The Desktop Audio Situational Awareness device monitors the surrounding environment for active audio sources including speech and other audio events. The device is intended to enable the Deaf and Hard of Hearing (DHH) to know when someone is speaking, what they are saying and where they are located in group discussions such as in the classroom, workplace and at social gatherings.

**Audio Situational Awareness**

Audio Situational Awareness provides spatial detection to identify and associate successive audio signals from the same source to present a consistent view of the audio environment. Audio sources are detected by a microphone array and the source direction is determined using Time Delay of Arrival (TDOA). The difference in TDOA of audio signals detected at different microphones in the array is used to determine the direction of the audio source relative to the microphone positions.

Microphone array beamforming is used to capture the audio signals originating from a detected source. Audio signal events from speech sources are translated to text while other audio sources may be identified by their distinct characteristics.

**Desktop Audio Situational Awareness Device**

The Desktop Audio Situational Awareness Device is similar in design to an Amazon Alexa device. The device case will be 3-5 inches in diameter and ½ inches thick. Thinner and smaller is better for portability.

Rounded edges are desired on both bottom and top with vent holes around the perimeter sufficient in size on the bottom for cooling (if necessary) and on the top for audio detection. There are two controls for power on/off and audio detection on/off (mute).

The device electronics comprise a single All-in-One motherboard (MB) with an internal battery as a standalone power source. A cost/ production design tradeoff may be to use a general purpose MB (e.g. Raspberry Pi) with a separate daughter board for the microphone array. This may impact the diameter and thickness of the device.

The All-in-One MB includes a low-power CPU (e.g. Arm Cortex A7 1.5 GHz quad-core processor with 1 GB memory), an array of six (6) MEMS microphones equally spaced around the perimeter, Bluetooth Low Energy / Wi-Fi (802.11n), a micro SD slot, and a USB port for programming, external power and charging. Wireless (Qi) charging is desirable. **The Respeaker Core V.2 is an example of a third party Alexa style MB.**

The battery power supply should be sufficient for at least 4-6 hours of continuous operation, although a longer period of 10-12 hours is desirable. Wireless and fast recharging capabilities are also desired.

**System Requirements**

The ASA will use the ReSpeaker Core vers. 2.0 as its hardware platform.

C++ will be the programming language of choice for the ASA.

Listen Mode

The ASA shall monitor the surrounding audio environment for speech sources. This will be called listen mode.

The ASA shall be able to detect speech audio. The goal is to detect speech up to a range of 30 feet in a classroom environment.

The ASA shall determine the location (direction of arrival) of a speech source relative to the device.

The ASA shall track active speech sources by location.

The ASA will track active speech sources as they move relative to the device.

A speaker identifier shall be assigned to each speech source.

Multiple speech sources collocated within 10 degrees of each other shall be considered one speaker. If there is a way to distinguish between collocated speakers, they may be assigned individual identifiers.

The ASA will convert speech to text for each speech source. A wireless connection to the Internet is permitted to achieve this.

The ASA shall start processing speech when the device is powered up.

User Interface

The user will address the system through a Bluetooth connected device, preferrably an Android tablet or phone.

The ASA shall transmit the speaker ID, location and text results to a Bluetooth connected device. Strong consideration should be given to encrypting this over-the-air transmission.

The tablet shall log the transmission record in a timestamped text file, displayable on request.

To enable debug capability, all Bluetooth-routed messages shall also be available on a monitor, and all Bluetooth-sourced commands shall also be sourced from a keyboard, connected directly to the ReSpeaker Core.

Record Mode

The ASA shall record an active speech source for a fixed period. This will be called record mode.

If a fixed period of silence (no active speech) is detected, the ASA shall stop recording and return to listen mode.

The recorded speech audio shall be translated to text using an online Speech-To-Text (STT) service.

Configuration Mode

The user shall have the ability to control the ASA device via the Bluetooth connected device.

The user shall have the ability to save a speech transcript to a file.

The user shall have the ability to reset or clear a speech transcript.

The user shall have the ability to select the language to be translated.

The user shall have the ability to select the language to be displayed.