Angle-of-Arrival (AoA) Triangulation

Amod Agrawal
University of Illinois at Urbana-Champaign
amodka2@illinois.edu

Mariam Vardishvili
University of Illinois at Urbana-Champaign
mariamv2@illinois.edu

ABSTRACT

Indoor localization has been a sought after problem in the area of mobile computing because it promises to enable location-aware and intelligent applications. As the smart assistant systems like Amazon Alexa and Google Home are gaining popularity, they are becoming an integral part of our daily lives and and home management. In this project, we place three Alexa devices at various positions in the room and localize the user based off her speech signal that is heard by all the devices.

1. MOTIVATION

The problem in this project is location triangulation with respect to the three anchors (microphone arrays) in the room. Such a solution can solve indoor localization and enable location intelligent application for smart home devices. For example, if the Alexa can localize you in the room - it can perform actions for your devices which are close to you. Location of the user provides a context-rich environment for the Alexa applications that can immensely improve the smartness and user experience of the device.

2. PROBLEM STATEMENT

The goal is to develop an algorithm which computes the location of the human speaker in a room with three Alexa devices (microphone-arrays) placed at pre-defined locations. Given that we know the locations of the three microphone arrays along with their dimensions and specifications, we use the signal recorded at these three arrays to compute the location of the signal source.

Assumptions: In this project, we assume that the source of the signal (the speaker) along with the three arrays lie in the same plane. That is, our entire search space is reduced to a 2D plane. However, we do experience multipath of the signal which is incident at the arrays with some elevation. We also assume that we donâĂŹt know the source signal and that all the microphones are clock synchronized.

DOI: 10.1145/1235

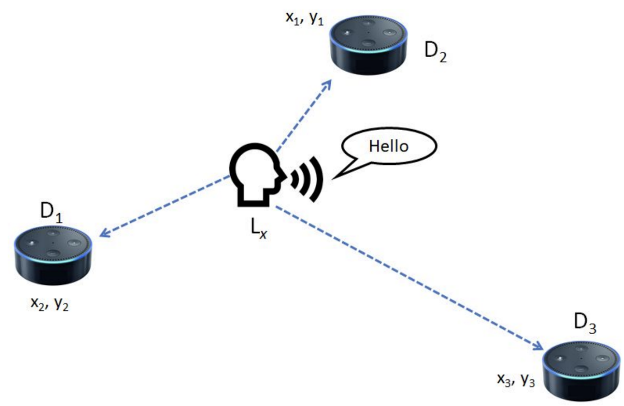

Figure 1: System Setup

3. APPROACH

Our approach to solving this indoor localization problem involves using the audio signals recorded at the three microphone arrays. The core idea is to compute the angle of arrival (AoA) at the three arrays and use these angles to extrapolate and find the location of the user.

4. SYSTEM SETUP

We first define the system set up for this project. The setup consists of three microphone arrays with six microphones each placed at three locations given by $\langle$(x1, y1), (x2, y2), (x3, y3)$\rangle$. Figure 2 shows the setup of the microphone array.

The pairwise angle between the microphones is 60 degrees and theyâĂŹre arranged in a regular hexagonal structure of side length d = 4.75cm. The figure also shows the orientation of the microphone array. We measure θ as the angle with the positive x-axis.

We find the location of each microphone as:
x-coordinate location of mic = x_1 + d $\times$ $\cos(\theta)$
y-coordinate location of mic = y_1 + d $\times$ $\sin(\theta)$
where x_1 and y_1 are the center coordinates of the microphone array.

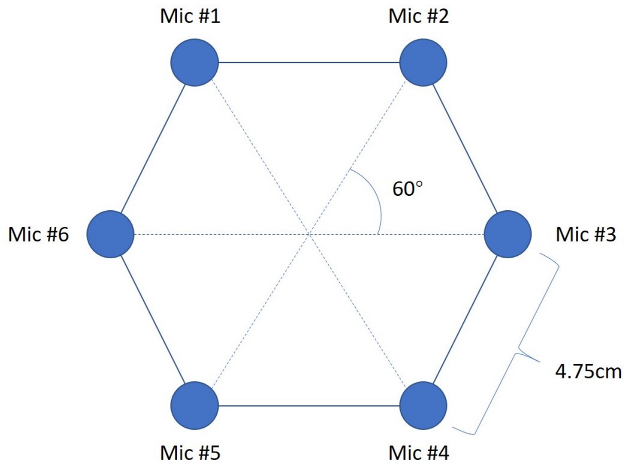

Figure 2: Microphone array configuration

5. PART 1: ANGLE-OF-ARRIVAL (AOA) ESTIMATION

AoA estimation is the core focus of this project. It is challenging to accurately estimate the AoA of a speech signal using a microphone array for several reasons:

1. Multipath and echoes of the source signal along with the background noise can heavily pollute the received signal.
2. Array dimensions are very small, with a radius of 4.75 cm, which limits our resolving power.
3. To accurately estimate the AoA, we need a spatial resolution of around 1mm. At a sampling rate of 44.1kHz (sampling rate of provided data), and using 343m/s as the speed of sound, our spatial resolution is only around 7.77mm, several times too large. In such cases, naive techniques for AoA estimation, such as delay-and-sum work poorly.

We use the Generalized Cross Correlation (GCC) method to solve the AoA problem. The algorithm is formally known as GCC-PHAT, Generalized Cross Correlation Phase Transform, which is precisely used to find the delay between two signals. We also perform an approximated quadratic splining function to the correlation result to find the time delay at the sub-sample level.

Since the given audio signal is a wide-band signal - it consists of frequencies ranging from 0Hz to 22.05KHz. However, most of the human speech lies in the spectrum of 500Hz to 1.5KHz, which is an important factor that can be used to pre-process the signal to discard frequencies that can pollute the AoA correlation.

The fundamental idea of AoA algorithm is that each microphone at the array will receive the same signal from the source, however, due to spatial distance between microphones: each microphone will receive the signal with a delay. Microphones closer to the signal source receive the signal first and the microphones farther away receive it later. For each direction from the array, we can calculate a unique set of delay values for each microphone irrespective of the source signal. Using the recorded signals we can compute our estimated delays and then do a space search over a dictionary of all directions to find the one closest to our computed values.

5.1 Algorithm

We define our algorithm as follows to solve the problem:

1. Preprocess the input signals
2. Find delays between signals using GCC-PHAT for every pair of microphones
3. Convert the delays into a displacement vector
4. Do a space search for 0 to 360 azimuth and 0 to 90 elevation to find the closest displacement vectors
5. Find the peak where the distance is the minimum and plot AoA spectrum along with heatmap
6. Use the three angles alpha1, alpha2, alpha3 from all three microphone arrays to formulate three line equations
7. Solve the line intersection problem by formulating it as an optimization problem and finding the least squares solution

Instead of using just one microphone as a reference microphone and finding five time delays with respect to it, we find delays between every possible pair of microphones (36 pairs). This increases our distance dimension space from R^5 to R^{36}.

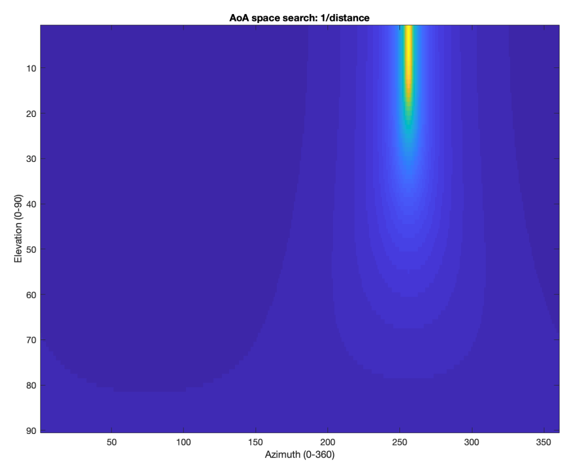

Figure 3: Angle of Arrival spectrum: precise

6. GENERALIZED CROSS CORRELATION PHASE TRANSFORM (GCC-PHAT)

The most error-prone step in our algorithm is to compute the time difference of arrival (TDoA) between two microphones. The TDoA represent the lag between the audio signals recorded at two different microphones. The most

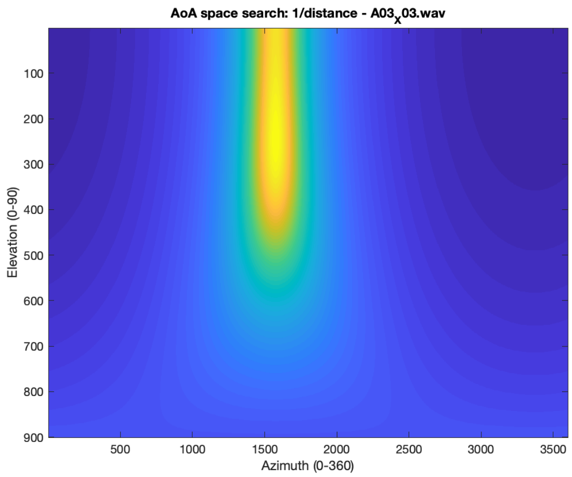

Figure 4: Angle of Arrival spectrum: polluted by multipath

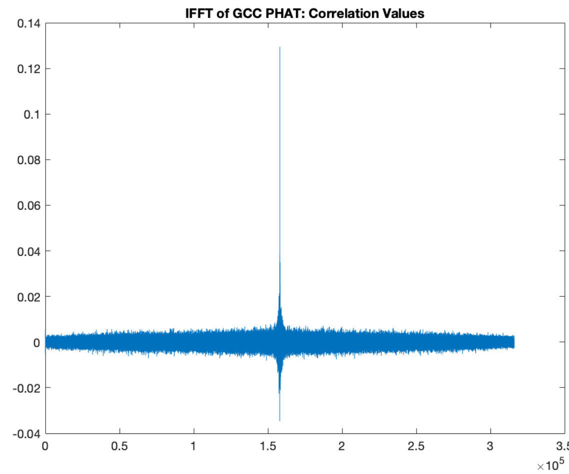

Figure 5: GCC-PHAT Result: Correlation Peak

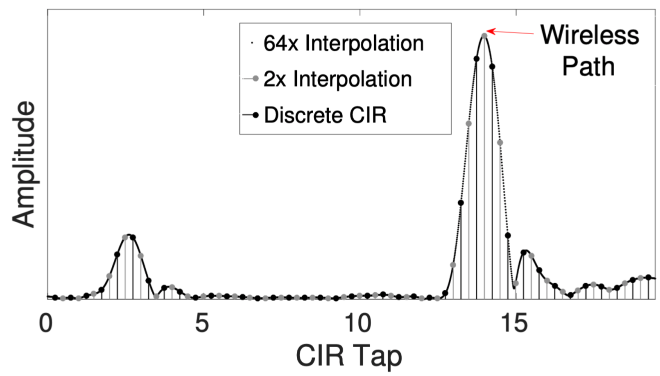

Figure 6: Interpolation affecting the maxima peak

common method of estimating TDoA is cross-correlation in time-domain, however, it often produces poor results due to signal pollution by noise and multipath. It also yields poor time-resolution. GCC-PHAT performs correlation in the frequency domain and "whitens" the signals by normalizing their spectral power in the FFT.

GCC algorithm uses Fourier transform representation of the two signals to find the delays between the line-of-sight paths of the signal. The highest peak in the correlation spectrum represents the LOS path delay. The smaller peaks may represent the delays in multipath signals which can be very arbitrary. We follow the algorithm as defined here [2] to implement it.

There is work which performs GCC-PHAT in time domain as well but due to maturity of this technique we use the frequency domain version. Frequency domain version is also faster and the weighting scheme can easily be modified to give more weight to human-centric frequencies in the range of 500Hz to 1500Hz.

7. INTERPOLATION

We interpolate the correlation values near the maxima peak to precisely find the maxima point. We add zeroes between correlation points and then run a low-pass filter. This interpolates the values at sub-sample time resolution.

8. QUADRATIC SPLINING

A closed form of quadratic splining is used to interpolate between the neighboring points of the max peak in the correlation spectrum. There are many ways to interpolate and precisely find the sub-sampled value - we use the following closed form to find the offset from the max peak:

$$\delta t = \frac{1}{2} \times \frac{(x_1 - x_2)}{(x_1 - x_2 + 2x)} \tag{1}$$

where x is the max peak point, and x1 is the point prior to x and x2 is the point after x. The above equation is derived from a generic quadratic interpolation between two points and represents its approximated closed form [1].

9. PART 2: LOCATION ESTIMATION IN A 2D SPACE

We use the angles provided by the Part 1 (AoA estimator) to formulate equations of three lines in the 2D space. With three linear equations, our goal is to find the point of intersection. However since AoA resolution is poor, there may be errors in the angles from part 1 and the lines may not intersect at one point, because of this we are deriving a least-squares solution to solve the problem.

Given the center of the array of microphones a=$[a1, a2]^T$ and an angle , a line can be defined by next form:

$$p = a + tn, -\infty \leq t \leq \infty \tag{2}$$

where n=$[cos\theta, sin\theta]^T$ is a direction vector of the line.

We solve the intersection of the three lines problem by formulating it as an optimization problem: minimize the sum of distances from the solution point to all the lines. As mentioned, we are deriving solution using the least-squares solution for the intersection of lines.

We get the squared perpendicular distance $D(p;\ a,n)$ from a point to a line:

$$D(p;a,n) = ||(a-p) - ((a-p)^T n)n||_T^2 \tag{3}$$

Given 3 lines we can find the point that is closest to all the lines by minimizing the sum of squared distances:

$$\hat{p} = argminD(p;a,n), A = \{a_1, a_2, a_3\}, N = \{n_1, n_2, n_3\} \tag{4}$$

where,

$$D(p;A,N) = \sum_{i=1}^{3} D(p;a_i,n_i) = \sum_{i=1}^{3} (a_i - p)^T (I - n_i j^T (a_i - p)) \tag{5}$$

To find the optimal p we need to take the derivative with respect to p:

$$\sum_{i=1}^{3} -(I - n_i j^T (a_i - p)^T = 0 \tag{6}$$

From here we get a linear system of equations that can be solved directly:

$$Rp = q \tag{7}$$

$$R = \sum_{i=1}^{3} (I - n_i j^T), q = \sum_{i=1}^{3} (I - n_i j^T) a_i \tag{8}$$

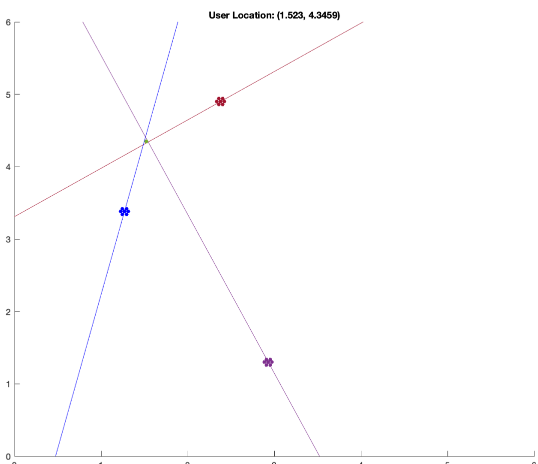

Figure 7: User Location: Localization using triangulation

10. EVALUATION

We calculate the L2 norm error for our localization points using the 10 samples of ground truth. We get a median localization error 7cm of and a mean localization error of 11cm. Since the core error in this problem is introduced from poor angle of arrival estimate, we also find the AoA estimation error. The median angle estimation error is 1.04 degrees and mean angle estimation error is 1.84 degrees.

During the exam, the errors for home test cases and exam test cases was 7cm and 18cm respectively.

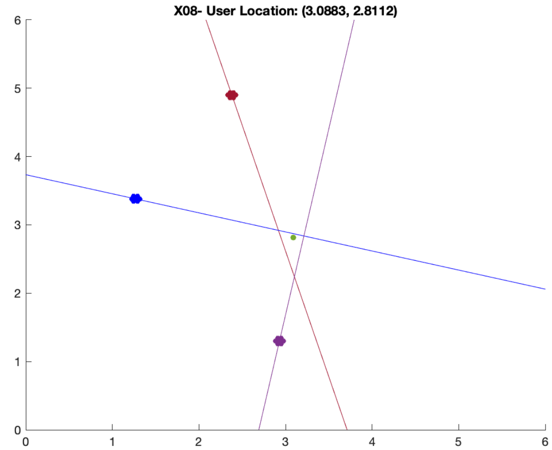

Figure 8: User location with angular error (least squares)

11. REFERENCES

[1] Quadratic interpolation using splining.

[2] C. Knapp and G. Carter. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 24(4):320–327, August 1976.

[3] W. Mao, J. He, and L. Qiu. Cat: High-precision acoustic motion tracking. In *Proceedings of the 22Nd Annual International Conference on Mobile Computing and Networking*, MobiCom '16, pages 69–81, New York, NY, USA, 2016. ACM.

[4] Z. Zhang, D. Chu, X. Chen, and T. Moscibroda. Swordfight: Enabling a new class of phone-to-phone action games on commodity phones. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, MobiSys '12, pages 1–14, New York, NY, USA, 2012. ACM.