Kardome

KARDOME'S VOICE USER INTERFACE SOLUTION FOR INTERACTIVE KIOSK

Rev. 1.1.3

ABSTRACT

This manual describes Kardome's voice user interface demonstrator for interactive kiosks. The demonstrator includes Audio Device (Mallet™) and a PC-based monitor. The manual describes the installation of the Monitor, how to set-up the demonstrator and how to set-up a multi-user voice user interface.

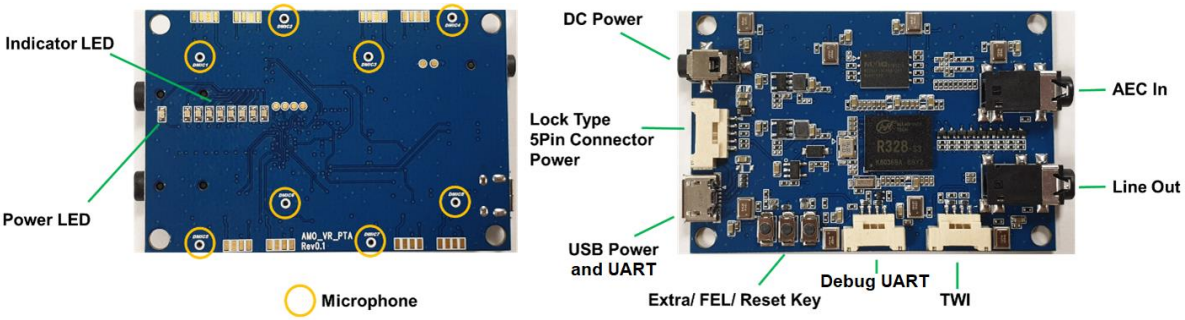
Kardome
2021

Table of Contents

1 MALLET-USB™

In the box, you will find:

1. MALLET-USB™ Microphone Array
2. micro-USB cable



Description

The MALLET-USB™ by Kardome is an 8-microphone array designed for applications such as voice user interface (VUI), acoustic surveillance, acoustic radar, hearing aids research, conference calls and sound fields analysis. 8 synchronous omnidirectional high-performance digital MEMS microphones are arranged on a surface of a rectangular board with 60*40 mm. 8 microphones are arranged in symmetric on the board. The deployment of the 8 microphones is suitable for a range of array processing algorithms including beamforming, interferometry, source localization, space-time adaptive processing (STAP), NULL steering and matched-filed processing (MFP).

Features	R328-S3 Dual-core ARM Cortex A7 Processor 8-microphone array High performance Digital MEMS microphones Low Distortion of 3% at 112.5dB SPL Signal-to-Noise Ratio of 64.5dB(A) Flat Frequency Response 20 Hz – 20 kHz 8 indication LEDs and 1 multi-purpose push-button Reset, push-button
Connectivity	USB - Micro USB USB Audio Class 2.0 driver for Windows 3.5Φ Stereo Line out UART TWI (Two Wire Interface)
Mechanical dimensions	Length and Width - 60 * 40 mm Weight – 9.7 g
Electrical specifications	Powered by USB port, DC Jack, Lock type 5 Pin Con. Current consumption - 150 ± 50 mA @ 5V Operating voltage - 5 ± 0.5 V
Environment specifications	Operating temperature - 0° to 70°C Relative humidity - 0% to 80%, non-condensing

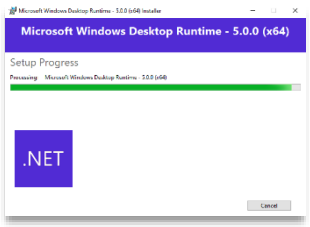
2 System Requirements

Operating System Windows 10

3 QuickStart Guide

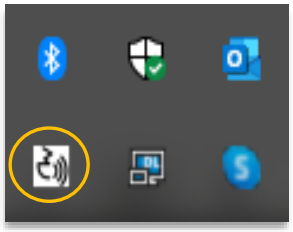
3.1 Installing the Monitor Application

Run the *Kardome.Monitor.Mallet.Kiosk.USB.1.1.3.exe* setup application. The setup will automatically install the following components:



Microsoft Windows Desktop Runtime

If not already installed



Kardome's USB Audio Device Driver

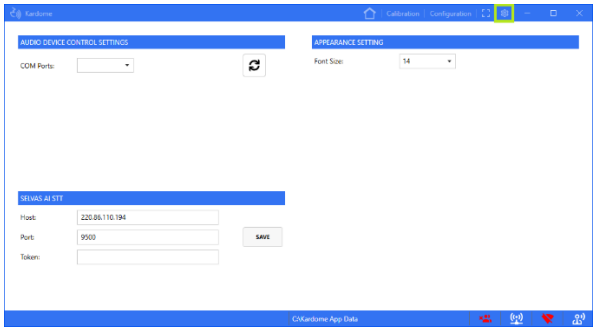
If not already installed

Kardome's Monitor

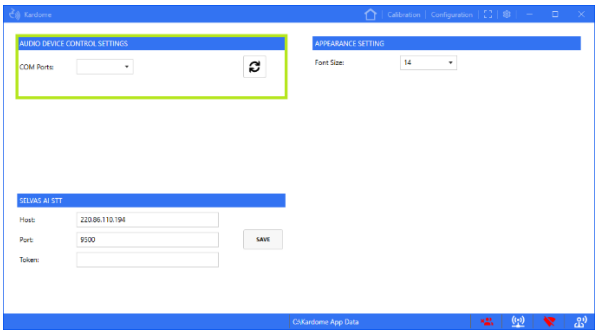
3.2 Connecting the Mallet

Connect the Mallet to the computer via USB cable. Once connected, a red indicator light would turn on.

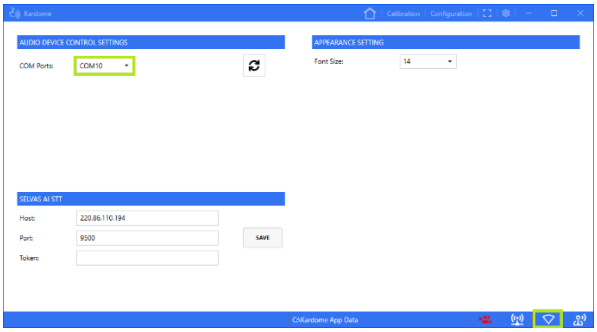
1. Select the Settings (⚙) Tab in the monitor.



2. Select the appropriate COM Port to which the Mallet card is connected. If the port is not in the list, press on the refresh (↻) button.



3. If connection was successful, the Audio Device icons in the status bar turns white. otherwise, please check chapter 5 for troubleshooting.

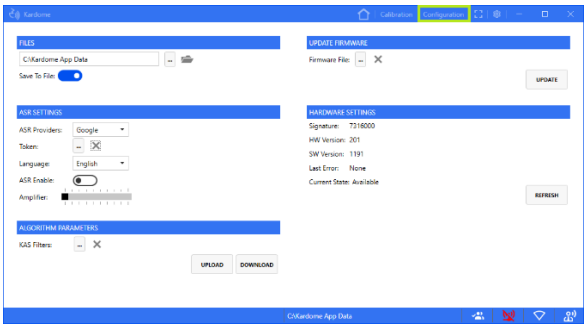


3.3 ASR Engine Setup

Note, without ASR (Automatic Speech Recognition) engine connectivity in place, the system can still process the speech signals and save them to files, provided that the option of "Save To File" is enabled in the Configuration tab.

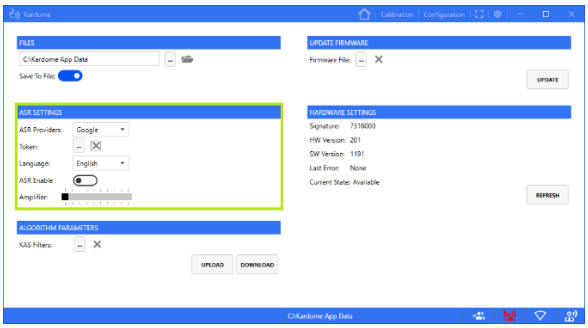
3.3.1 Google Speech to Text API

1. Select the Configuration Tab in the monitor.

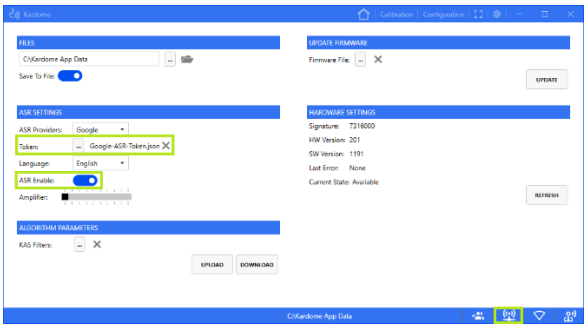


2. Configure the ASR Settings:
   - Set ASR Provider to Google.
   - Set a valid token by selecting an appropriate JSON file.
   - Select the desired Language.
   - Enable the ASR select.

   Information for making a Google token can be found here.

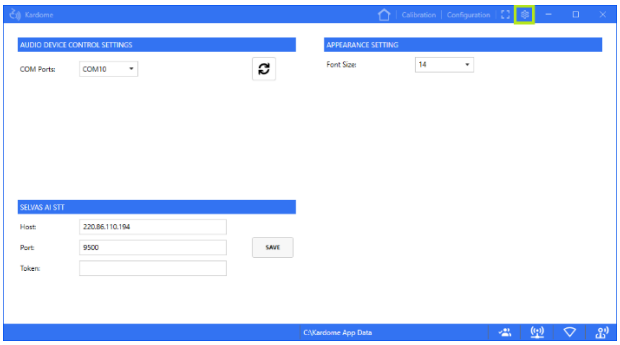


3. If ASR configuration was successful, the middle icon in the status bar turns white. otherwise, please check chapter 5 for troubleshooting.

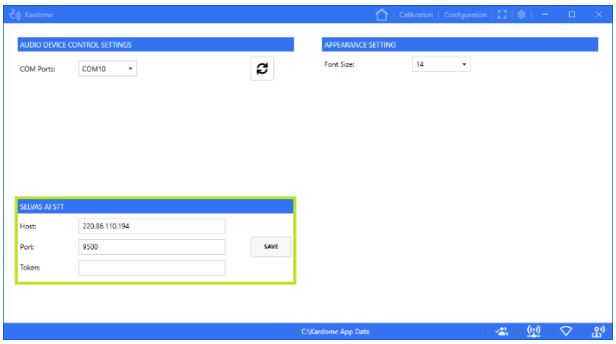


3.3.2 Selves AI Speech to Text API

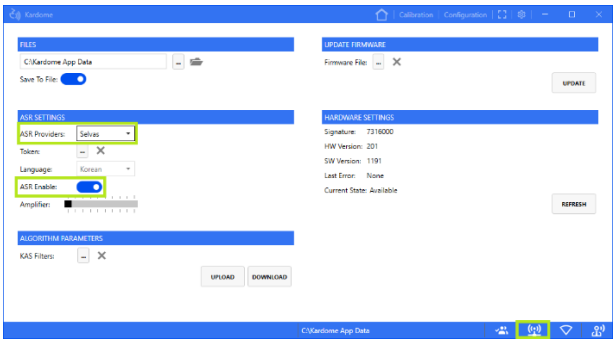
1. Select the Settings (⚙) Tab in the monitor.



2. Configure the Selvas AI ASR Settings:
   - Set the Host IP
   - Set the Host Port
   - Set a valid token
   - Save the settings



3. Select the Configuration Tab and activate the ASR enable select. If successful, the middle icon in the status bar turns white. otherwise, please check chapter 5 for troubleshooting.



3.4 Initial Calibration

During calibration and onwards, please remain static and refrain from disposition the Mallet.

It is advised that you take a few minutes to read chapter 4.2 in this manual. This chapter explains the details of the initial calibration process and define the required environmental conditions to complete the calibration successfully.

1. Select the Calibration Tab in the monitor.

2. Reset the MALLET by selecting the *Reset* button and wait until the process is completed.

3. Calibrate the initial ambient noise characteristic by activating the *Ambient Noise* switch and wait until the process is completed. Upon completion, the switch turns to blue.

Please refer to chapter 4.2. A green indicator light would turn on as soon as the Mallet start recording.

4. Calibrate the initial user acoustic signature by activating the *User* switch and wait until the process is completed. Upon completion, the switch turns to blue.

Please refer to chapter 4.2. A green indicator light would turn on as soon as the Mallet start recording.

3.5 Performance Evaluation

1. Select the Home Tab in the monitor.

2. To use the Kardome algorithm turn on the 'Kardome Enable' button. If the button is turned off the Mallet will be used as a passive microphone array.

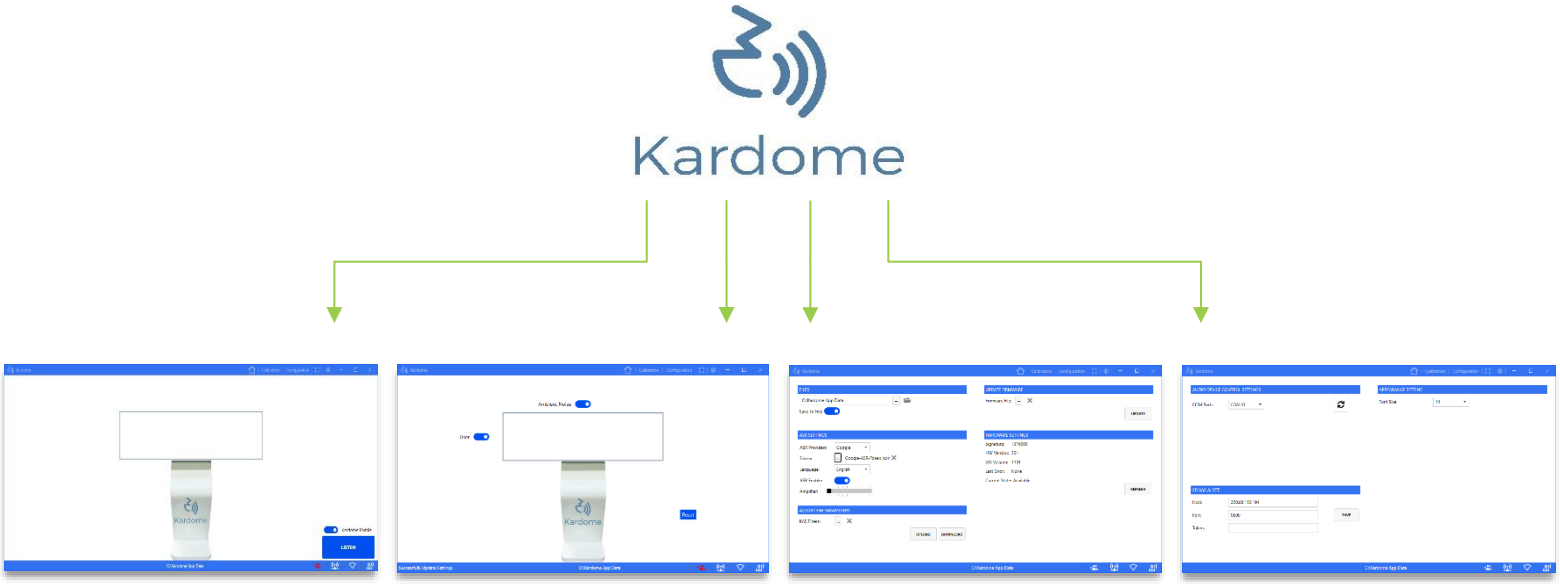
3. Before Activating the LISTEN button, keep quiet for 2-3 seconds.

Those 2-3 seconds allow the algorithm to take in mind the current environmental noise.
If the speaker will speak through those 2-3 second, the algorithm will consider him as a noise and the results will suffer.

4. Activate LISTEN button, wait until "Listening…" appears at the lower left side then ask a user to speak his commands from the exact same position as calibrated. You should be able to see the text in the display window.

All the audio files being saved in the working directory.

4 Monitor and Calibration Process Specifications

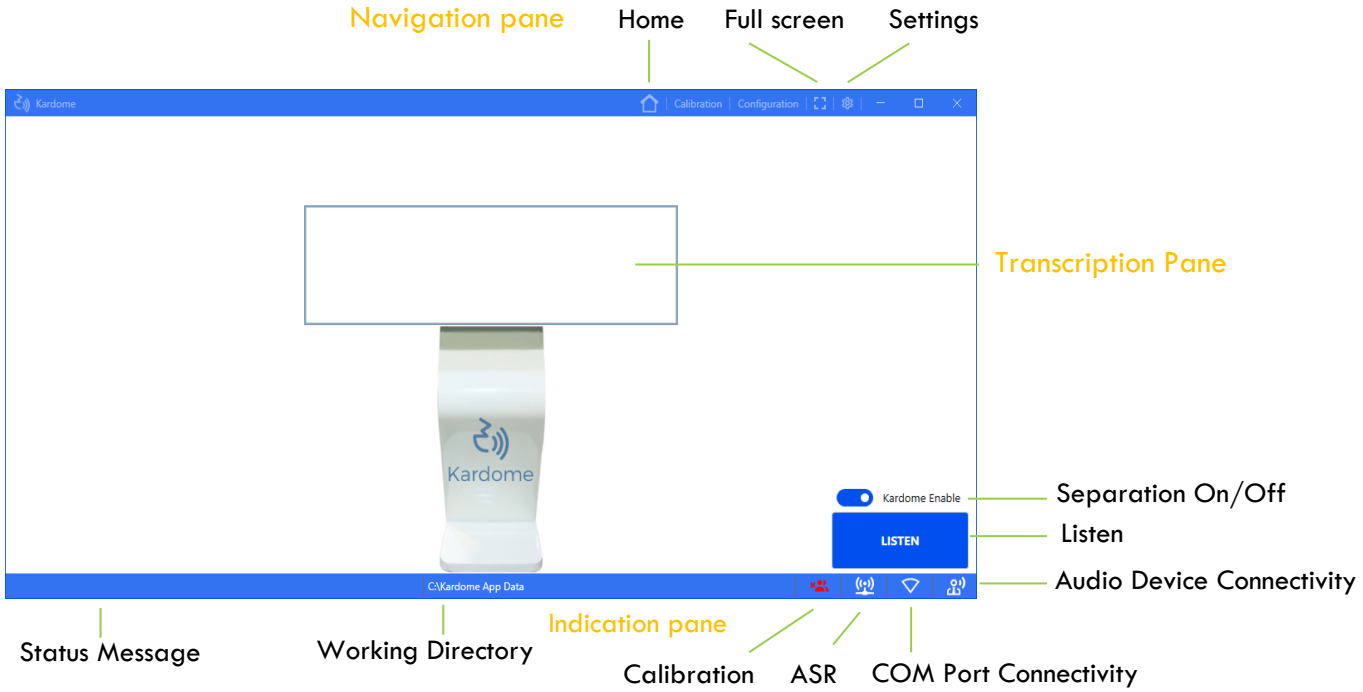


Home	Calibration	Configuration	Settings
Operate the multi-user voice user interface	Measure the acoustic signature of the speech sources	Configure the system	Network, Appearance and Selvas.AI Settings

The demonstrator can do any combination of the following operation:

1. Separation - if Kardome switch (Home tab) is enabled
2. Transcription - if ASR token is provided and ASR is enabled (Configuration tab)
3. Saving to file - if Save to File is enabled (Configuration tab)

4.1 Home tab



Listen

Start processing the audio signals from the microphones. The operation may be any combination of Separation, Transcription and Saving to file depends on the respective switches.

Kardome Enable

Activate/deactivate Kardome's algorithms for speech separation of simultaneous speech from several sources.

Calibration indicator

To be calibrated, the ambient noise and acoustic signature of at least one speech source should be measured.
Without calibration, you may not be able to operate the system while separating the speech sources. However, you may still be able to operate the uncalibrated system without separation by disabling the Kardome Enable switch.

ASR Indicator

Google ASR (Automatic Speech Recognition) is used to transcribe the speech signals and to notate the transcription on the transcription pane. To this end, in the Configuration tab, there should be a valid ASR token linked to the monitor, and the ASR switch should be enabled.

Note, without ASR, the system can still process the speech signals (separation) and save them to files, provided that the option of "Save To File" is enabled in the Configuration tab.

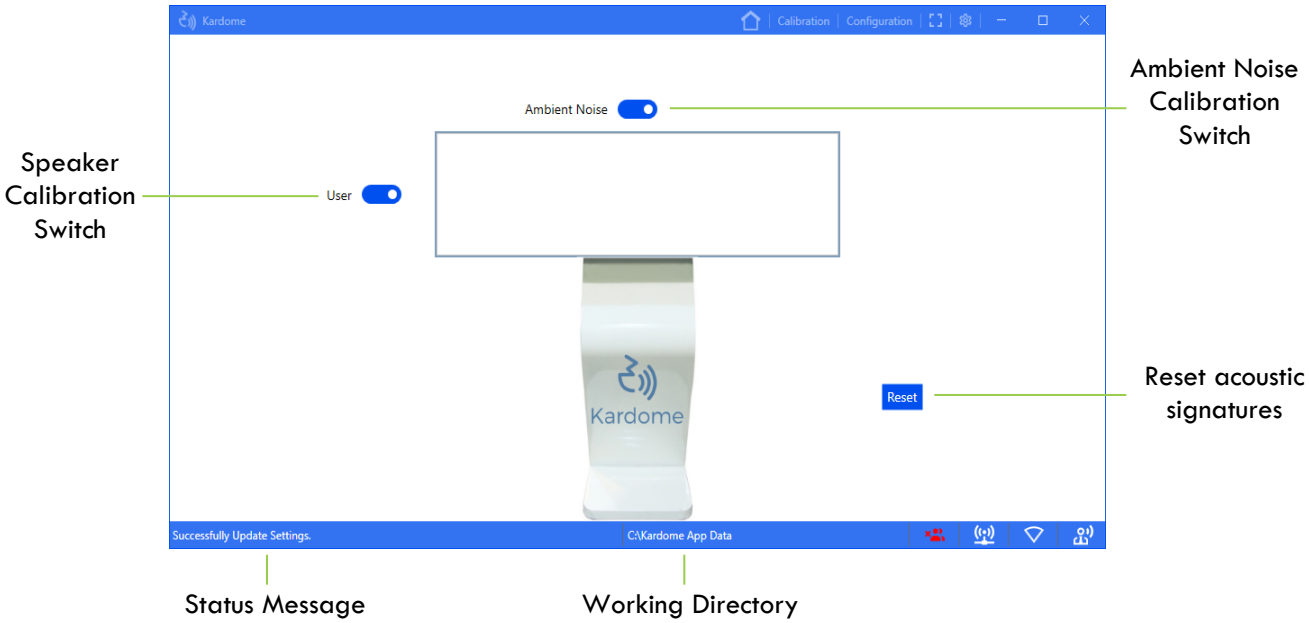
Audio Device Connectivity Indicator

For the system to work, the Mallet should be connected to the PC. If the icon is red, please check chapter 5 for troubleshooting.

COM Port Connectivity Indicator

A suitable port to which the Mallet is connected need to be selected in the Setting Tab. If the icon is red, please check chapter 5 for troubleshooting.

4.2 Calibration tab



(Initial) Ambient Noise Calibration

Must be calibrated first at all terms. It is used to estimate the acoustic signature of the user when it is being calibrated.

It is advised that during the calibration, the characteristic of the ambient noise is as similar as possible to the expected characteristic during the user calibration.

Here it must be taken in mind to be quiet and not speak until the dialog box disappears. If the calibration is spoiled, you may repeat the calibration by turning the Ambient Noise Calibration switch Off and then On.

(Initial) User Calibration

Measure and estimate acoustic signature of the user.

You will need a help of a colleague to complete this task.

Please ask your colleague to stand in the typical location in front of the Kiosk and to start speaking continuously, and then turn the respective switch On. Through the whole calibration, the speaker should speak continuously, for example: count from 1 to 30 without breaks.

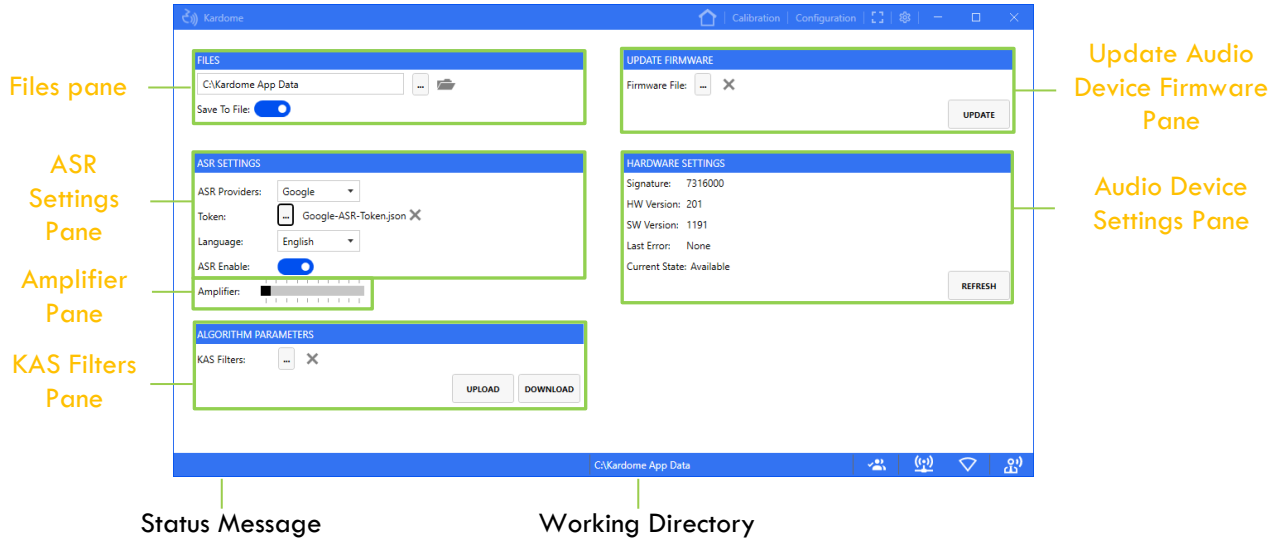
Please verify that only a single speaker is speaking, and the speaker does not move significantly while calibrating.

If the source calibration is spoiled, you may repeat the calibration by turning the Source Calibration switch Off and then On.

Reset

Reset the ambient noise measurement and the acoustic signature of the user. Reset the calibration when starting a new scenario, or when the performance is insufficient.

4.3 Configuration tab



Files

Upon enabling the Save To File option, the Monitor may save the processed speech signals to wav files located in the selected directory.
When Kardome is Enabled (in the Home tab), the speech signals are separated, the separated signals are saved into separated wav files with the name:
Kardome_Audio_yyyy-mm-dd_hh-mm-ss_out_1.wav
where yyyy, mm, dd, hh, mm and ss stand for the year, month, day, hour, minute and second of the beginning of the operation.
When Kardome is Disabled (in the Home tab), the speech signals are not separated, the raw signals from the eight microphones are saved into a single wav files with the name:
Kardome_Audio_yyyy-mm-dd_hh-mm-ss_out_1.wav
with the previous conventions.
During calibration it will create a wav files of the defined duration length with the name:
Kardome_Audio_Calibration_yyyy-mm-dd_hh-mm-ss_out_#.wav
with the previous conventions.
Where 5 is ambient noise, and 1 is the speaker.

ASR Settings

ASR Provider
In the ASR provider you may choose between the Google and Selvas AI speech to text services.

Token (required for Google Speech to Text only)

The Token is a JSON file (stands for JavaScript Object Notation), which is used to enable Google Speech to Text. Creating and activating a gcloud service account will give you the required JSON file.

Language
In the Language select list you may choose the language to transcribe.
The languages that currently exist are:

- Google: English, Korean, Japanese, Chinese.
- Selvas AI: Korean.

ASR Enable
If ASR is disabled, then the transcription pane in Home remains empty. However, the processed speech signals are saved to files this option is enabled in the Files pane.

Update Audio Device Firmware

This pane is used to upgrade the firmware of the Audio Device. If you received a new release of the firmware, which is a file with the extension *.ldr, you may select this file by browsing to its location, and then pressing UPLOAD.

Note, for the update to take effect, the Audio Device should be reset by pressing the reset button or by disconnecting its USB cable and reconnecting again.

Caution! Any *.ldr file can be uploaded into the flash. However, uploading *.ldr file which was not provided by Kardome may invalidate the Audio Device.

Amplifier

The volume of the speaker can be amplified up to 10 times when 1 is the default value.

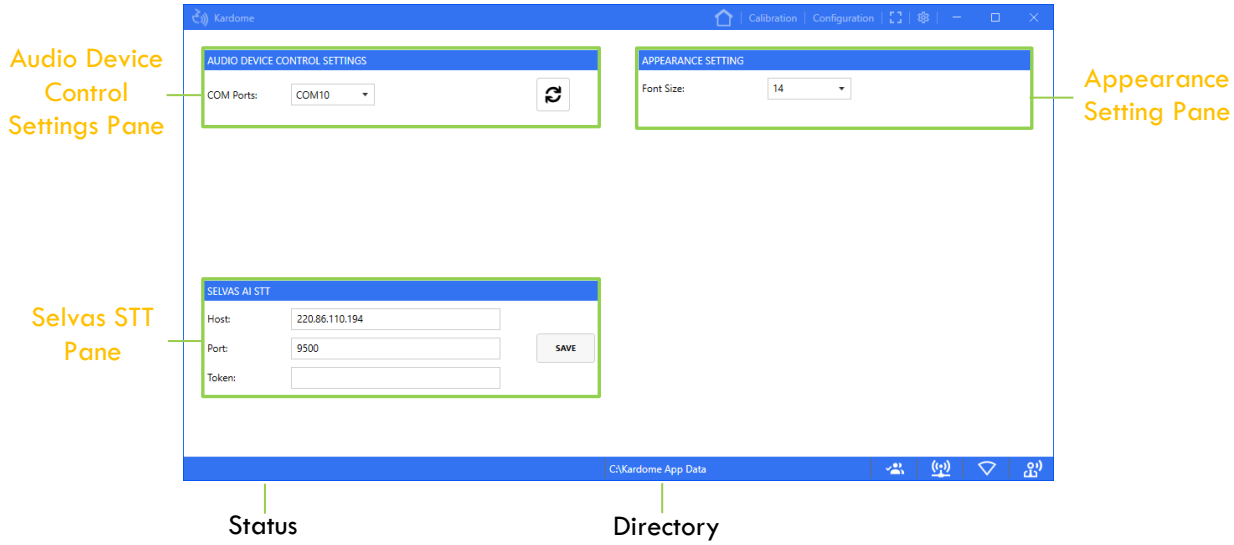
KAS Filters

This functionality is used to upload/download an initial parameter set, applicable when preparing the MALLET for deployment/mass-production.

Audio Device Settings

At this pane you can see the relevant setting of the device that connected to the system. The Hardware and Software version, the current state and the last error that occurred.

4.4 Settings tab



Audio Device Control Settings This pane is used to select the appropriate port to which the Mallet card is connected.

Selvas STT This pane is used to set the Host IP, Port, and the Token of Selvas to get their transcription services.

Appearance Settings This pane is used to set the font size of the transcription at home screen (between 14-24).

5 Troubleshooting

Problem observed	Suggested action
Calibration indicator is red 	This means that the Audio Device is not calibrated. 1. Go to the Calibration tab and measure the ambient noise and the user. Please refer to chapter 4.2 for instructions.
ASR indicator is red 	This means that the Monitor cannot use Google speech-to-text to transcribe the separated speech signals. 1. No internet connection. Set up or verify internet connection. 2. No ASR token is available in the Configuration tab to operate Google speech-to-text. Obtain and provide a valid ASR token in the Configuration tab. 3. The ASR is disabled in the Configuration tab. Enable the ASR in the Configuration tab.
Audio Device Connectivity indicator is red 	This means that the Monitor cannot communicate with the Audio Device. 1. Check the USB connection to the Audio Device. 2. Check the power LED and the working status LED on the Audio Device.
COM Port Connectivity indicator is red 	This means that the Monitor cannot communicate with the Audio Device. 1. Select the appropriate port to which the Mallet card is connected in the Settings tab.