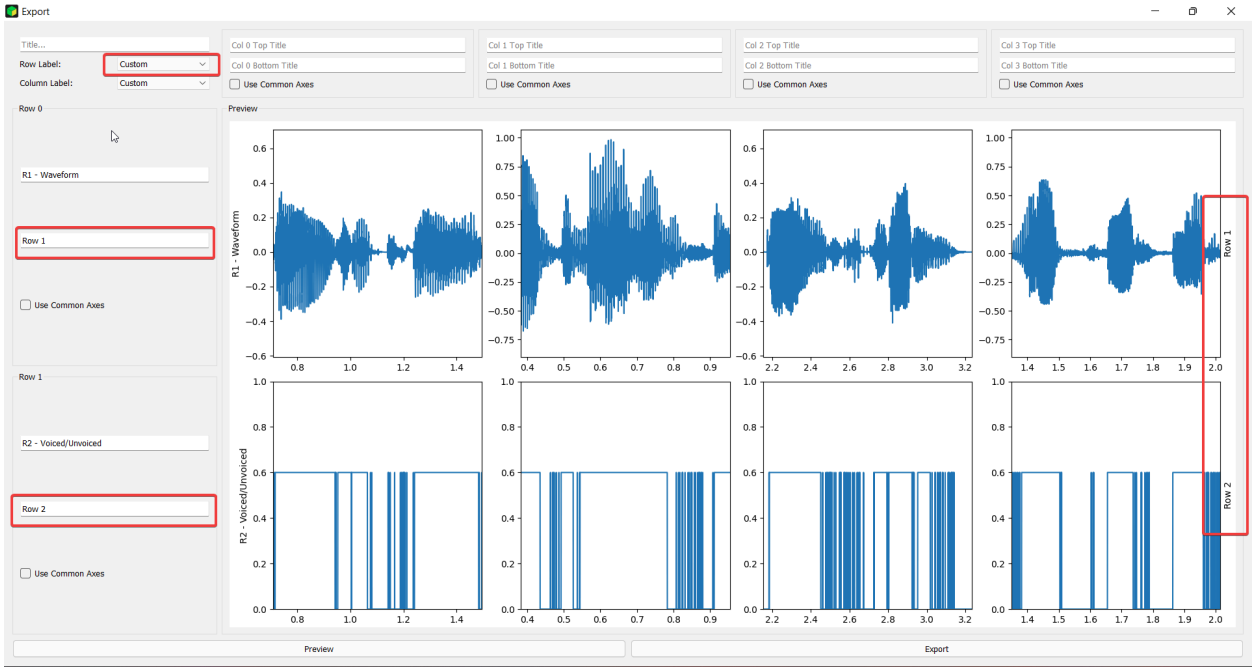
Row Configurations:



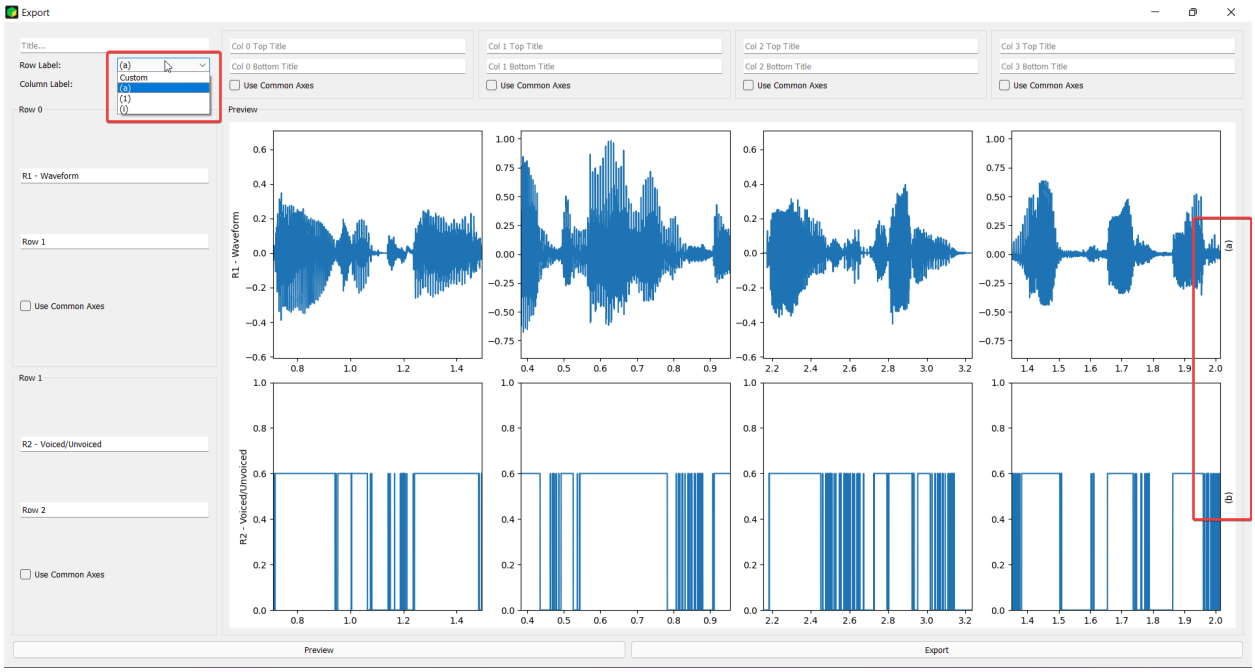
Row Custom Titles:



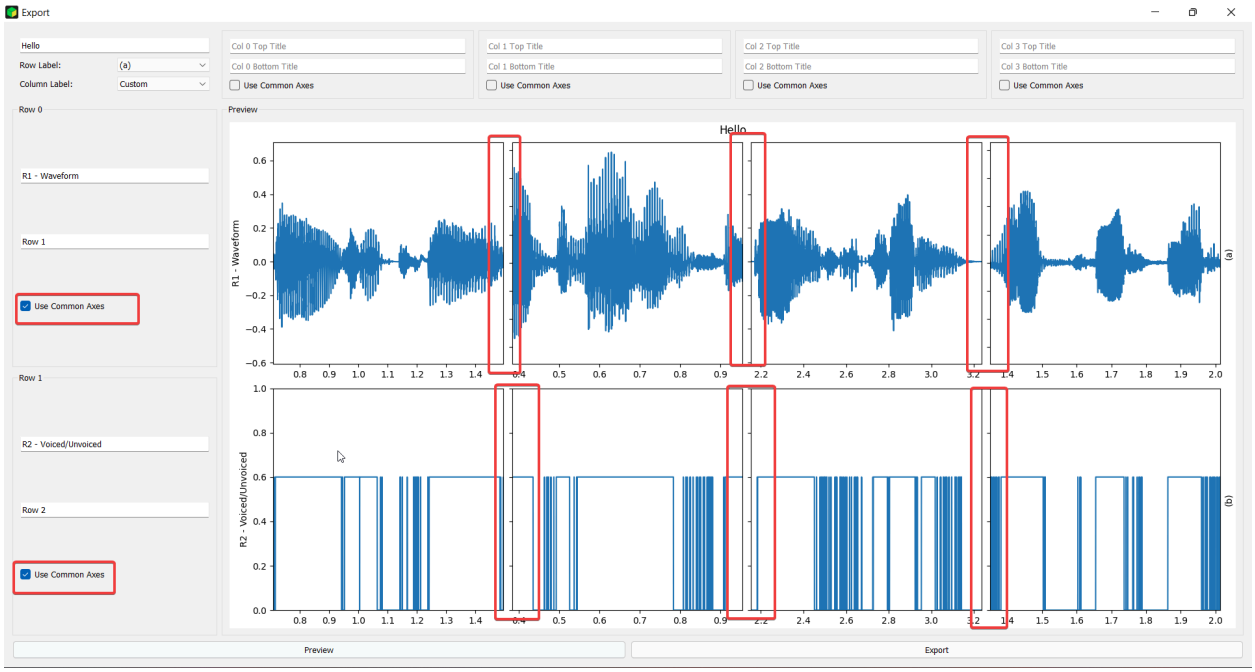
Custom Row Labelling:



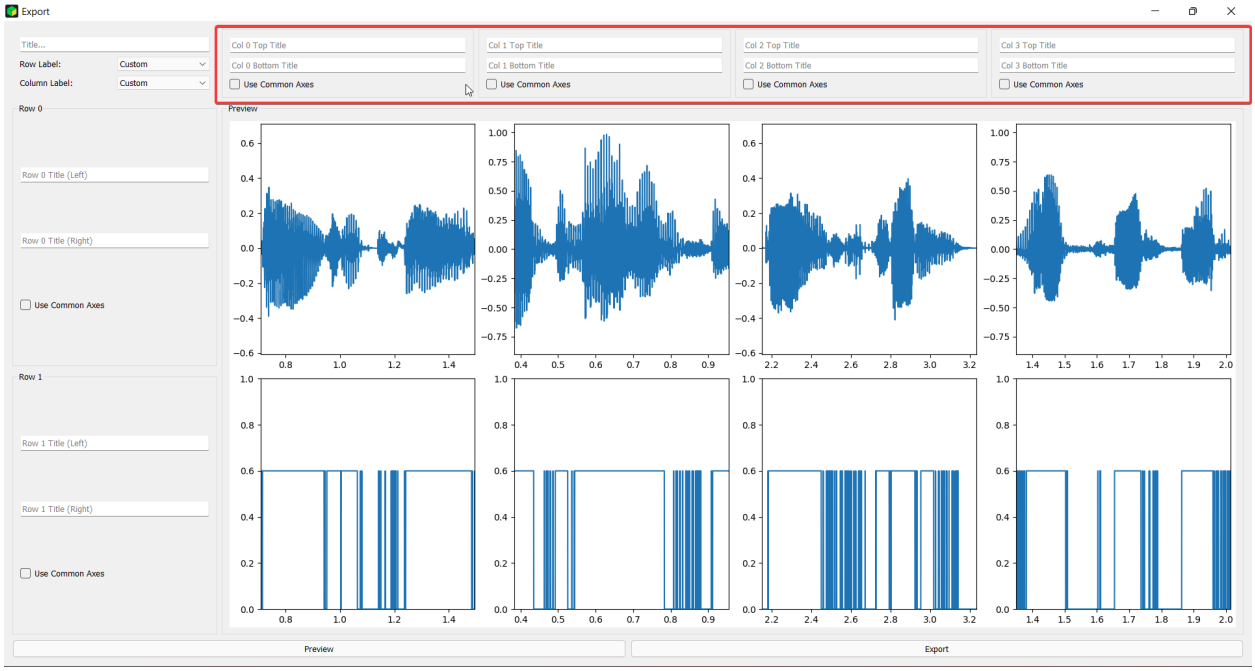
Autolabelling:

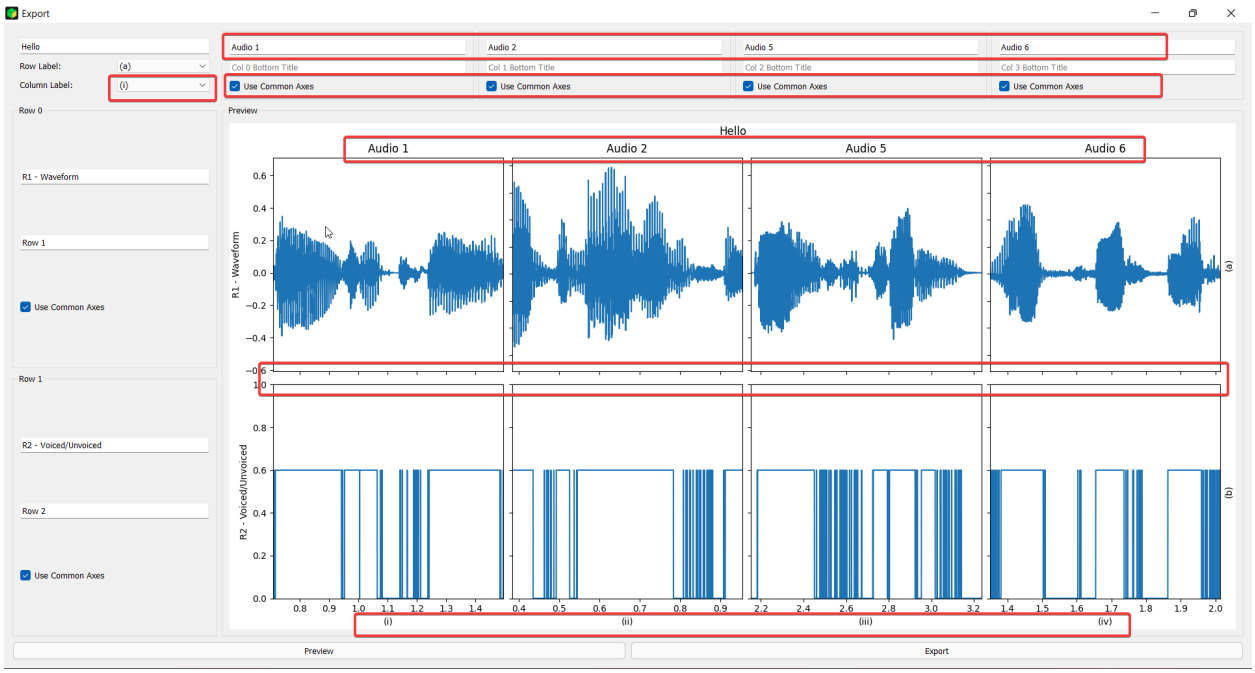


Row Common Axis:



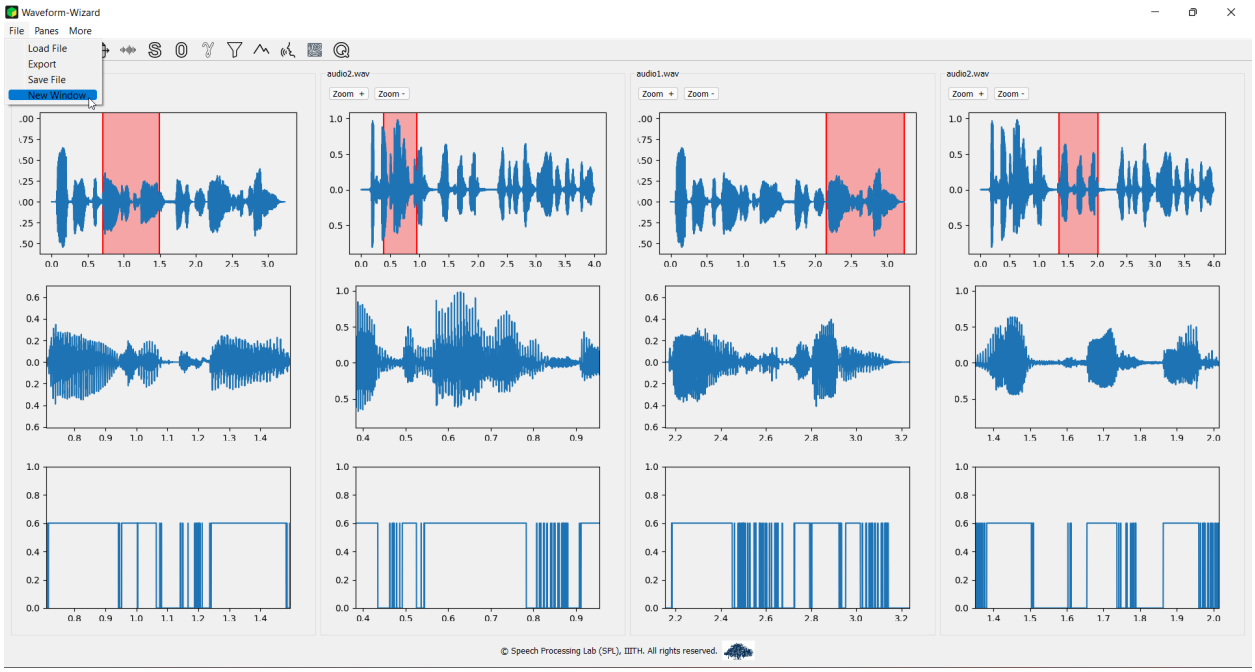
Similarly for Column:



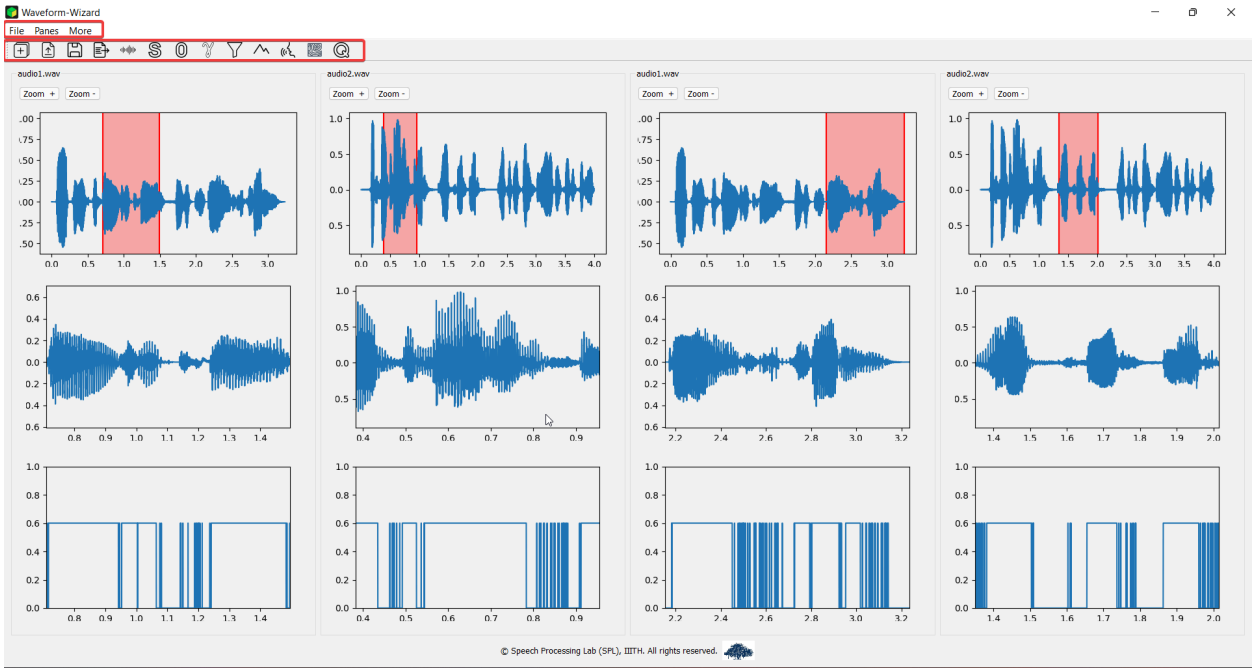


11. New Window:

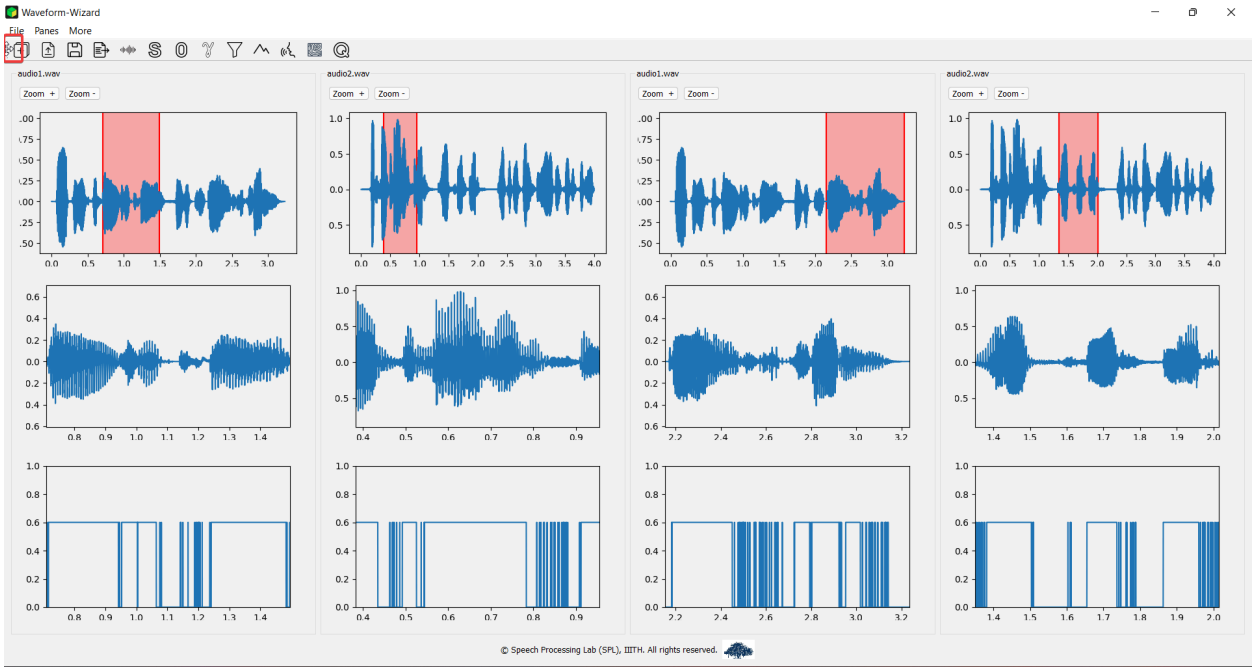
File -> New Window

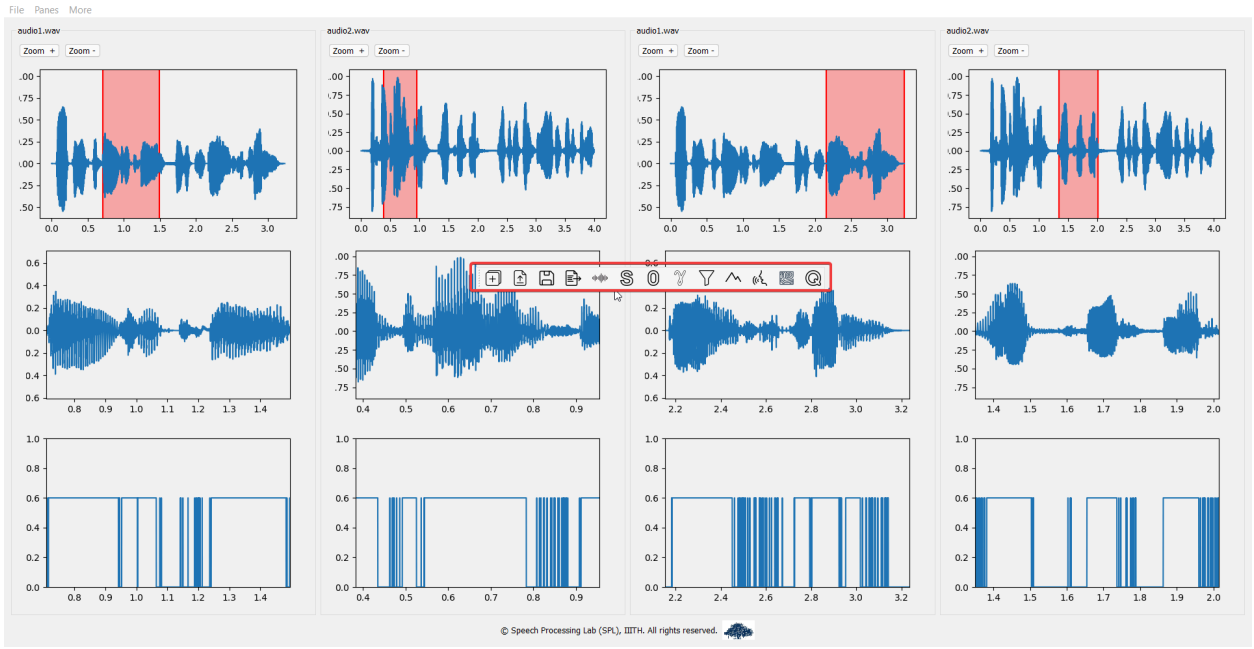


12. Menu Bar and Tool Bar:

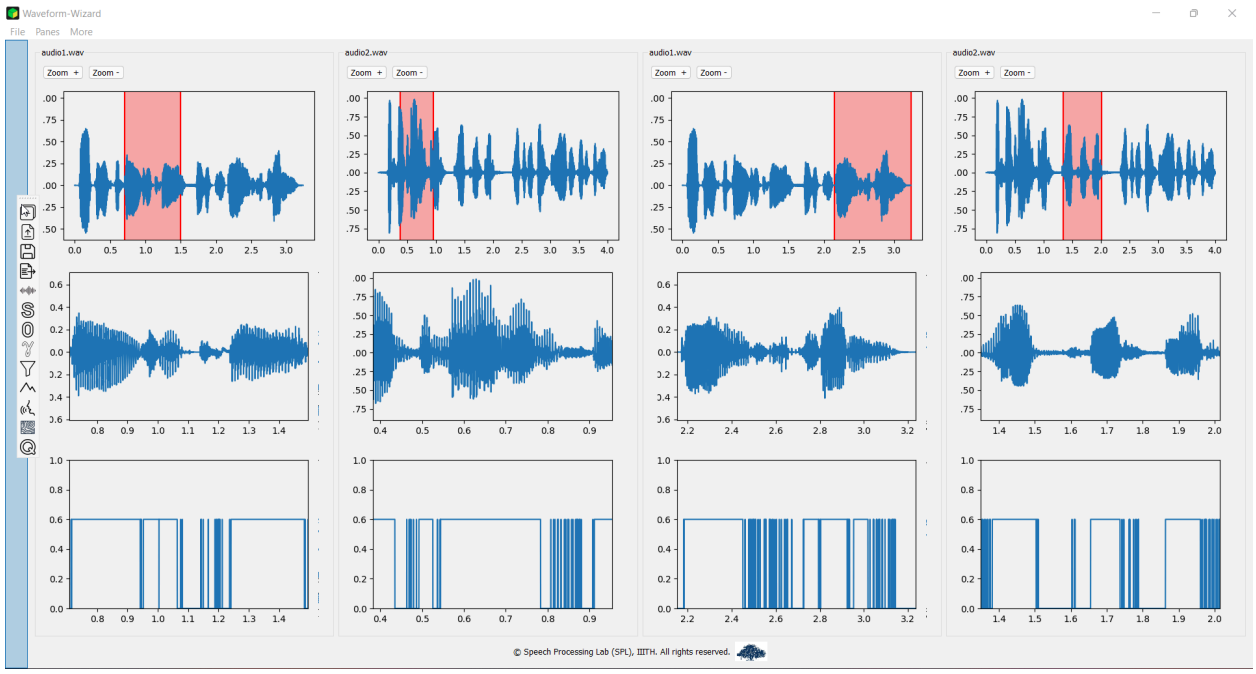


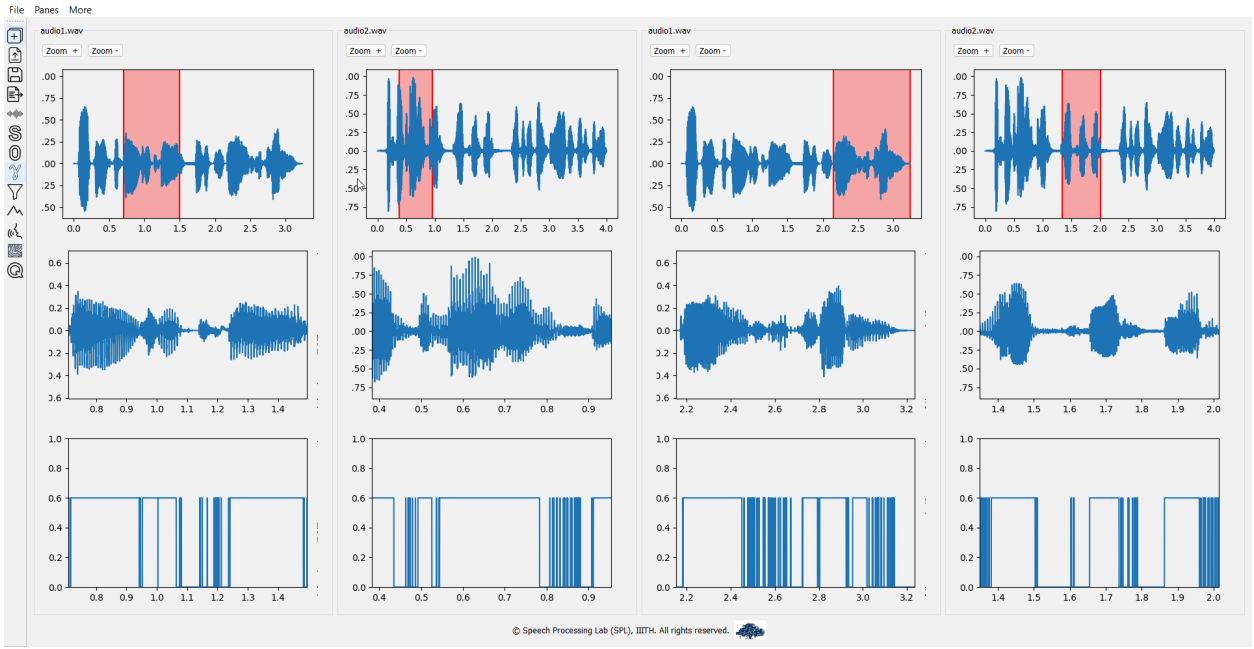
Tool Bar is Movable, just click on the 6 dots (on the left), tool bar pops out, and is movable.





Drag to edge of window to pin it back:




File
Panes
More
audio1.wav
audio2.wav
audio1.wav
audio2.wav
Zoom +
Zoom -
© Speech Processing Lab (SPL), IIITH. All rights reserved.

Acknowledgement

We express our sincere gratitude to all contributors who made the successful completion of this Application possible. First and foremost, we extend our thanks to our project supervisor, Dr. Anil Kumar Vuppala, for his guidance and support that were instrumental in completing this project. We also acknowledge the initial assistance provided by Anirudh, which helped us kickstart the work. Their invaluable input, constructive feedback, and unwavering encouragement were pivotal in shaping the success of this project.

Conclusion

In conclusion, we've achieved the development of a primary Python-based GUI Application aimed at visualizing transformations utilized in Speech. Throughout the development process, we encountered significant challenges. Despite these obstacles, we've managed to overcome them and deliver a functional application.