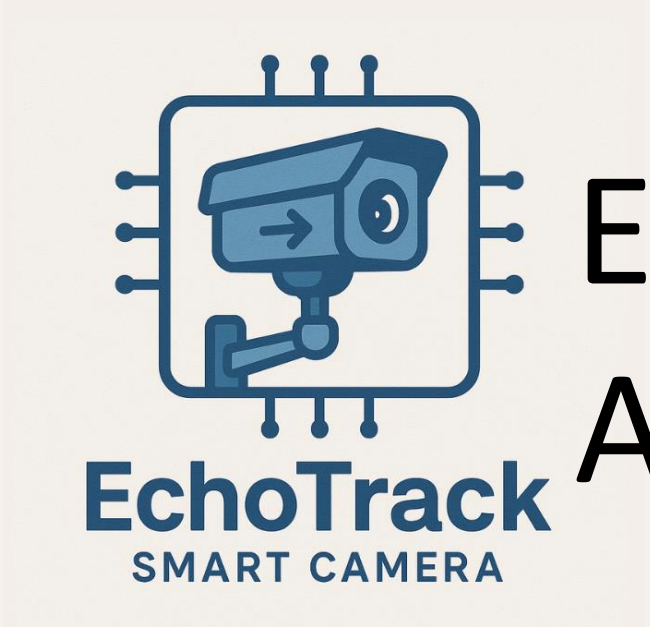


Echo-Track:
AI Smart Camera

Team Members

Jonah Cheyette	Rohan Koshy	Keerthi Kalyaan	Andrew Chang	Yuxiang Zhang
				
Peripheral Integration, CAD design	DSP, Hardware Integration	Software, ML	DOA Algorithm	Software, ML

01

Introduction

Introduction

- “The Cocktail Party Problem”
- Example: Board meetings happening over video call
    - Camera always be on who’s speaking
- Edge AI in an embedded system

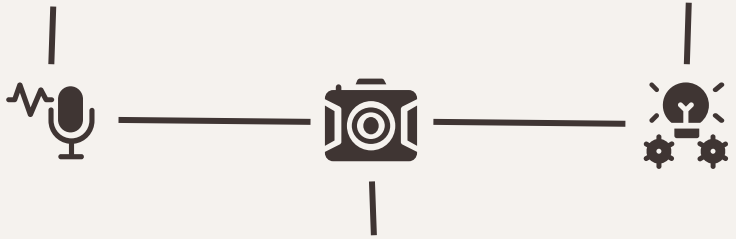
Solution

Microphone and DOA

Finds the loudest speaker in a room

AI Model

Uses keypoint detection to monitor speaker(s)



Motor Points Camera

Towards the DOA predicted Direction

02

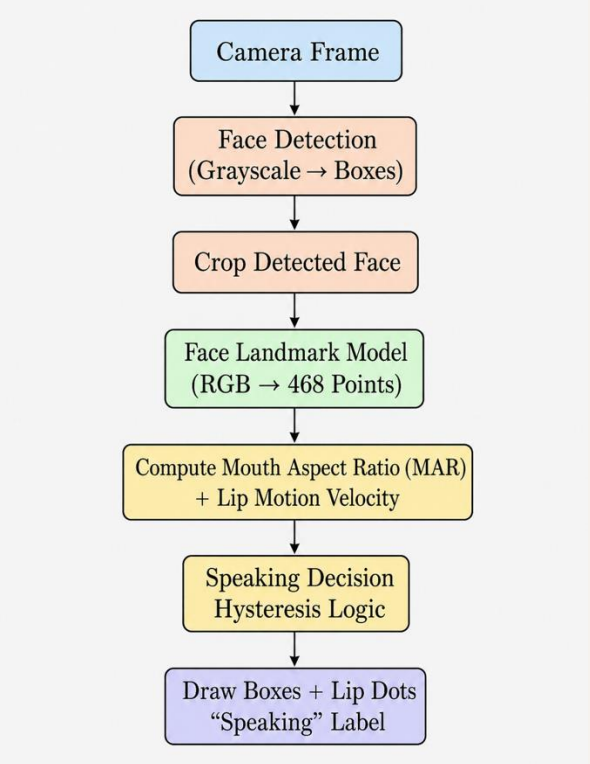
Software

Software Workflow

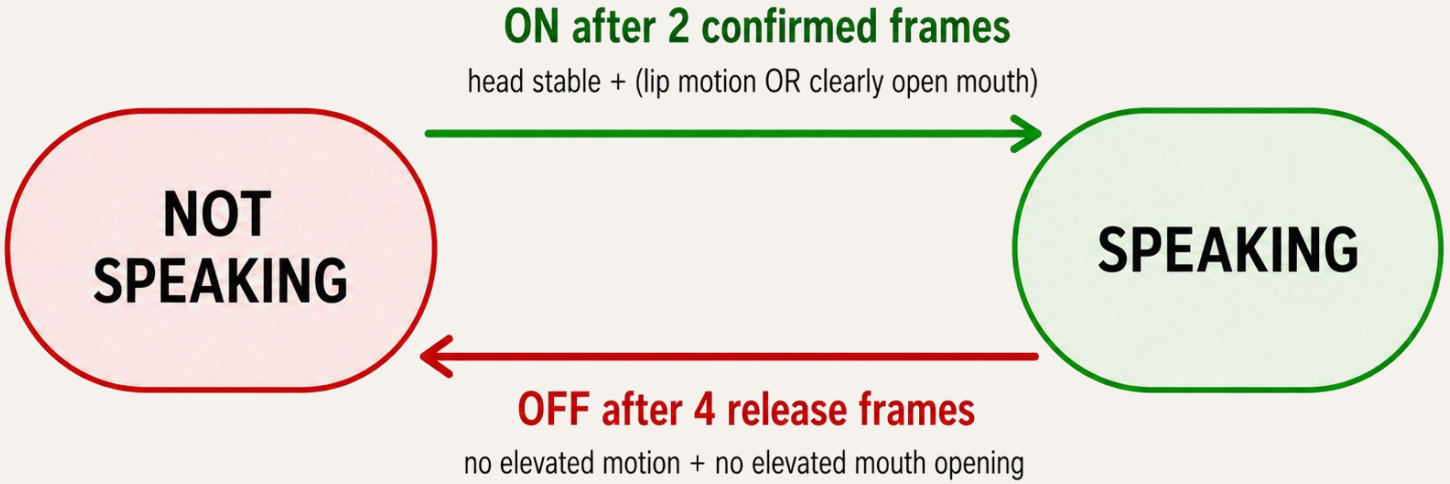
- **Inputs**: camera frames
- **Outputs**: face boxes + lip points + “Speaking” label overlay

Models:

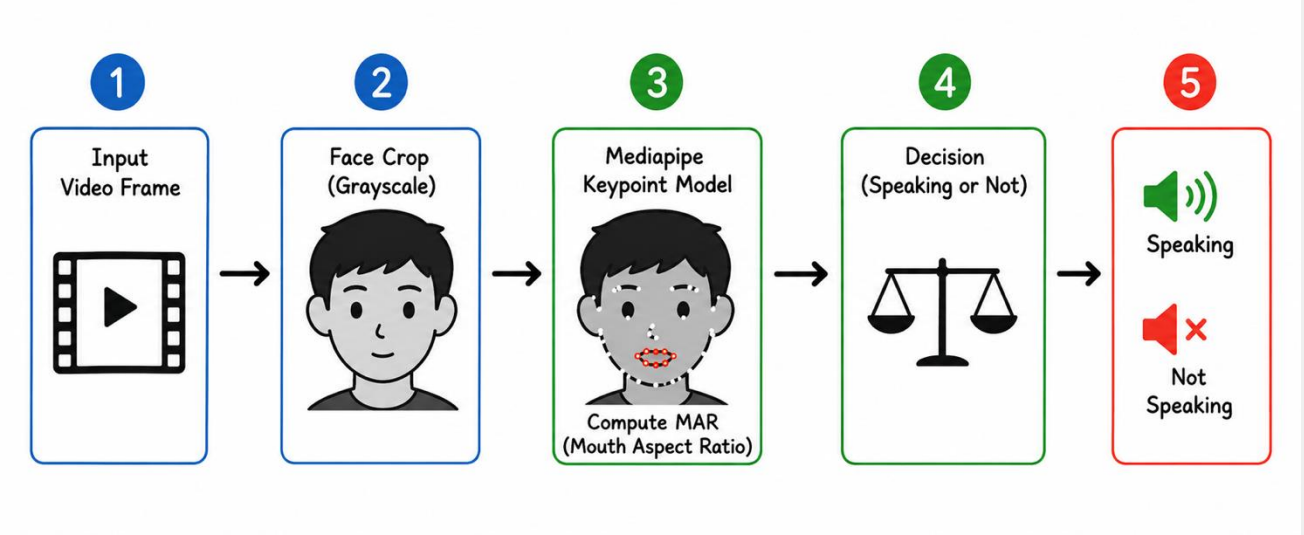
- **Face detection** (grayscale)
- **MediaPipe Face Mesh** (468 landmarks, RGB) -> Crop down landmarks to 24



Architecture



Visual Architecture

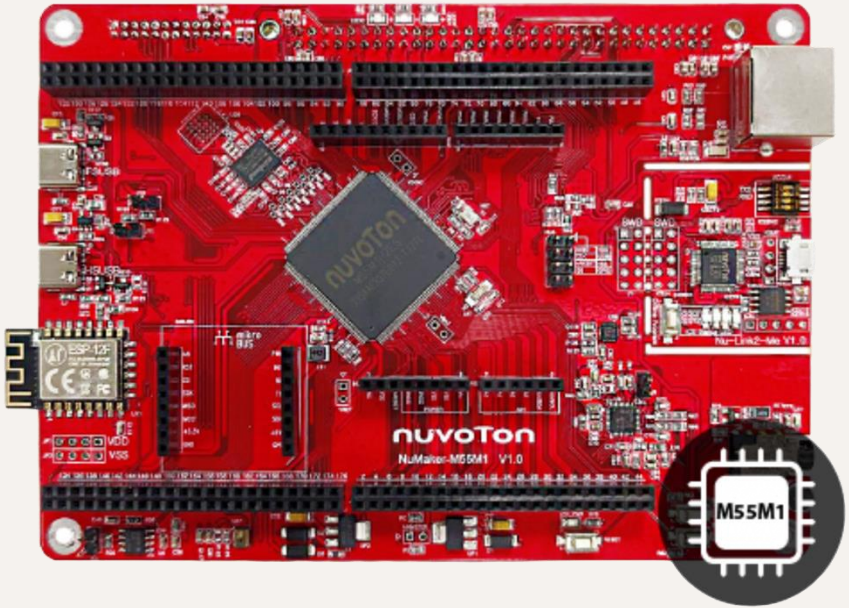


03

Hardware

M55M1 Board

- Nuvoton MCU includes
  - ARM Cortex-M55 CPU
  - ARM Ethos-U55 Micro NPU
- Memory
  - 1.5 MB SRAM
  - 64 MB HyperRAM
- 2 PDMA (16 channels each)



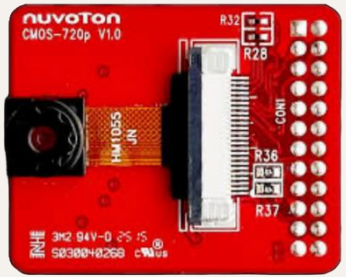
Peripheral Components

Motor:

SF3218MG Sunfounder
Digital Servo

Microphones:

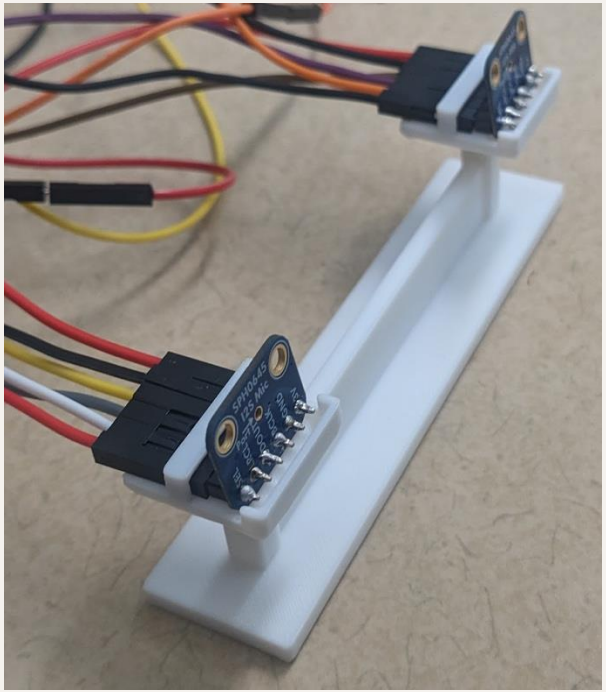
Adafruit I2S Breakout
Boards x 4

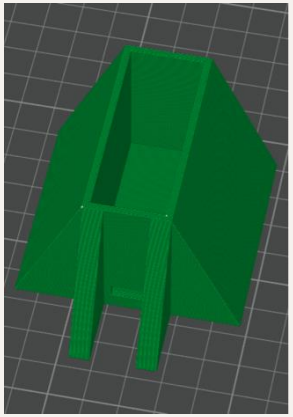


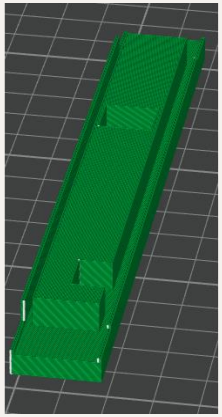
Camera:

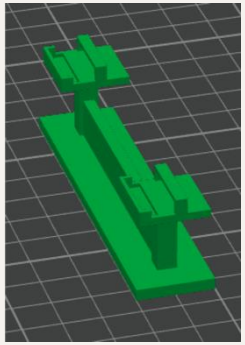
Nuvoton 720p CMOS
Camera

3D Printed Assembly







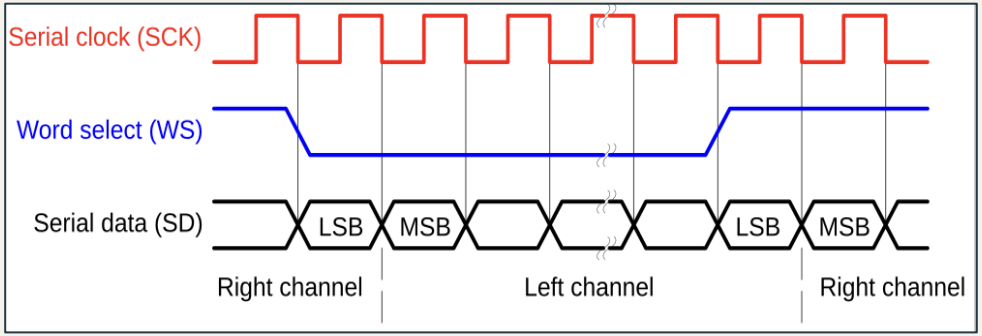


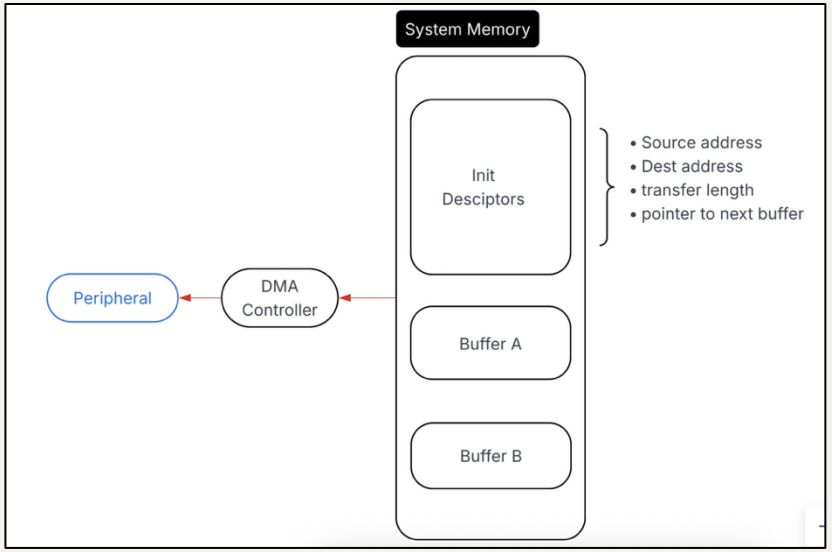
Hardware Diagram

Microphone Connection

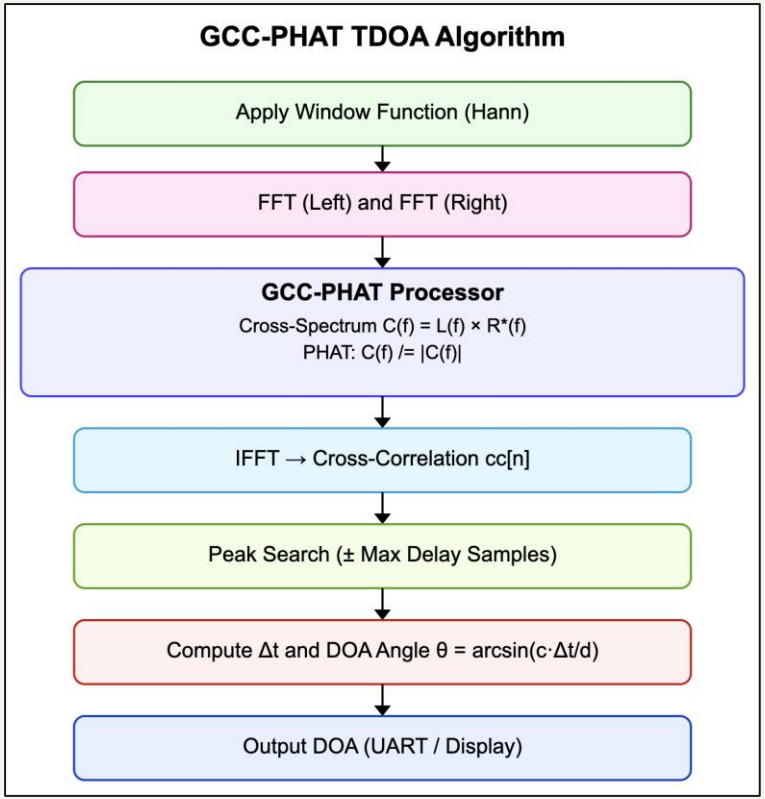
Two I2S microphones

- 48KHz stereo PCM data
- 256 frames per DoA detection





DOA
(Direction of Arrival) Algorithm



Algorithm Explained

Time Difference of Arrival (TDOA)

Hann Window:

- Smooths boundaries of window
- Reduce spectral leakage

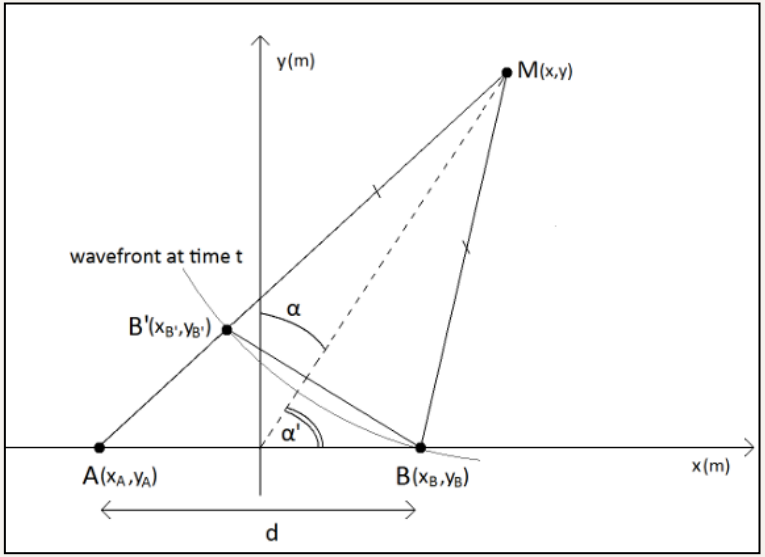
Uses GCC-PHAT, a cross-correlation technique:

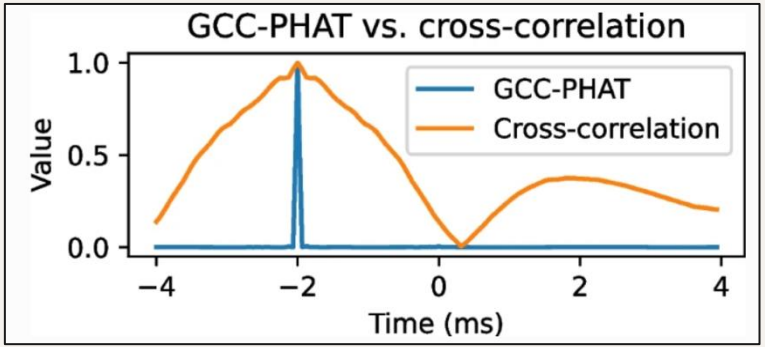
- Reduces background noise
- Good for speech detection

FFT and IFFT:

- Frequency domain for GCC-PHAT
- Time domain for angle calculation

Angle calculation: $\theta=\arcsin(c\Delta t/d)$



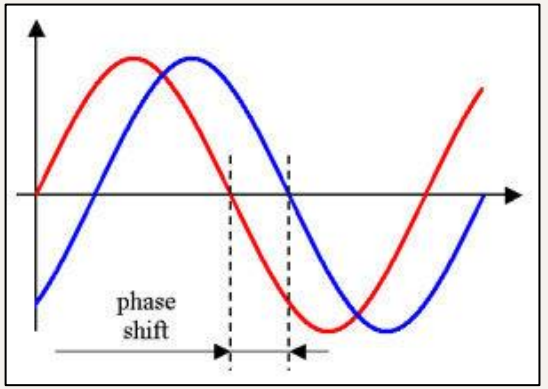


DoA Testing Phase

Artificial Data Test

Setup:

- Create snippet of artificially created, phase shifted audio data
- 2 channels of data
- DOA algorithm in C



Procedure:

- User can change the target angle
- Use equation given angle to write phase shift
- Run DOA to find angle approximation

Result:

- ~8 Degrees of error at -90 and 90

```
Enter test angle (degrees) for phase shift: 30
DOA: 28.45 deg
Computation Time: 0.001220 seconds
```

DoA Phase 2

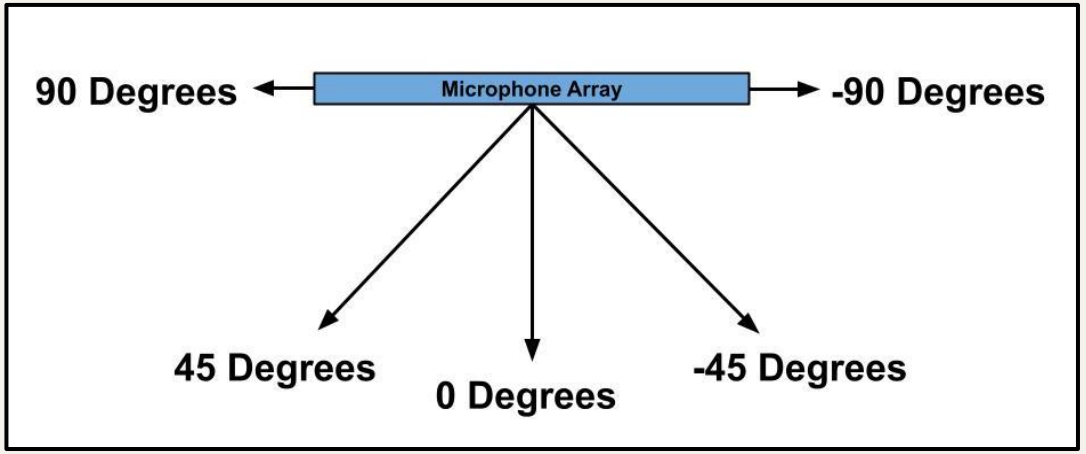
180 Degree Test

Setup:

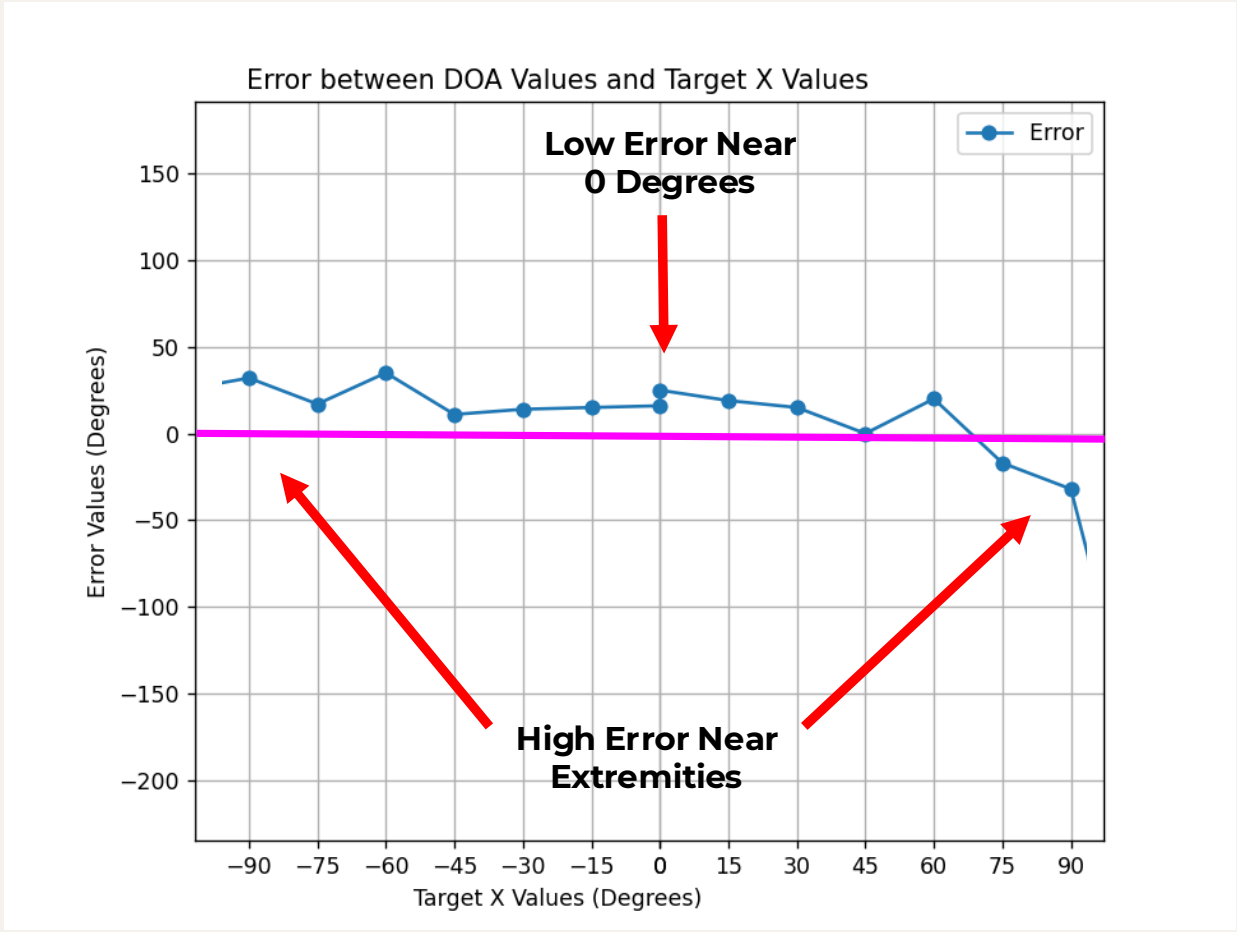
- Laptop with 2 microphones
- Protractor to measure angles

Procedure per angle:

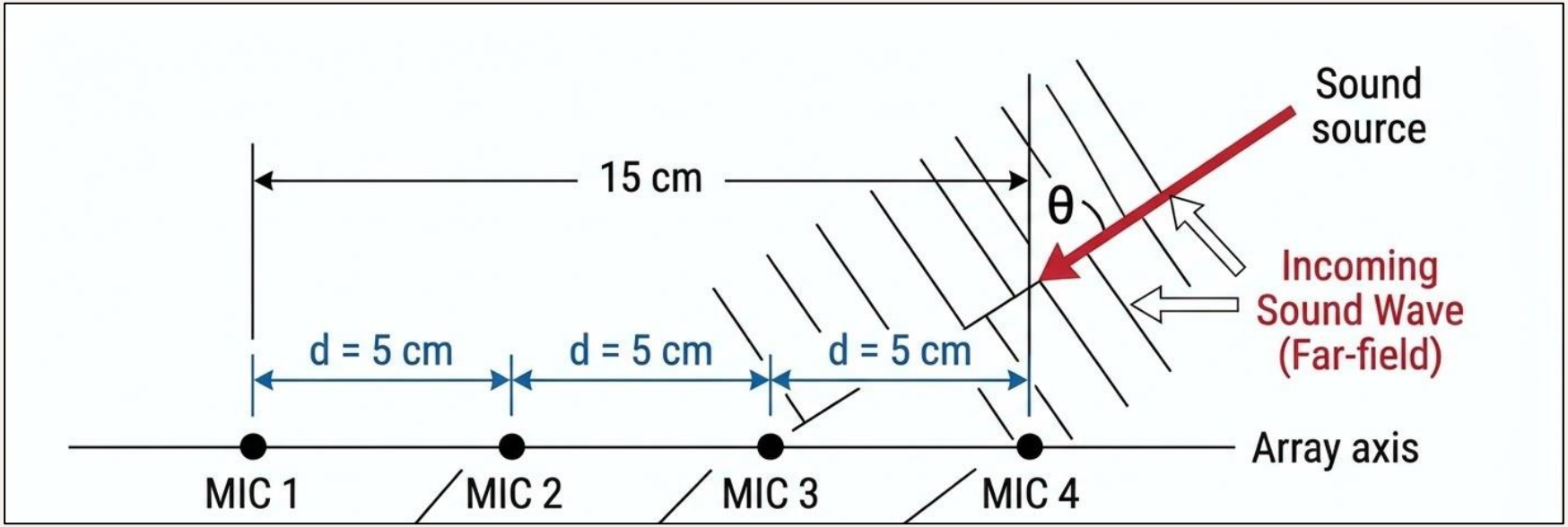
- Record microphone input
- Split left and right channels
- Calculate DOA



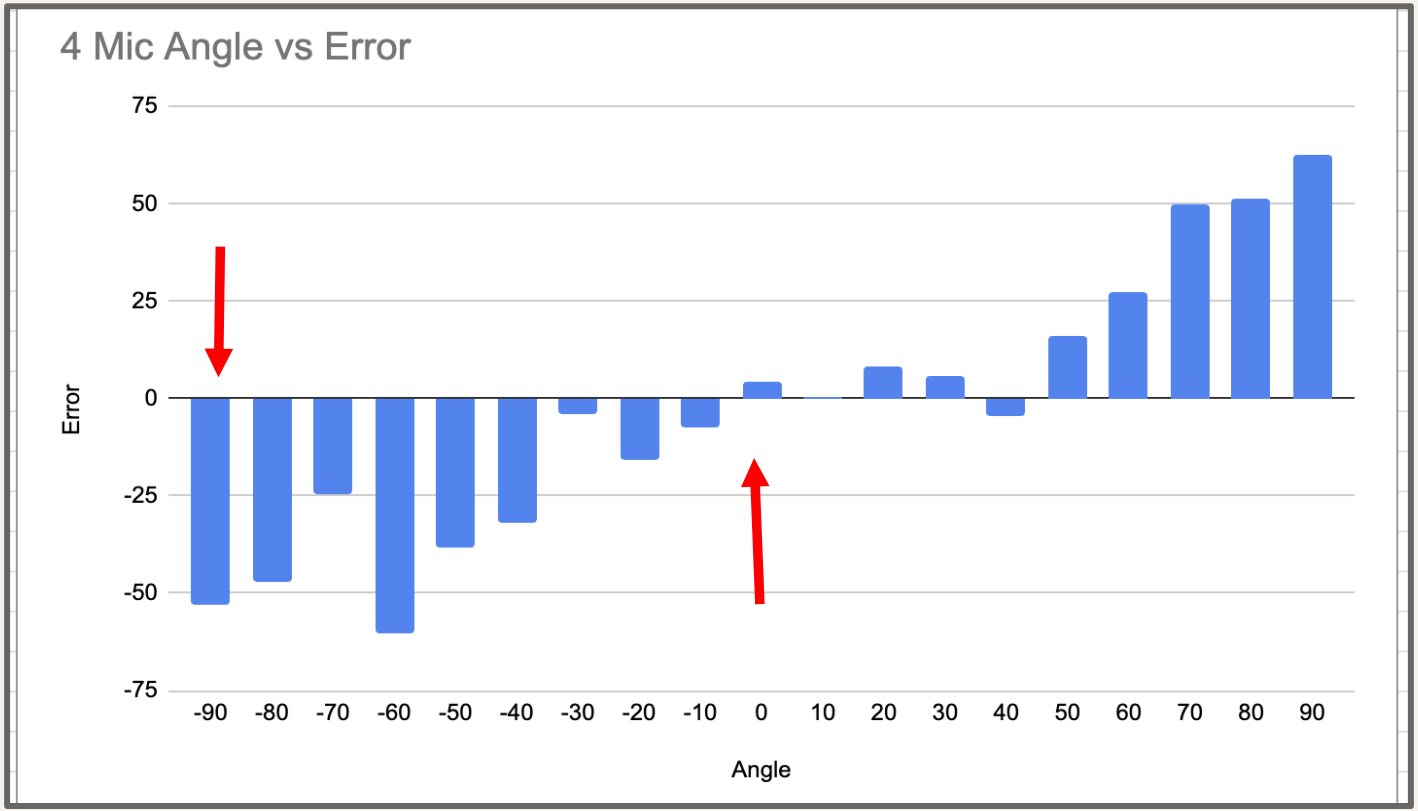
Results


Error between DOA Values and Target X Values
Error
Low Error Near
0 Degrees
High Error Near
Extremities
Error Values (Degrees)
Target X Values (Degrees)
150
100
50
0
−50
−100
−150
−200
−90
−75
−60
−45
−30
−15
0
15
30
45
60
75
90

4 Microphone Array



4 Microphone Array Comparison



Final Setup


M55M1 Dev Board
Camera
AI Model
Inference
Motor
Microphones

Demo

04

Conclusion

Project Deliverables

Our current Microphone and Camera system offers these unique solutions to the problem:

180 Degree Range

Camera can detect all speakers facing it

Lightweight

DoA and ML inference are able to run on single MCU development board

Resistance to Noise

Both ML model and DoA are configured to handle robust scenarios

Thank You For Listening

Special thanks to Professor Isukapalli, Max Dokukin, Kriteen Shrestha, Raman Sargis, and Aaron Zhou