AUTOMATING AUDIO LABELING WORKFLOW USING DEEP LEARNING FOR VOICE ACTIVITY DETECTION

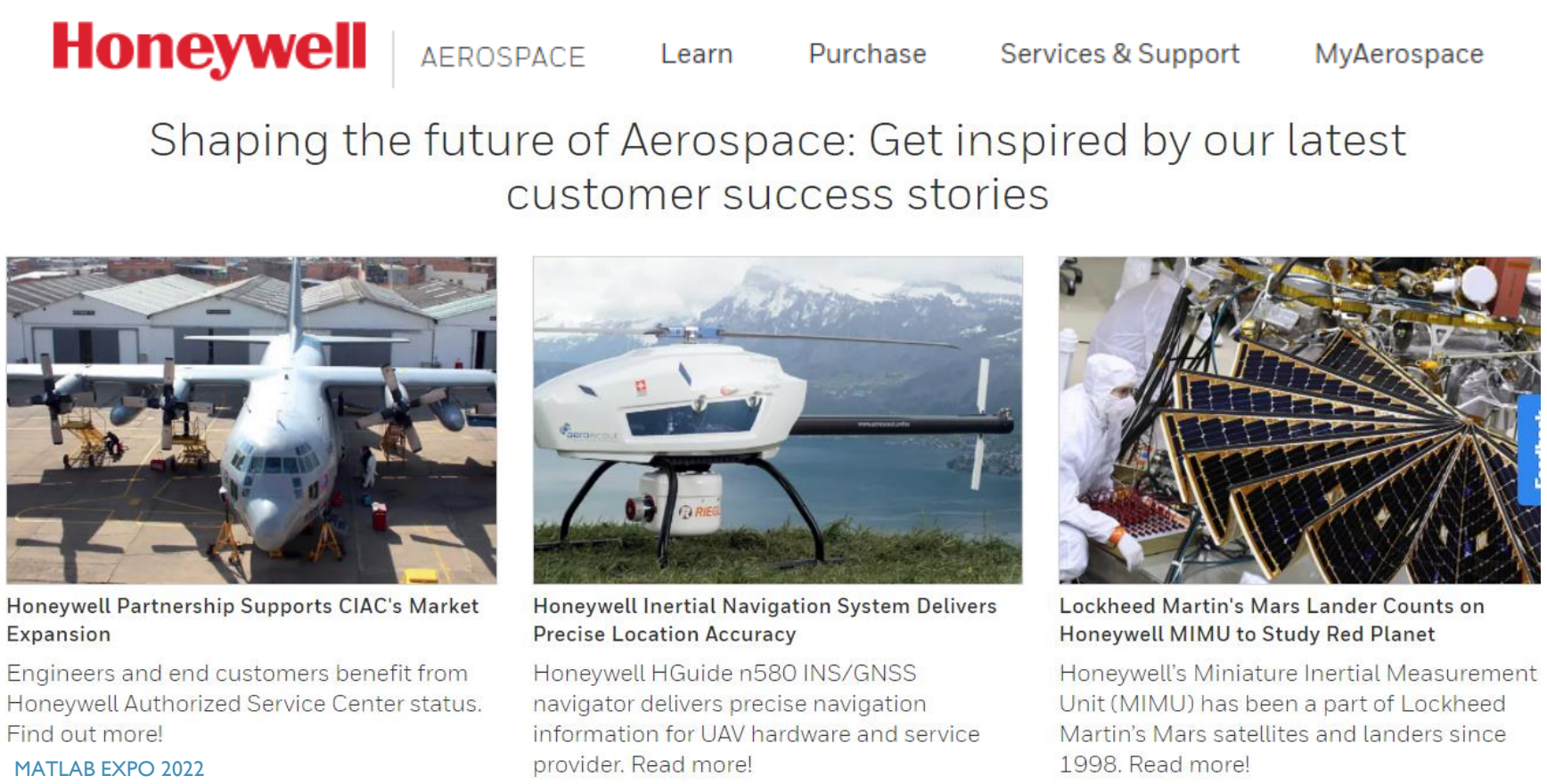
RAMAKRISHNAN RAMAN
FELLOW
HONEYWELL TECHNOLOGY SOLUTIONS

VASANTHA SELVI PAULRAJ
LEAD EMBEDDED ENGINEER
HONEYWELL TECHNOLOGY SOLUTIONS

AUTOMATING AUDIO LABELING WORKFLOW USING DEEP LEARNING FOR VOICE ACTIVITY DETECTION

- This presentation discuss our experience in using a pre-trained Deep Learning model in a labeling algorithm and demonstrates how we have adopted automation of Audio Labeling workflow towards audio Voice Activity Detection(VAD) using MATLAB.

HONEYWELL AEROSPACE - HTTPS://AEROSPACE.HONEYWELL.COM/



Honeywell Aerospace products and services are found on virtually every commercial, defense and space aircraft in the world. With an unmatched heritage of innovation that spans more than a century, our aim is to solve the greatest challenges CEOs, pilots, operators, passengers, finance, maintenance and cabin crews face -- and transform the way we all fly. We deliver improved fuel-efficiency, more direct and on-time flights, safer and more comfortable travel and better flight planning and traffic management. We do this through one of the industry's broadest and most advanced portfolios including world-class engines, cockpits, cabin design, wireless connectivity and enterprise performance management services – even for emerging market segments like flying taxies, while providing world-class customer service, repairs and technical support.

HONEYWELL: SPEECH TECHNOLOGIES

Speech Mag

Speech TECHNOLOGY

Speech Recognition | Customer Self Service | Virtual Assistants | Analytics | Artificial Intelligence | More Topics

HOME | SUBSCRIBE | NEWS | IN DEPTH | WHITE PAPERS | WEBINARS | RESOURCES | CONFERENCES | ABOUT

Honeywell Tests Voice Recognition in Plane Cockpits

Honeywell Aerospace has begun testing voice recognition inside aircraft cockpits.

The company is currently using its Honeywell Innovative Prototyping Environment (HIPE) to flight-test voice recognition in an Embraer ERJ170 aircraft and is working with pilots and customers to assess its usability, safety, and efficiency in real airborne scenarios.

Method and system for remotely training and commanding the speech recognition system on a cockpit via a carry-on- device in a connected aircraft
Patent number: 10672385
Date of Patent: June 2, 2020
Assignee: Honeywell International Inc.
Inventors: Bharath Sundararajan, Vasantha Selvi Paulraj, Rengarajan Mohan

Audio processing system and method using push to talk (PTT) audio attributes
Patent number: 10278034
Date of Patent: April 30, 2019
Assignee: HONEYWELL INTERNATIONAL INC.
Inventors: Sivaraman Saptharishi, Vasantha Selvi Paulraj

Methods and apparatus for post-processing speech recognition results of received radio voice messages onboard an aircraft
Patent number: 10157616
Date of Patent: December 18, 2018
Assignee: HONEYWELL INTERNATIONAL INC.
Inventors: Jitender Kumar Agarwal, Vasantha Selvi Paulraj, Kiran Gopala Krishna, Chaya Garg, Roger W. Burgin, Robert E De Mers

Sample Patents

PROBLEM CONTEXT

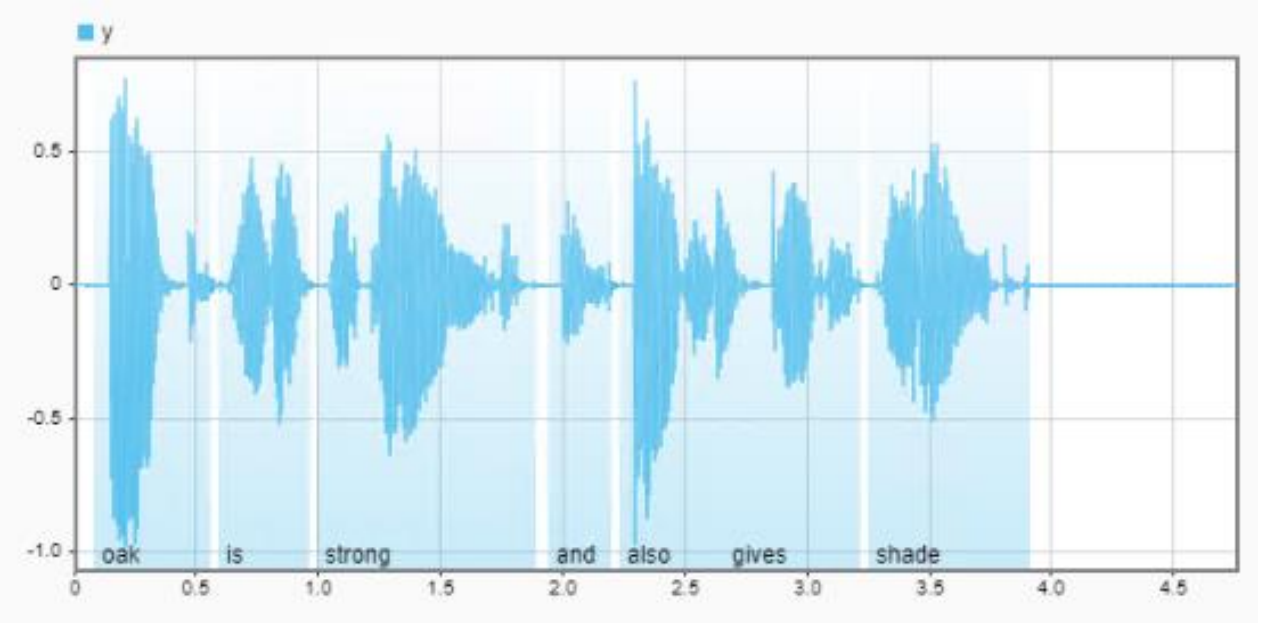
- Speech technologies for Aviation domain has a unique challenge as it involves heavy domain specific vocabulary.
- There is need for large amount of domain specific labeled corpus for each accent and language
- Larger vocabulary use cases require huge training data as well to be labeled

CHALLENGES IN LABELING OF AUDIO DATA

- For audio files, labeling involves analyzing segments of audio and listening to them and manually assigning appropriate labels for specific time slots in the audio files.
- However, such an Audio Labeling workflow is a labor-intensive process and slows the development cycle time
- For instance, if we are planning to train a Deep Learning Model to classify segments into speech or noise (for VAD - Voice Activity Detection purposes), we need to label the speech and noise segments in the audio files.
- Manual labeling of audio data is a resource intensive and time-consuming process

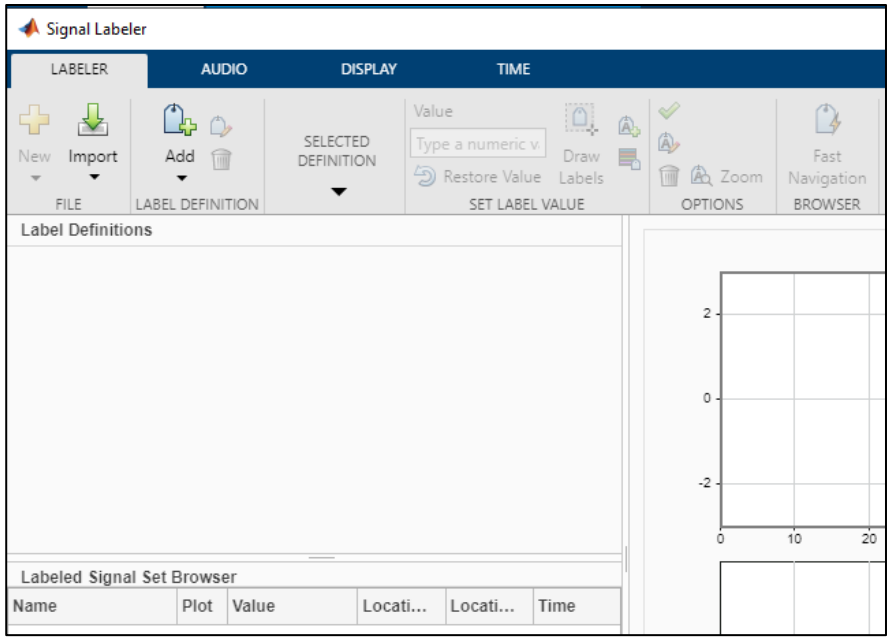
LABELING OF DATA

- Deep Learning models require labelled data for training purposes. Labeling of the data is a crucial and important step.
- To overcome the challenges, feasibility of automated labeling of the data was explored, using MATLAB Signal Labeler



LABELING WORKFLOW

Import audio file to signal Labeler → Define Label → Custom Labeling Function → Auto Labeling → Examining the labels → Exporting Labeled signals

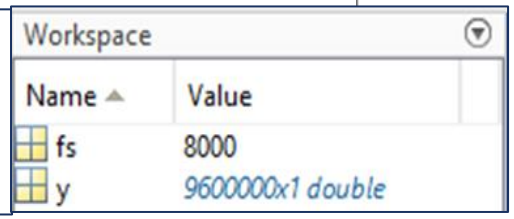


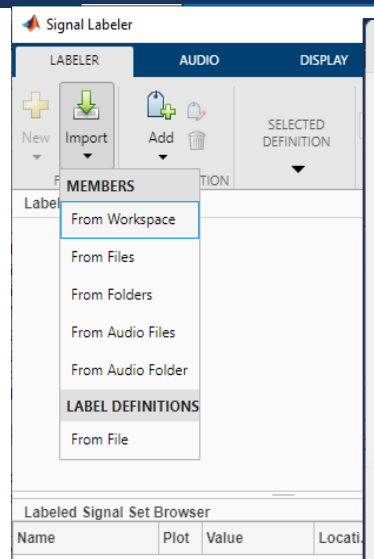
STEP 1: LOADING AUDIO FILE & IMPORT TO SIGNAL LABELER

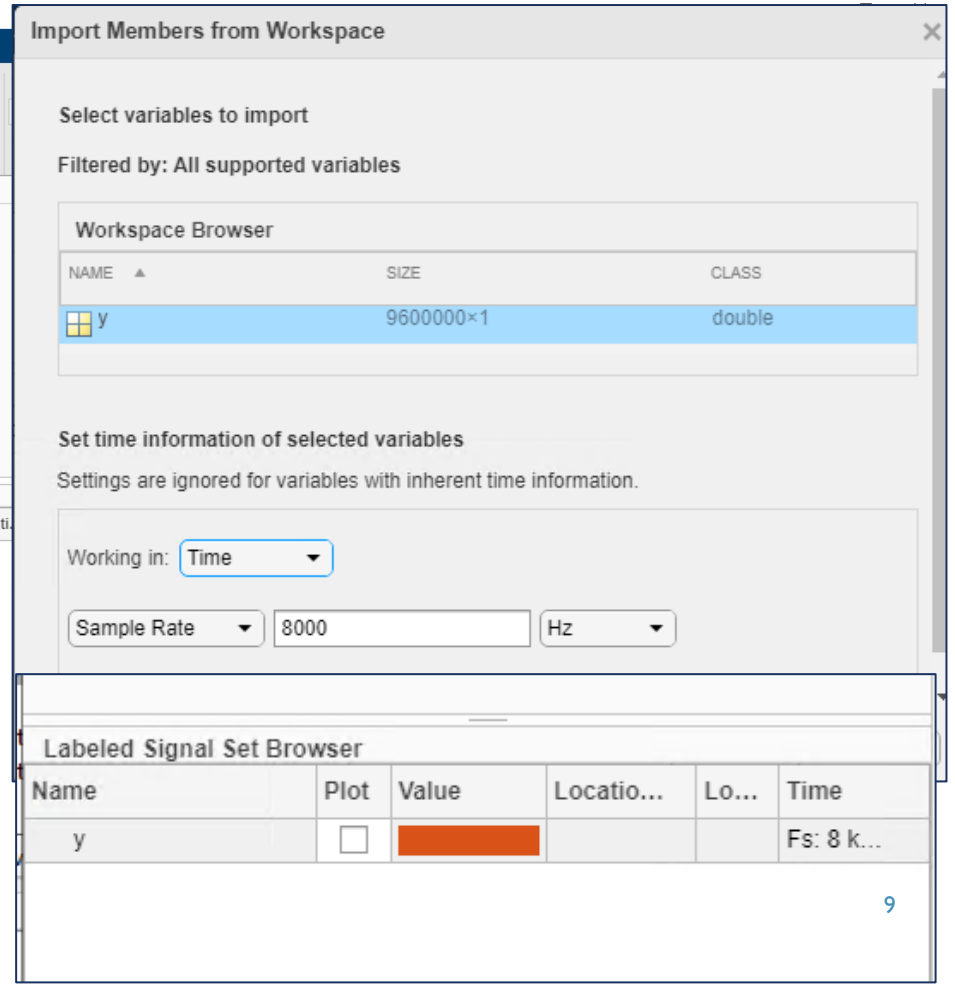
- Open Signal Labeler. On the Labeler tab, click Import and select file (use 'select from workspace' in the Members list, if loaded prior).
- In the dialog box, select the signal, y.
- Add time information: Select Time from the drop-down list and specify fs as the sample rate, which is measured in Hz.
- Close the dialog box ("Import & close"). The signal appears in the Labeled Signal Set Browser.

To load audio file in workspace:

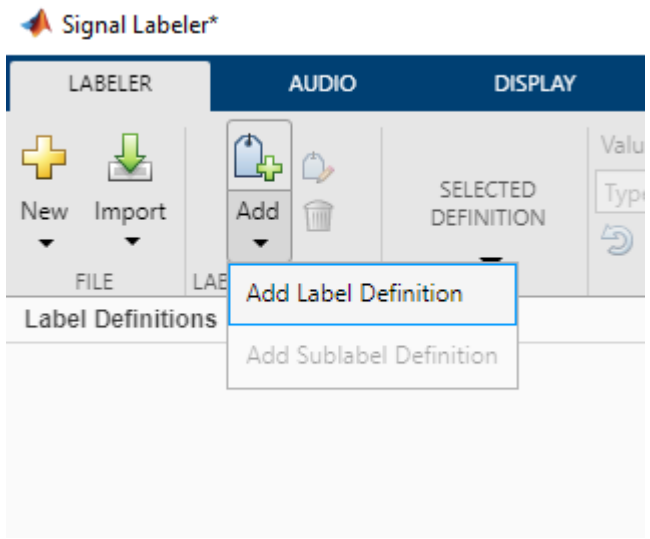
- [y,fs] = audioread('myAudioFile.wav');
- % To hear, type soundsc(y,fs)



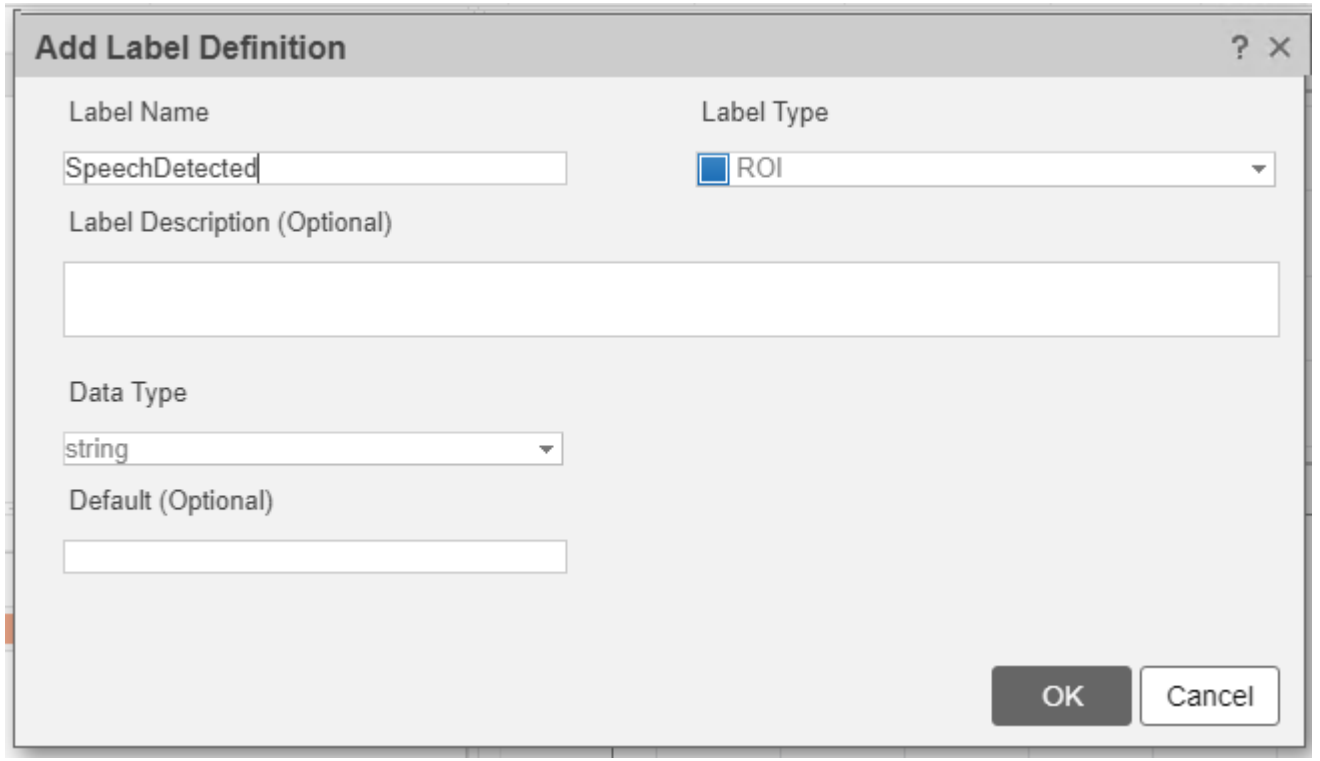




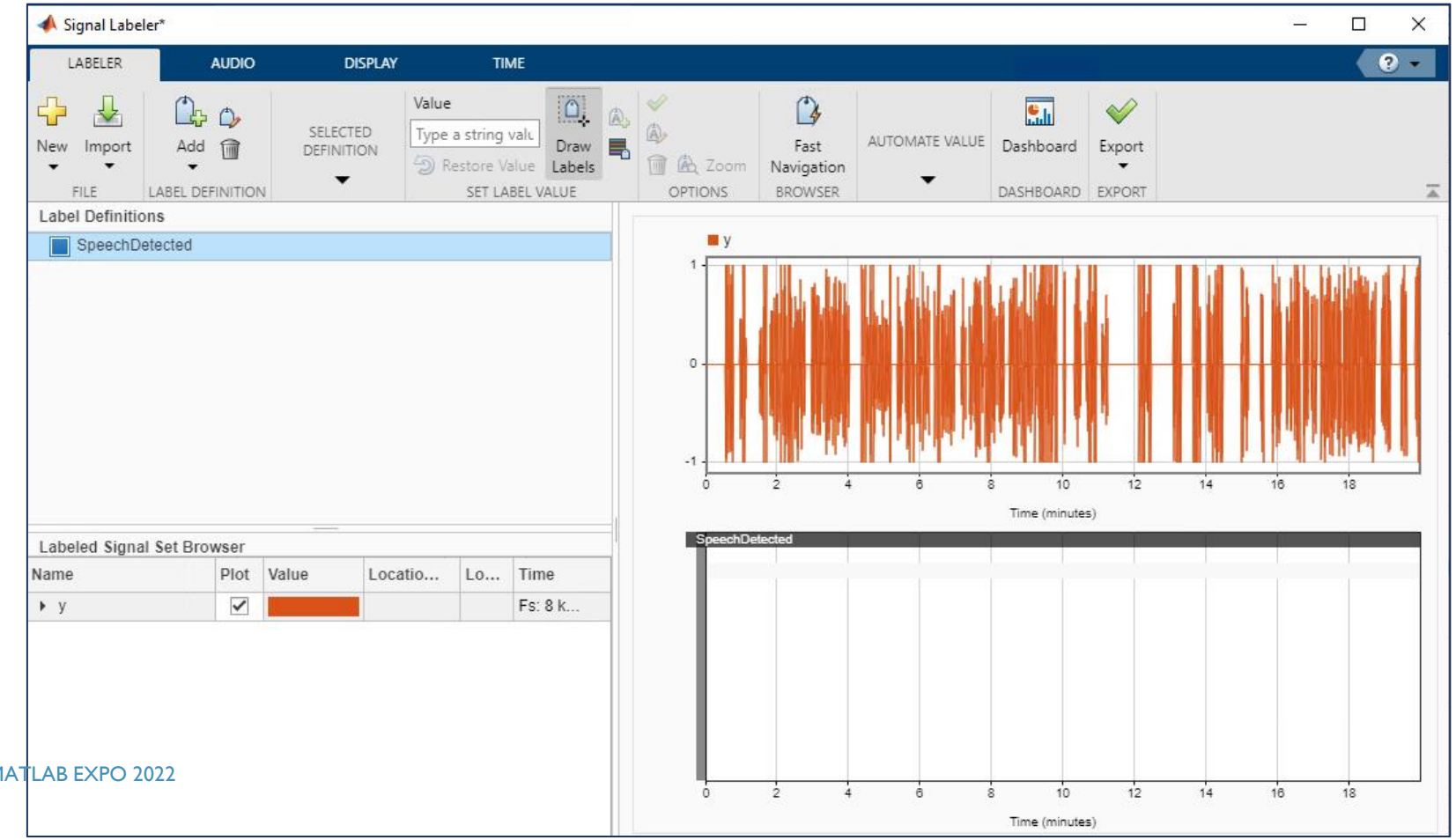
STEP 2: DEFINE LABEL

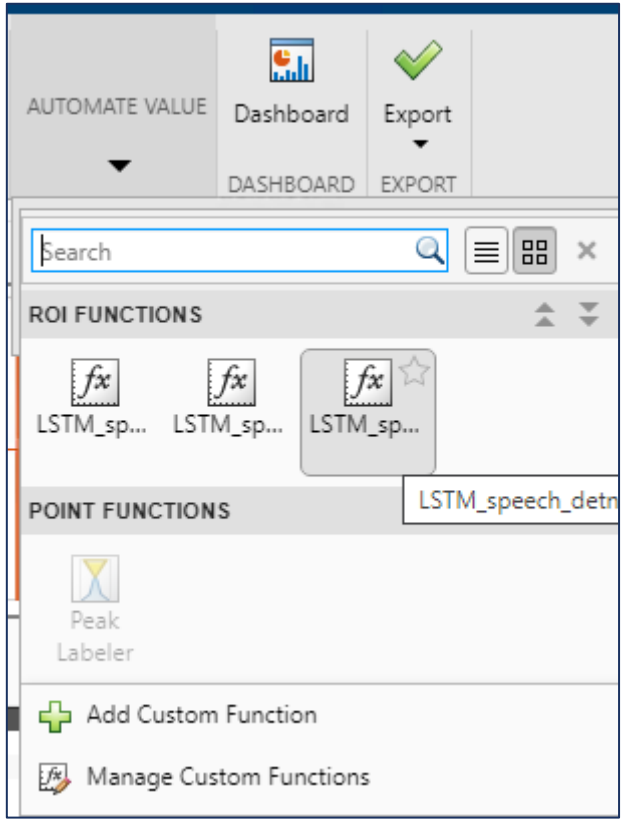


- Click Add Definition on the Labeler tab.
- Specify the Label Name as "SpeechDetected", select a Label Type of ROI, and enter the Data Type as string.

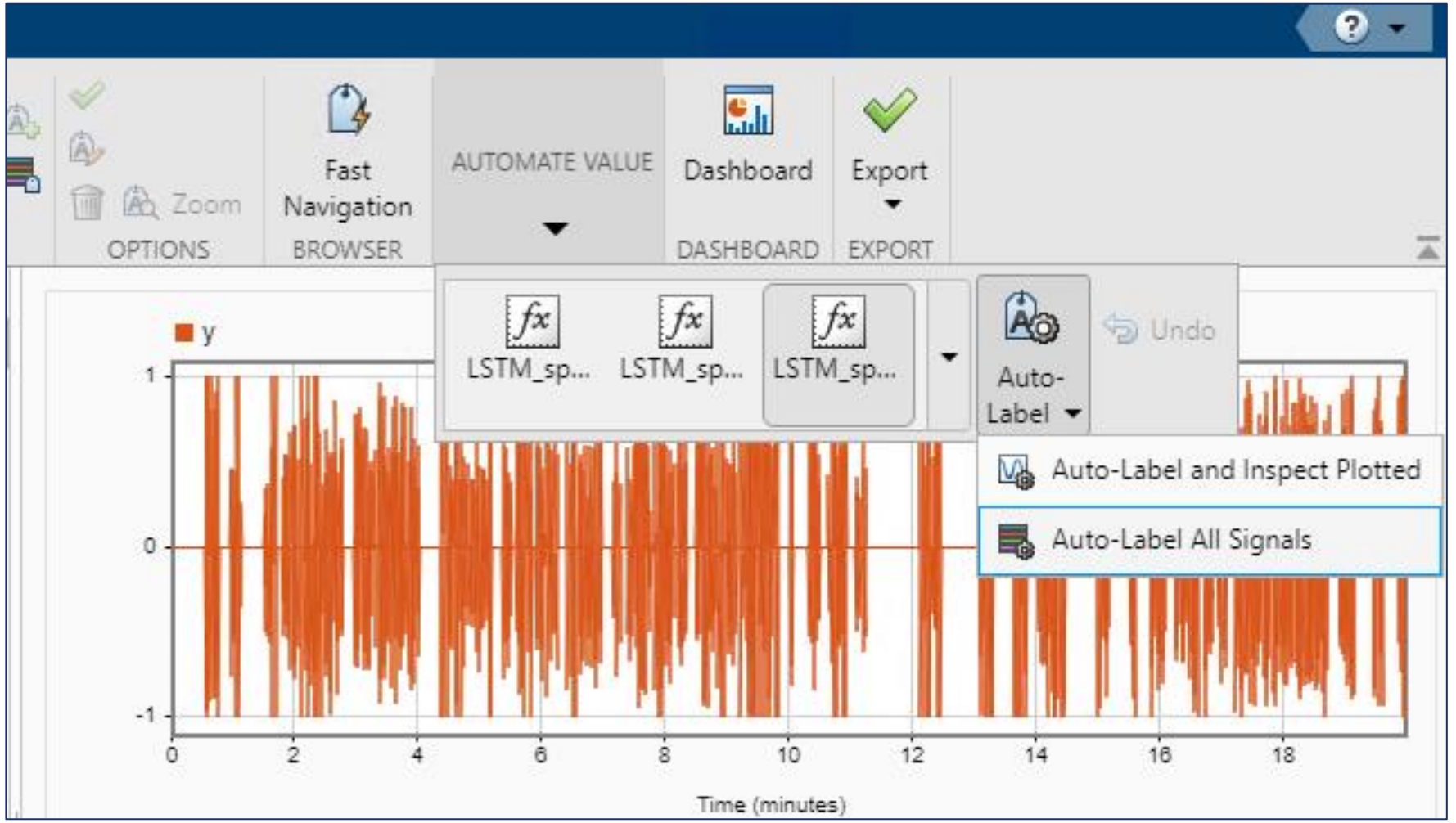


STEP 3: CUSTOM LABELING FUNCTION

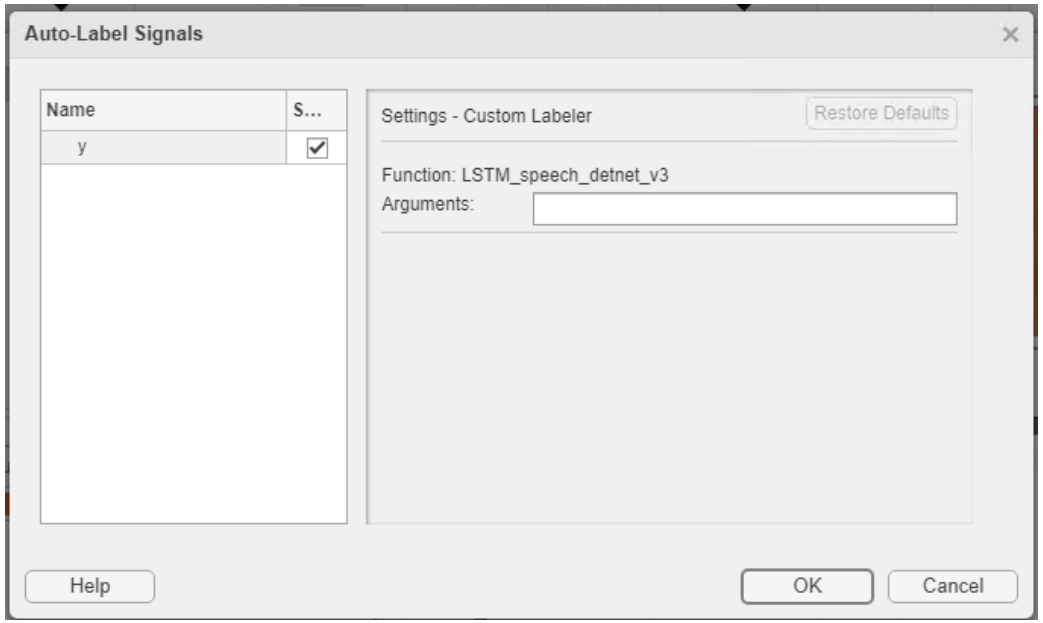


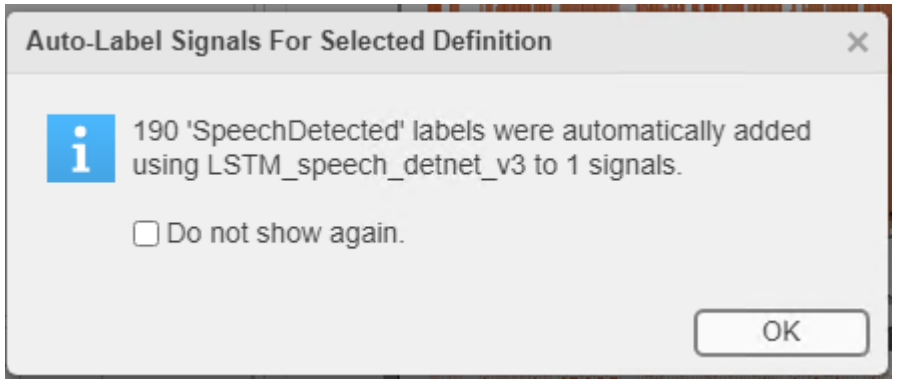


STEP 4A: AUTO LABELING

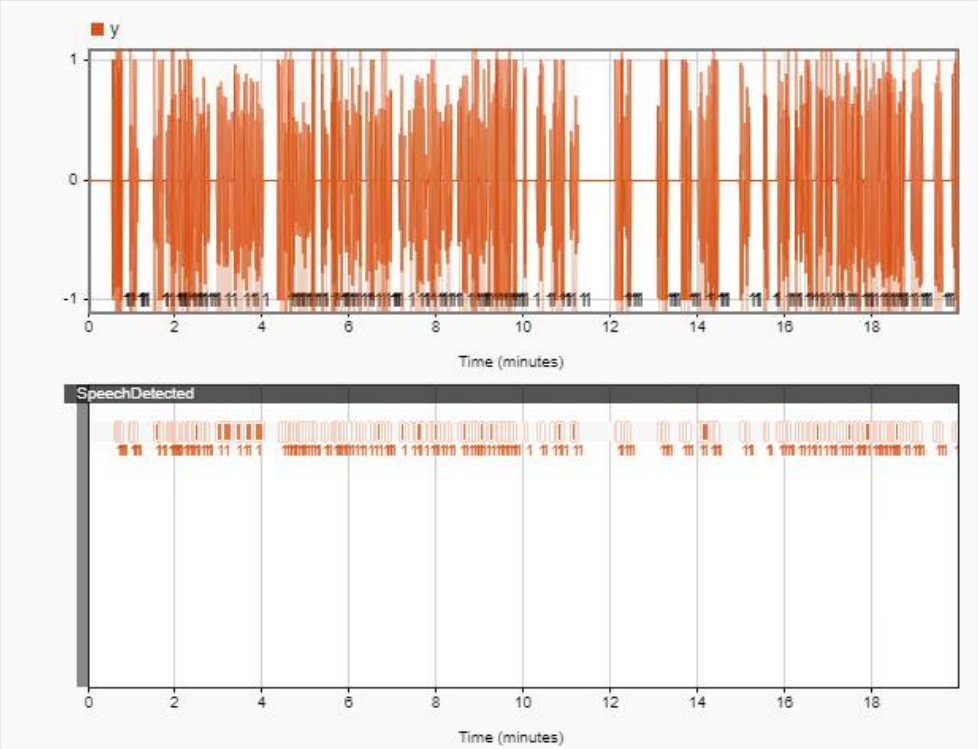


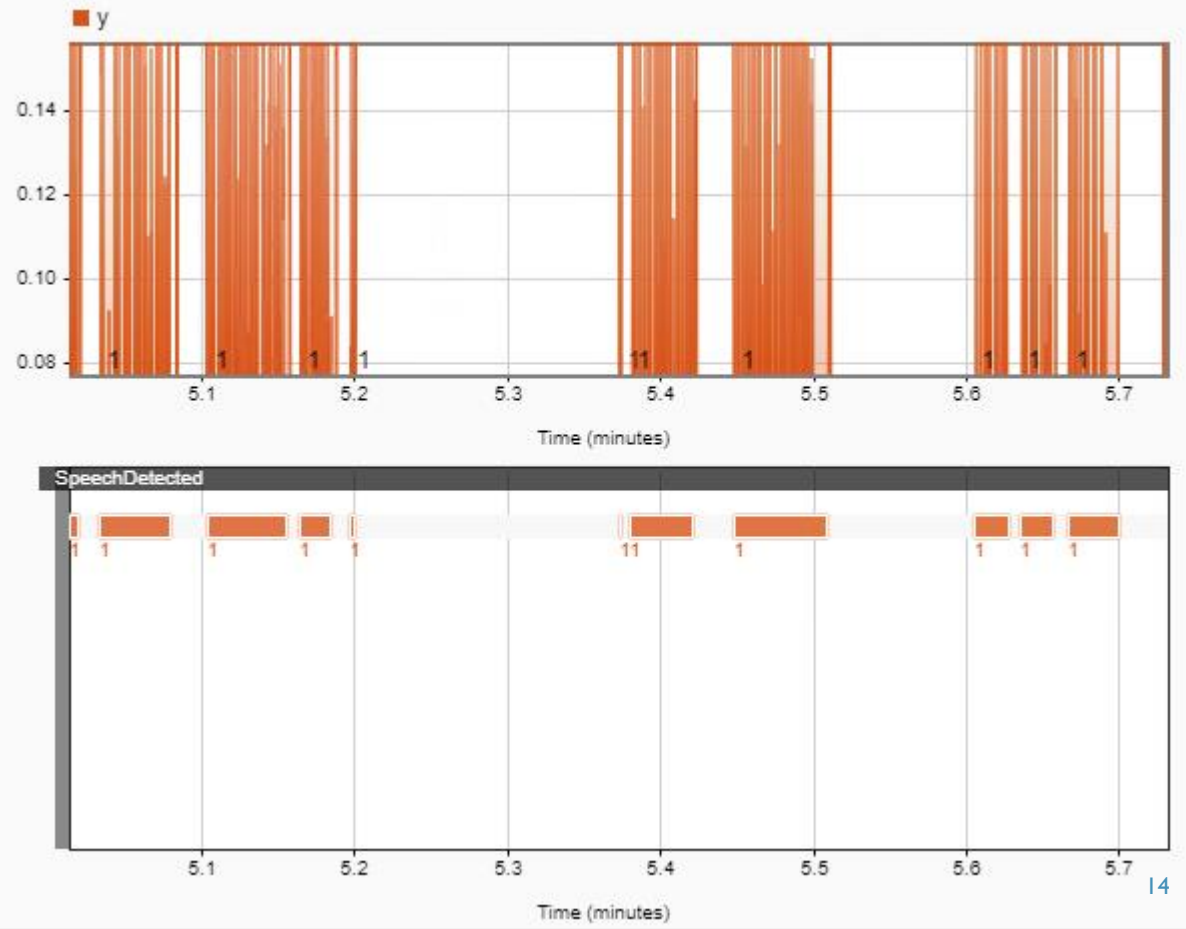
STEP 4B: LABELING COMPLETE



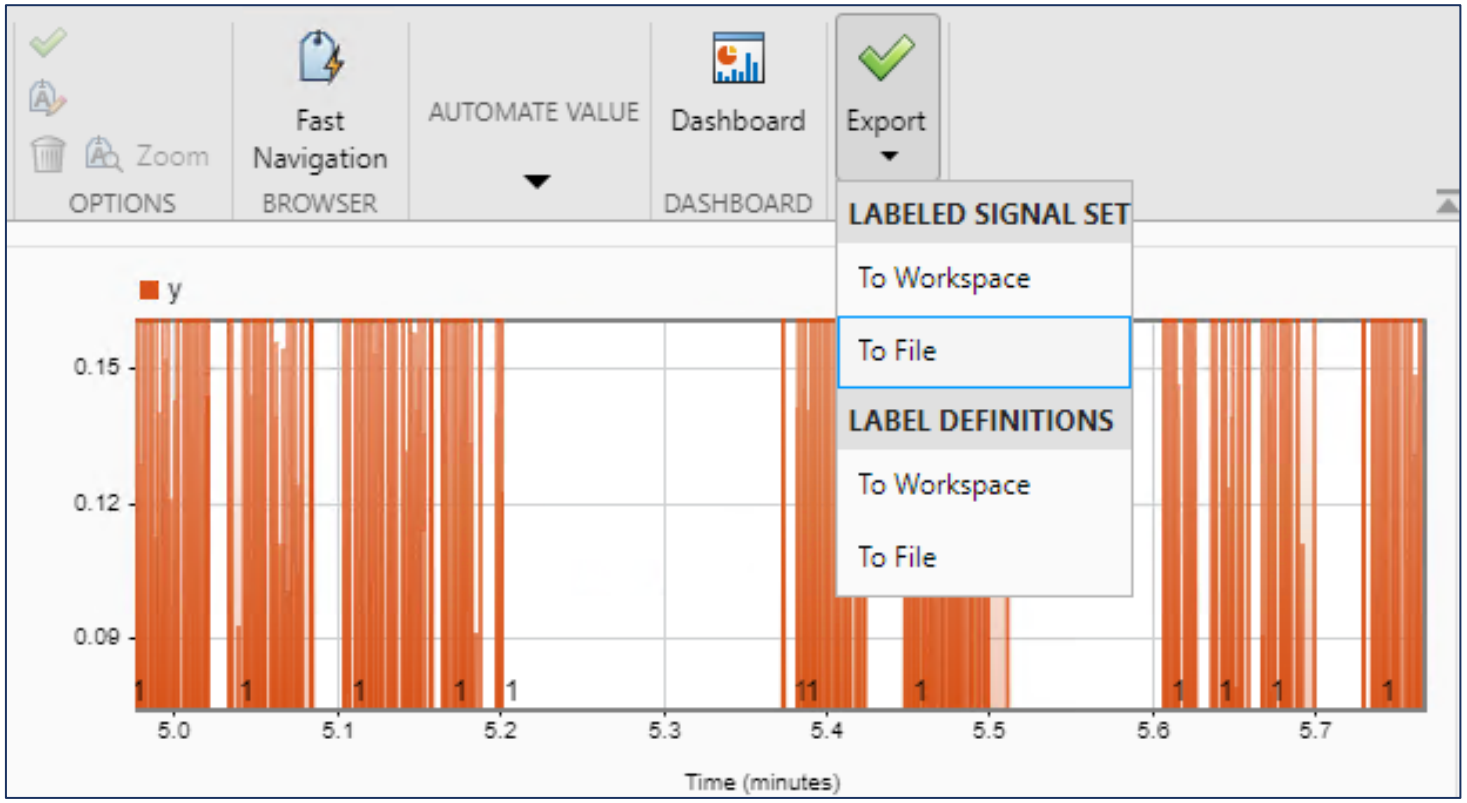


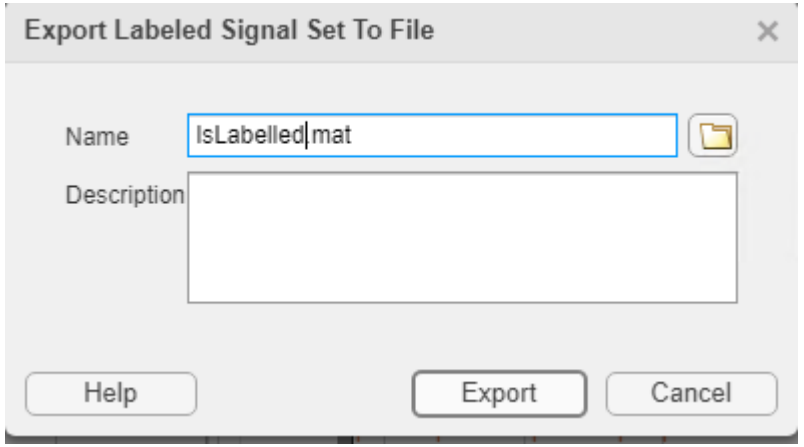
STEP 5: EXAMINING THE LABELING





STEP 6: EXPORTING LABELED SIGNAL SET





LABEL SET GENERATED

1x1 labeledSignalSet

Property	Value
SampleRate	
SampleTime	
TimeValues	
TimeInformation	"inherent"
Description	""
NumMembers	1
Source	*1x1 cell*
Labels	*1x1 table*

ls.Labels.SpeechDetected{1,1}

1 ROILimits		2 Value
36.1760	38.6880	"1"
39.3760	40.2560	"1"
40.5280	41.9520	"1"
42.6720	44.7520	"1"
45.0240	46.6880	"1"
57.6480	58.7360	"1"
58.8320	59.3440	"1"
59.5040	62.8160	"1"
63.5040	65.4720	"1"
65.7760	67.9520	"1"
91.5520	92.6400	"1"
92.7360	98.0160	"1"
98.7520	101.0400	"1"
101.3280	102.1600	"1"
108.7040	111.2000	"1"
111.7920	112.7840	"1"
113.4880	115.2000	"1"

CONCLUSIONS

- Speech recognition systems requires large amount of domain specific labeled corpus
- Automation of Labeling can provide immense benefits in significantly reducing the cycle time
- This presentation illustrated automation of Audio Labeling workflow towards audio Voice Activity Detection(VAD)

ACKNOWLEDGEMENT

- Presenters gratefully acknowledge the help & support received from Jayanth Balaji A., Application Engineer, MathWorks

THANK YOU