Indoor Acoustic Localization Project

Bowei Liu, Chen Yang, Zecheng He

Team D10

Carnegie Mellon University

March 8th 2018

# Project description

With both home and working environment getting smarter these days, interaction between people and smart devices in a convenient way is expected to greatly enhance work efficiency and make people more comfortable. Imagine if you are able to control the room temperature, adjust lighting, or open the door for someone by just simply sitting at your desk and voicing our thoses commands.

Our project aims to build a localization system using acoustic information to allow people in the room to issue different voice commands. The environment/smart devices will respond accordingly by performing the given task to suit a person’s exact need without being confused by which individual in the room is actually giving out the commands.

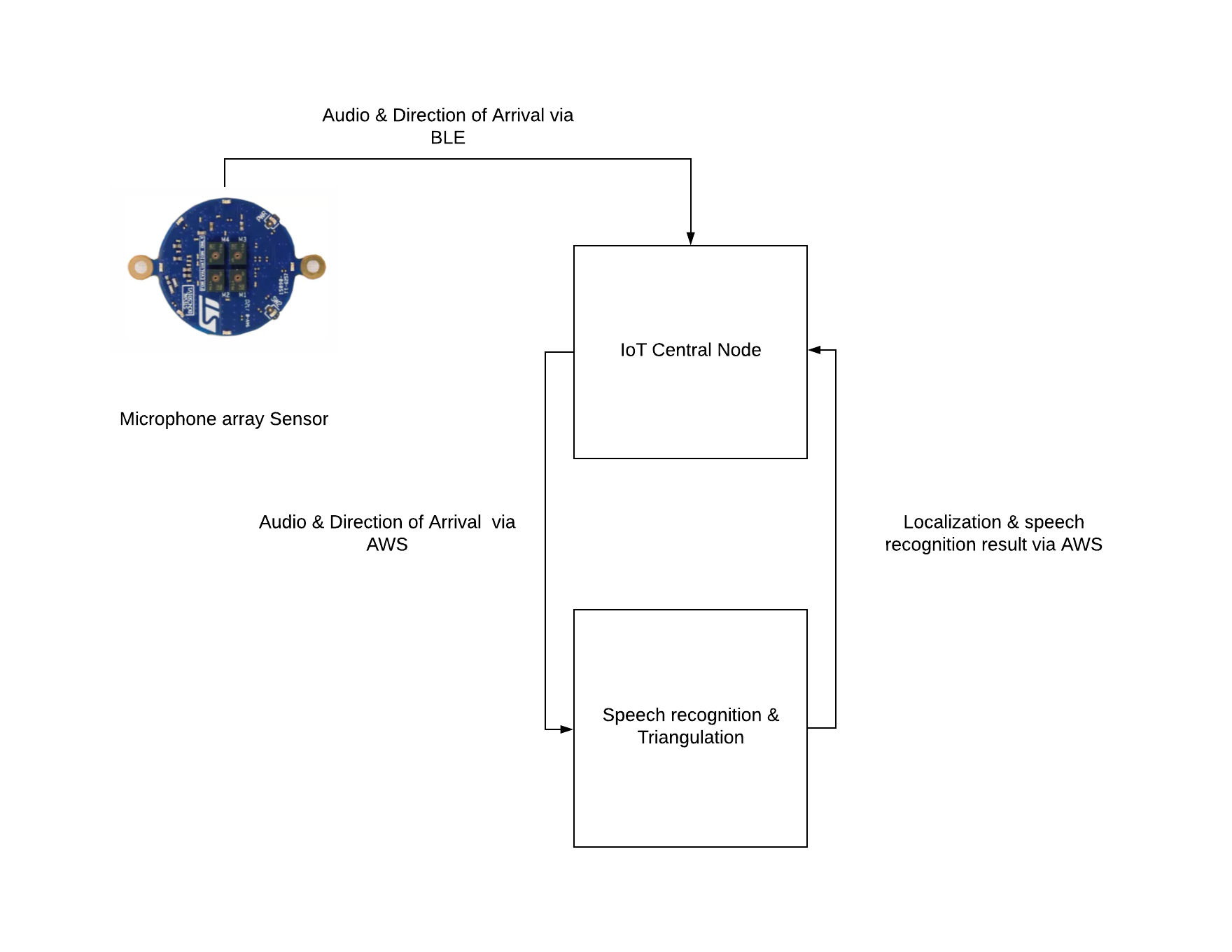
# Design requirements

The network of sensors should be able to localize the sound source accurately enough to distinguish personal spaces (less than 20 cm away from the person) of two different people.

Every sensor node should be able to stream audio information to the IoT central node in real time via BLE.

The IoT central node should be able to upload sensor outputs (localization and audio) to AWS and obtain speech recognition result after processing with a tolerable delay (less than 500 ms).

# Functional Architecture

****

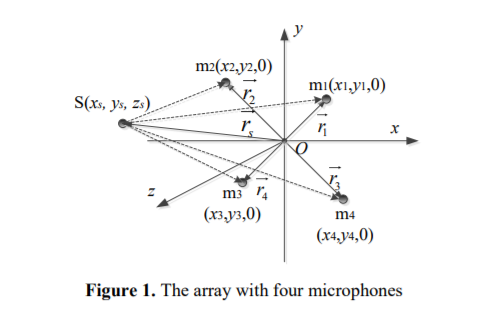
In this design, each microphone array sensor collects audio information and calculates a direction of arrival (DOA) of the sound source. Multiple sensor nodes transmit these audio and DOA outputs to the IoT central node which is connected to WiFi. The outputs are forwarded by the IoT central node to the AWS for speech recognition and triangulation processings. The result is sent back to the IoT central node for application actions.

# Design trade studies

Time difference of arrival (TDOA) is an alternative localization method to DOA. In this case, only one microphone will be used on each microphone array sensor node.

**The triangulation method using an array with four microphones [1]**

The array is positioned in the x-y plane, with an X shape, as shown in Fig. 1. The microphones are numbered as 1 to 4. The microphone mi (i =1~4) is located at a fixed position, with coordinate (xi , yi , 0). The sound source is located at arbitrary point in the space, with coordinate (xs, ys, zs). i r (i =1~4) and s r represents the vector from the origin to microphones or the sound source.



Microphone m1 is chosen to be the reference microphone. The geometry relationship can be written as



where j = 2, 3, 4, 1, j t is the TDOA between mj and m1. If 1, j t is positive, mj is farther to sound source than m1, and negative means mj is nearer to sound source. And c is the sound speed, which is considered as a constant. Eq. (1) can be rewritten as



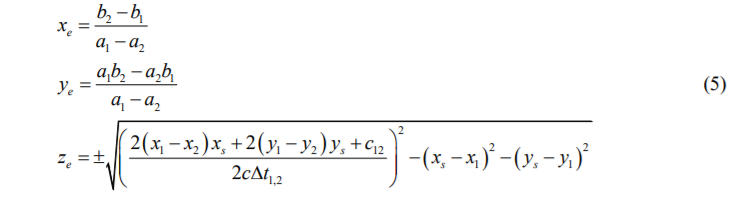
The square of Eq. (2) can be written as



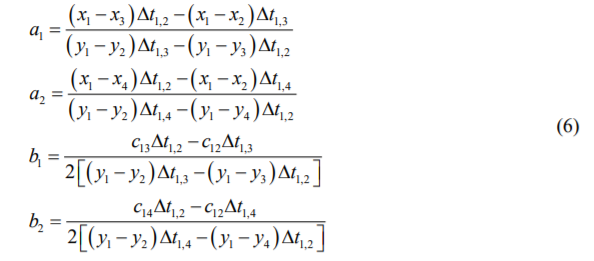
where



Eq. (5) can be obtained deriving from Eq. (3) :



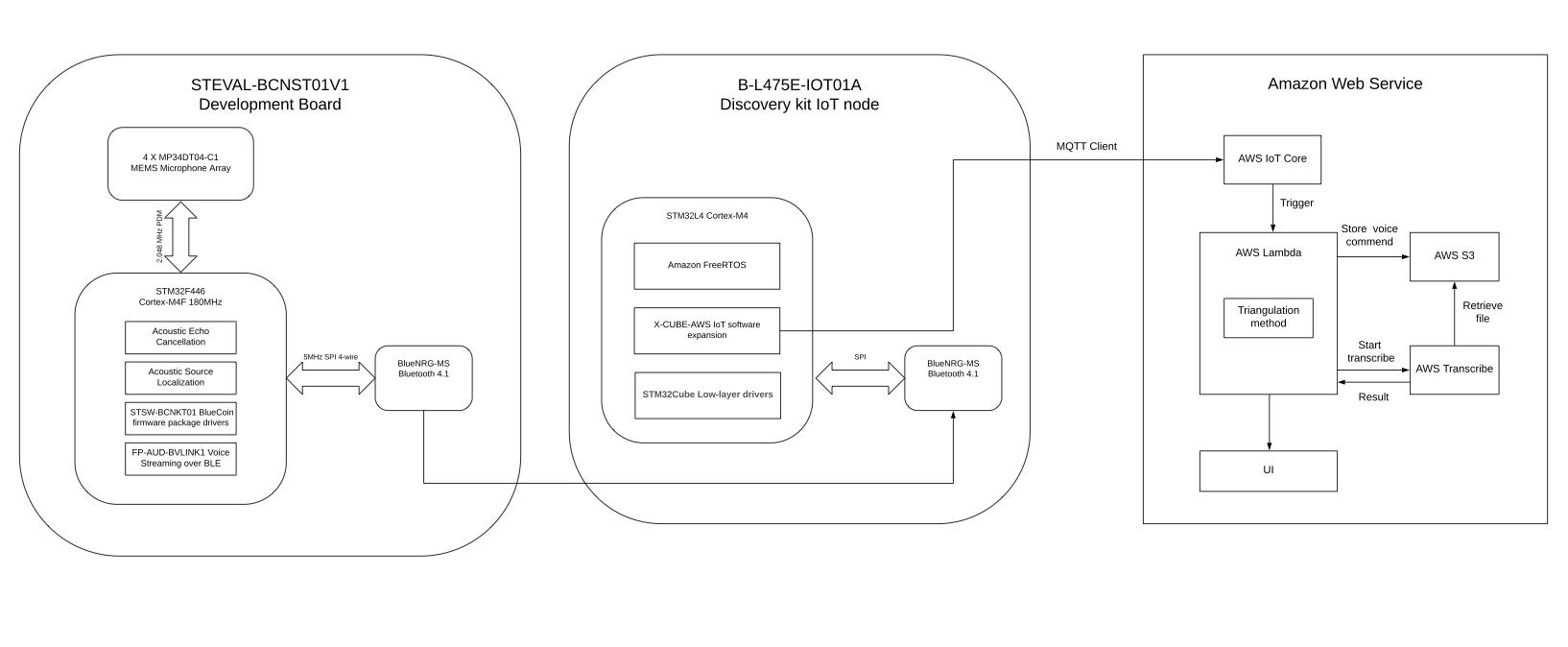
Where



The estimation of coordinates of sound source is presented as Eq. (5).

However this approach heavily depends on the time synchronization of the four microphones in the array in order to achieve an acceptable accuracies. Direction of arrival (DOA) is not as time sensitive as compared to TDOA.

# System description/depiction

****

## Microphone array sensor

On each sensor node, there is a 4-microphone array. Using the 4 sound signals collected from 4 microphones, we apply the GCC-PHAT algorithm to calculate the direction of arrival of the sound source. After calculation, the DOA output and the audio output from the microphone that is closest to the sound source will be transmitted to the IoT central node via BLE services.

## 

## Amazon Web Service

Through MQTT client, IoT central node sends data collected from all four Bluecoin sensors, which includes direction of arrival and users’ speech, to AWS IoT Core. Upon arrival, this event will trigger AWS Lambda. It is a serverless compute service that runs our code without provisioning any servers. AWS Lambda charges on a subsecond metering, which saves us from the cost of running idle servers.

Firstly, the Lambda function will store users’ speech in AWS S3 bucket(file storage service), then it will run the triangulation algorithm described in the next section. The algorithm yields location of the speaker, and Lambda function is able to choose the speech data that’s collected from Bluecoin sensor closest to speaker’s position. This data will be used by AWS Transcribe, which provides and result of speech recognition and gives us the text of user’s commend. Eventually, the location as well as the text of user’s speech will be presented in the UI interface.

## Localization algorithm

We are planning to try out two approaches. One approach is using time difference of arrival (TDOA) as described above in the design trade section. We are also going to try using the DOA information given by the four microphone arrays to find a intersection of the four directional output. Practically there will not be a perfect point intersection among the four direction but instead an area enclosed by the four directional lines. More exploration is needed to narrow down to an acceptable smaller area to identify individuals.

# Project Management

In order to keep our team project on track, we’ve adopted Asana[2] as project management software. For each task, there’s a clear task description, due date and assignee.



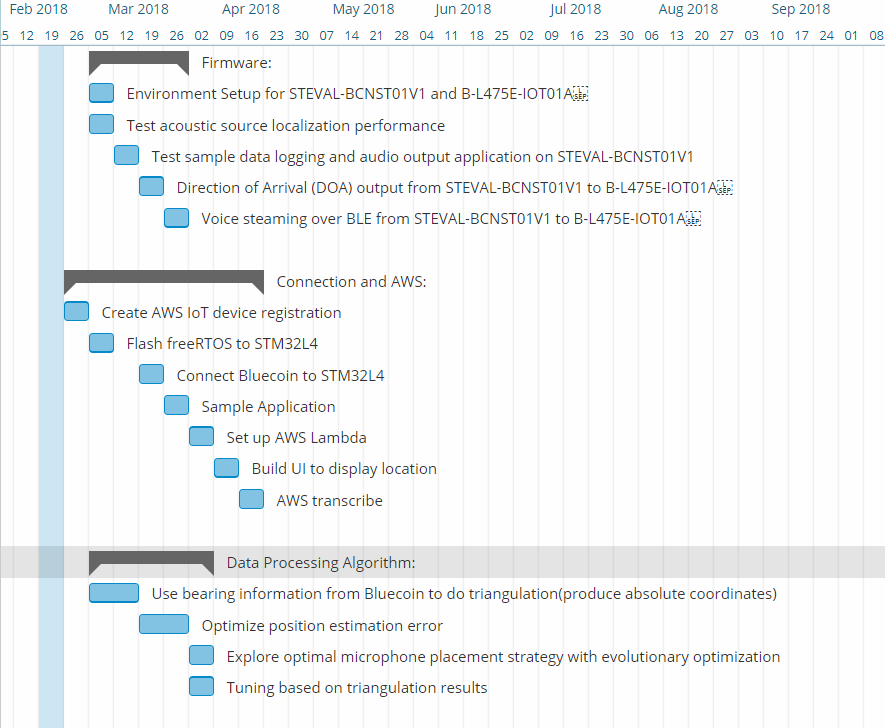
For example, this is the task of sending Directional of Arrival from the BlueCoin sensor(STEVAL-BCNST01V01) to STMicrocontroller(B-L475E-IOT01). It is assigned to Chen and has a due date of March 21st.

## Schedule

Here’s the complete schedule from Asana.



Another view of the timeline.



## Team member responsibilities

As described above, tasks are planned out in Asana with due date and assignee. In general, Chen will be focusing on the firmware, which includes four BlueCoin sensors(STEVAL-BCNST01V01) and one STMicrocontroller(B-L475E-IOT01). He’s also responsible for comparing different algorithms that acquire direction on arrival, and filtering out frequencies that are not in the range of human voice.

Zecheng will set up the communication between STMicrocontroller and Amazon Web Service(AWS) as well as computation running on AWS. This includes AWS IOT registry, Lambda Service and S3 bucket. He will also build the UI interface, which can visualize speaker’s location.

Bowei will work on data processing algorithm and testing for a optimal deployment configuration for the sensor array. For data processing, raw output from each sensor like acoustic signals and time difference of arrivals are transformed into bearing information which can determine from which direction the sound source is located. Using bearing information from the four sensor arrays, a triangulation algorithm will be developed to provide the exact position of the sound source. Finally a optimization method will be applied to get the deployment configuration which achieves the best localization accuracy.

# Related work (Competition)

[3] proposed a new indoor-localization method based on acoustic direction finding, which requires users to shake the phone while walking for a period of time. This scheme utilizes relative shift and velocity of the phone in order to find the direction of sound source. Although it achieves decent accuracy, this method takes a considerable amount of time due to shaking and walking. On top of that, the computing power of smartphones limits this scheme’s real-time performance. Our system is built on low-cost embedded devices instead of hand-carry smartphones, and it provides instant location of the speaker. Since all heavy-loaded computation is pushed to the cloud, users can expect immediate real-time experience.

There are also other approaches to localize the speaker, including [4]. Even though the use of infrared signal achieves high accuracy, it requires specialized devices and does not provide speech information.

# References

[1] Miao, Yang, Wang, Wen, Wang, Lian. (2014). A Triangulation Method Based on Phase Difference of Arrival Estimation for Sound Source. *The 21st International Congress on Sound and Vibration.* Retrieved from <https://www.iiav.org/icsv21/content/papers/papers/full_paper_447_20140129161505281.pdf>

[2] Retrieved from <https://app.asana.com/0/571424715060633/571424715060634>

[3] Huang, Xiong, Li, Lin, Mao, Yang, Liu. (2013). Accurate Indoor Localization Using Acoustic Direction Finding via Smart Phones. *Networking and Internet Architecture, Cornell University.* Retrieved from <https://arxiv.org/abs/1306.1651>

[4] Want, Hopper, Falcã, Gibbons. (1992). The active badge location system.

*Journal ACM Transactions on Information Systems,* Volume 10, Issue 1:91-102. New York, NY, America.

|  |
| --- |
|  |