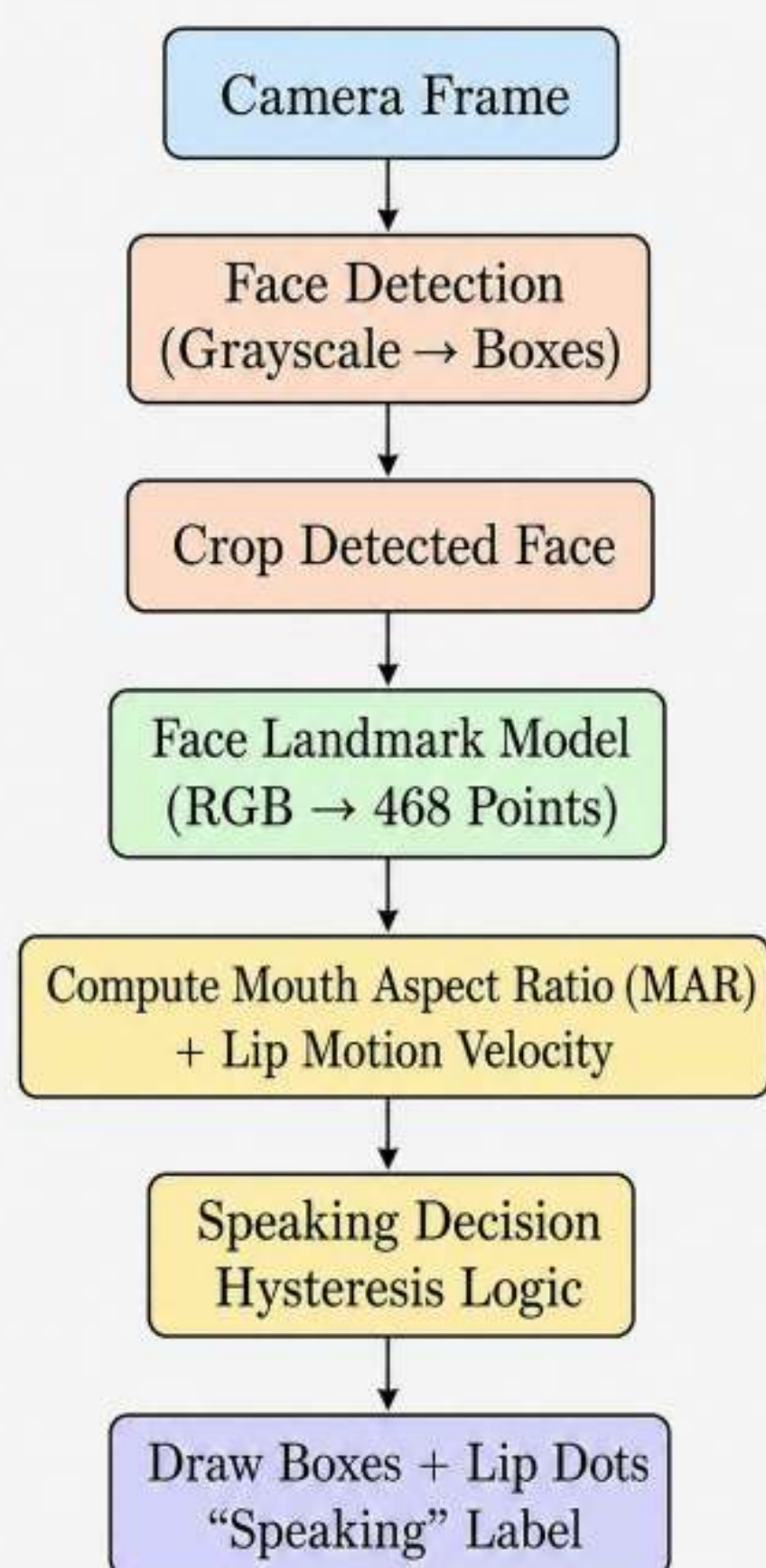
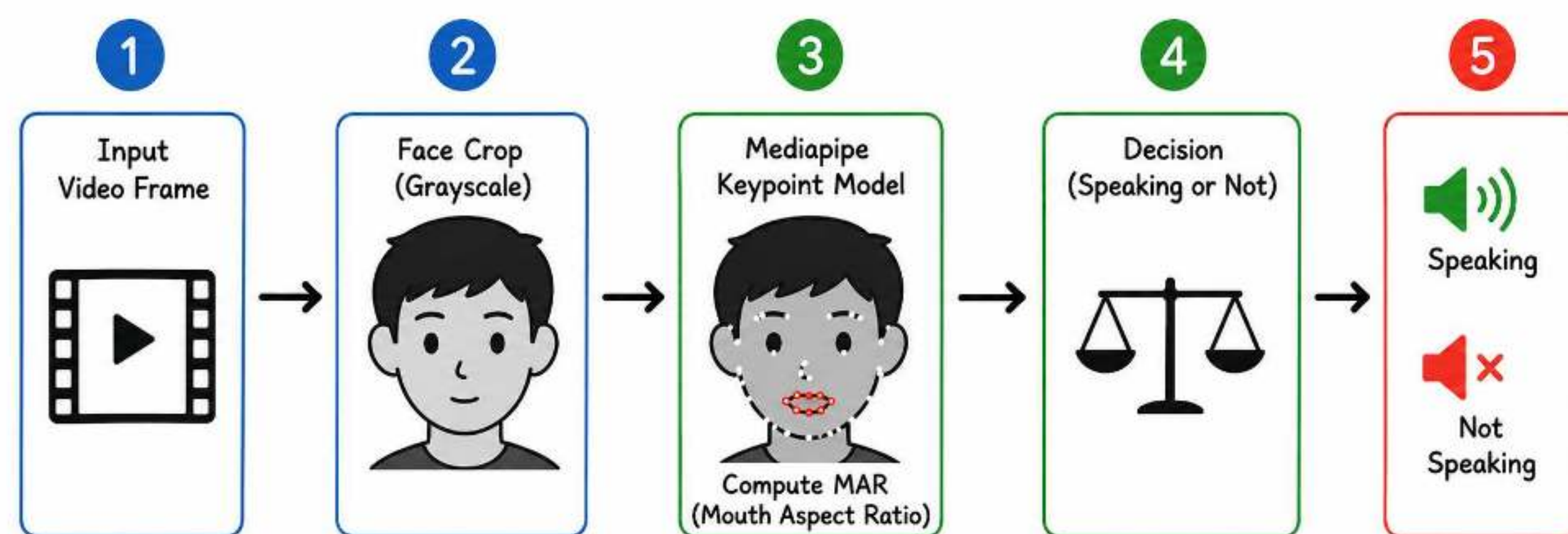


## Background

EchoTrack is a lightweight, real-time, embedded system that uses Direction-of-Arrival to rotate a camera toward the active speaker and a Edge AI Model to identify them on-screen, all running on a Nuvoton Development Board.

## ML Overview

### System Architecture



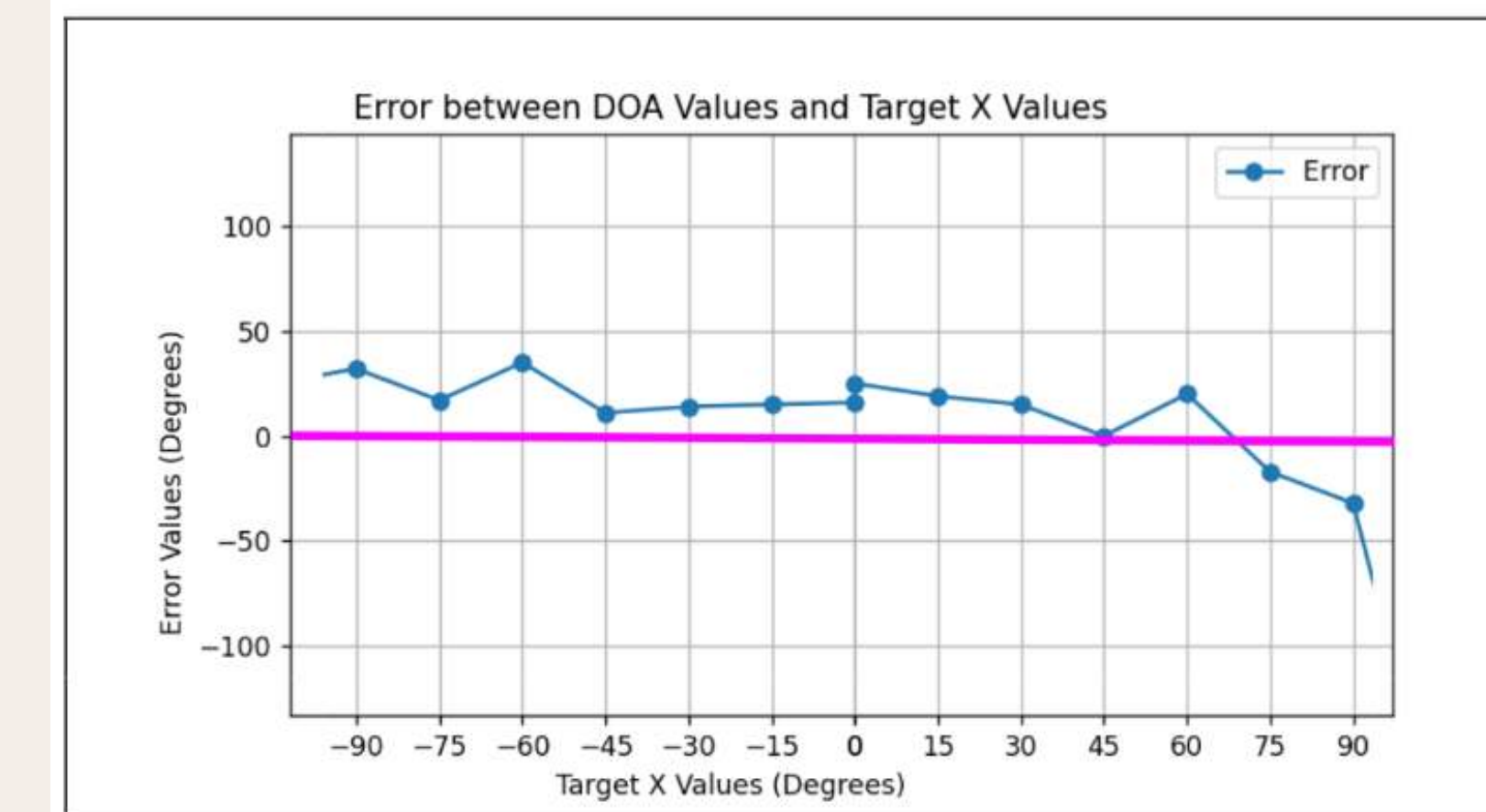
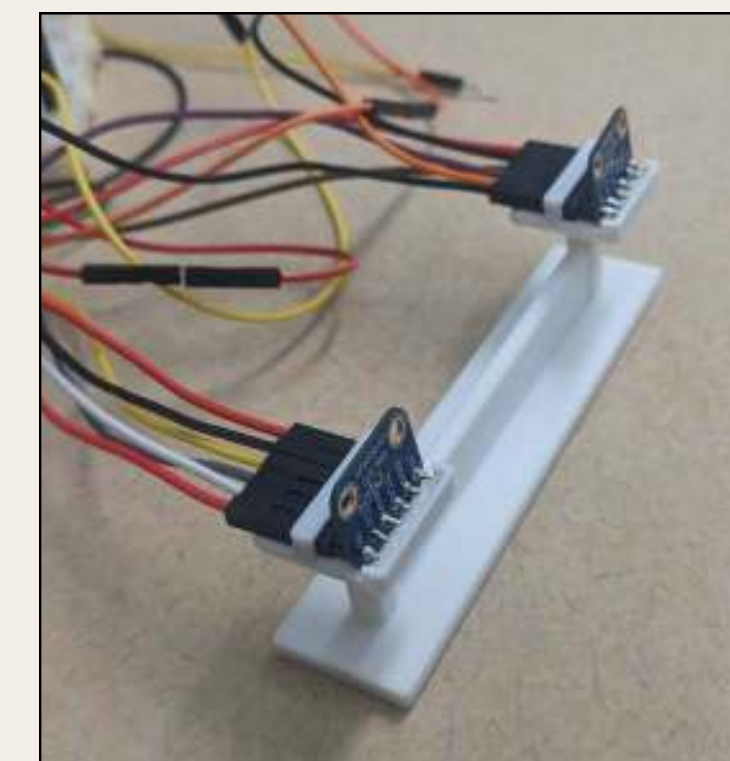
### EchoTrack's ML pipeline:

- **Find Face:** The system uses a black-and-white frame to detect faces quickly.
- **Crop Face:** It expands the face box by 1.4x, keeping the mouth and chin in view.
- **Track Mouth:** A keypoint model maps facial landmarks and uses MAR, or mouth aspect ratio, to measure lip movement and detect speaking.
- **Act:** It combines mouth movement with audio direction to steer the camera toward the active speaker

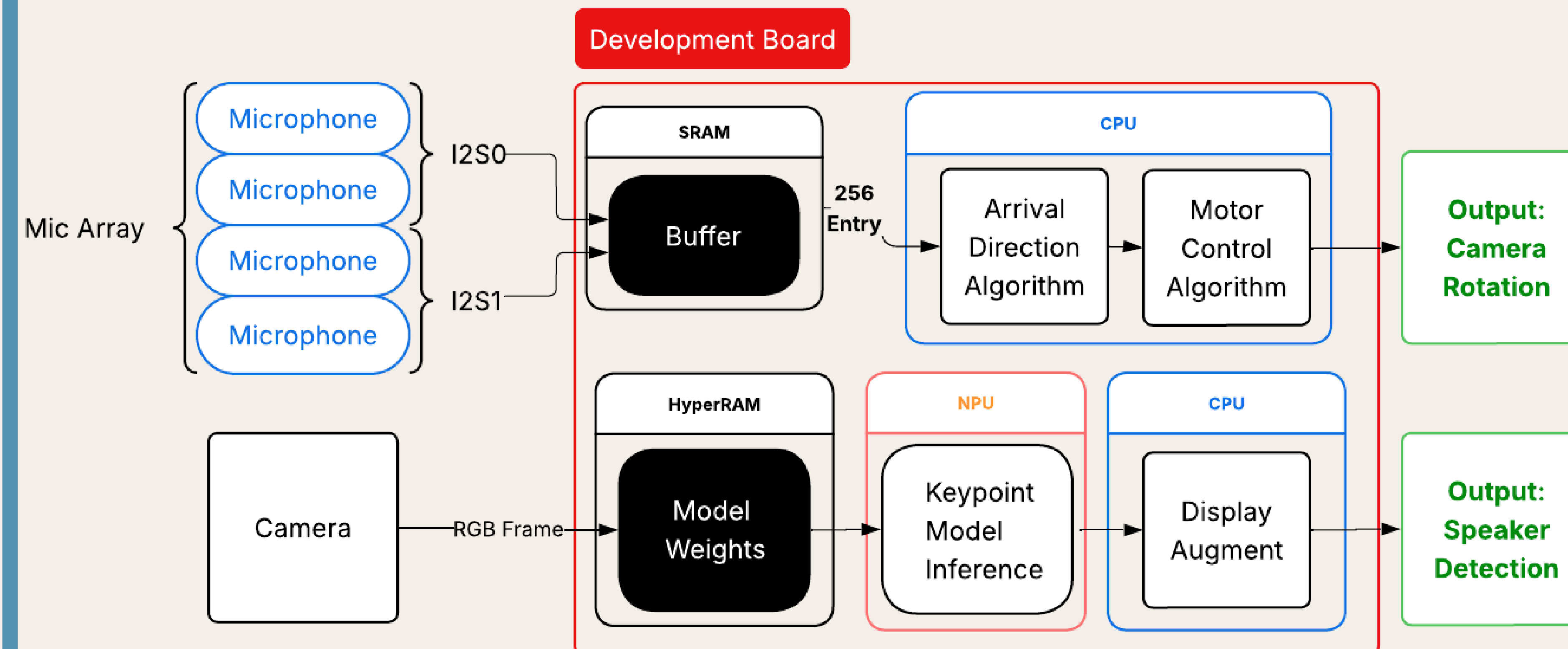
## Direction of Arrival

The project uses the TDOA (Time Difference of Arrival) algorithm using GCC-PHAT because it is computationally efficient for microcontrollers while remaining robust against noise. The setup uses a 2 microphone array and was tested by playing speaking sounds every 15 degrees until it spanned 180 degrees in front of the microphones.

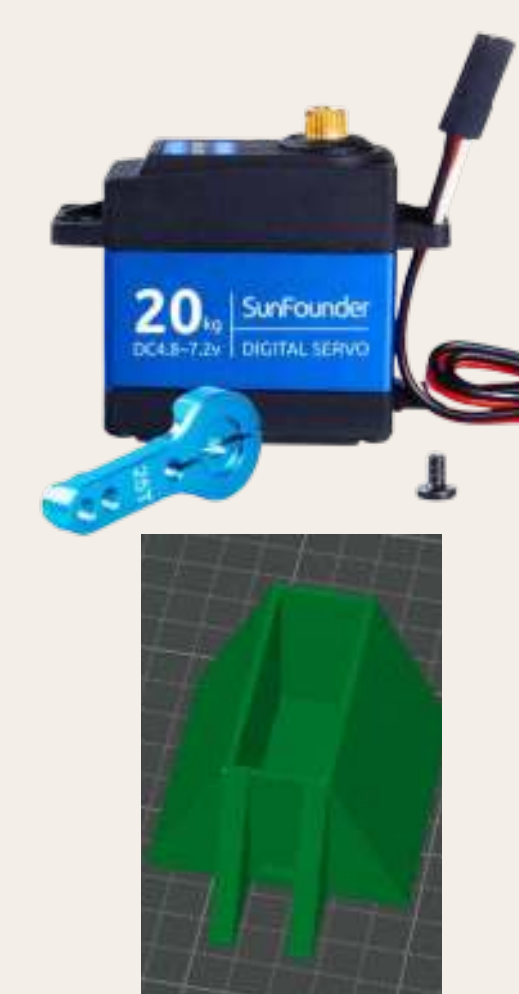
$$\theta = \arcsin\left(\frac{c\Delta t}{d}\right)$$



## System Overview

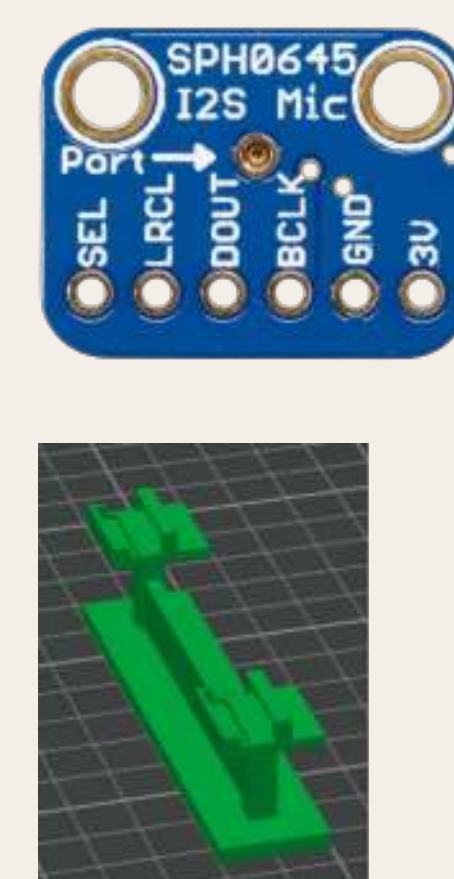


### Peripherals



#### Motor

- High torque digital servo
- EPWM driven



#### Microphones:

- 12S Connection
- Digital PWM Output



#### Camera

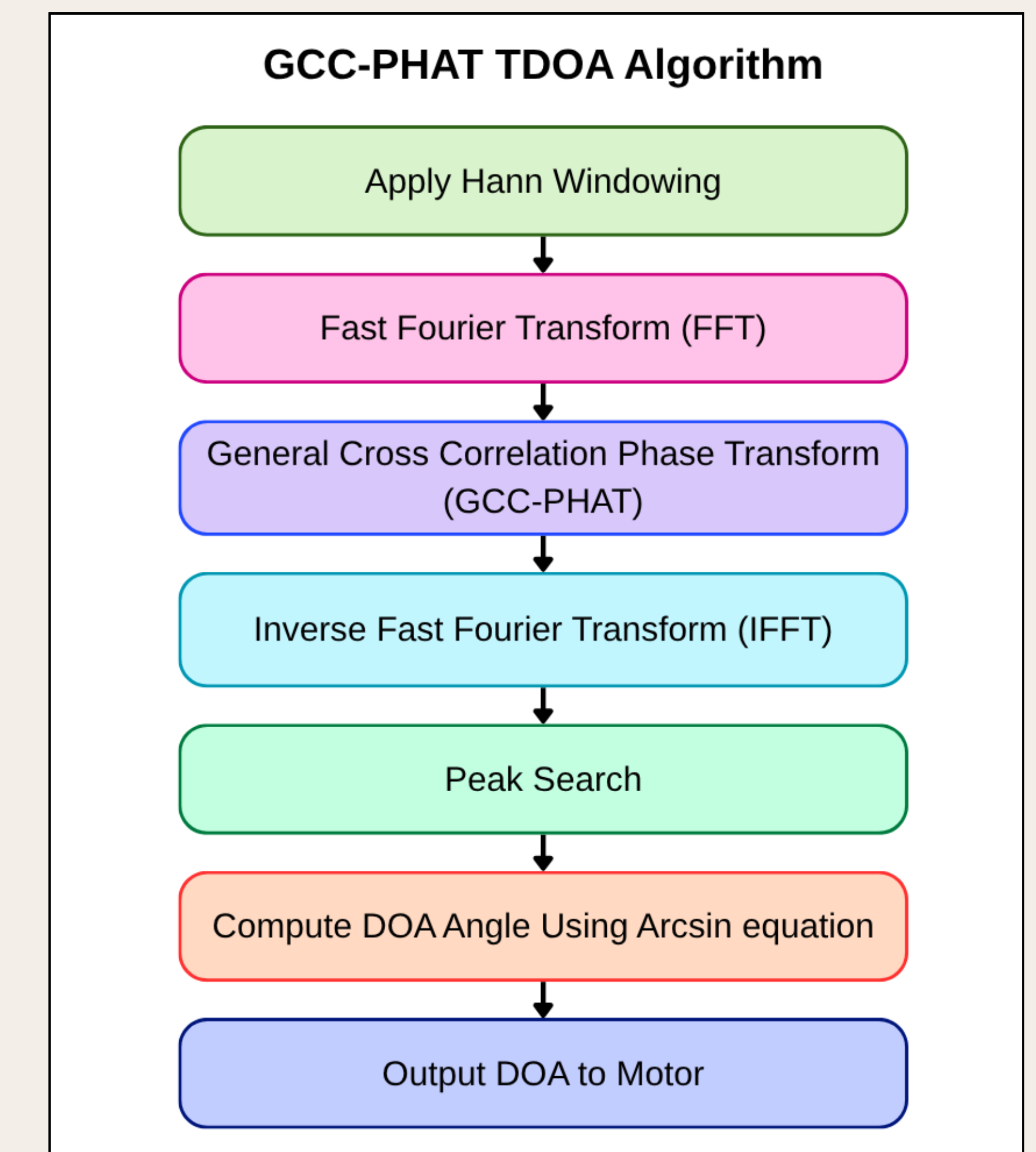
- 30 FPS
- Parallel Communication

#### NuMaker M55M1 Board

- ARM Cortex-M55 MCU
- ARM Ethos-U55 Micro NPU
- 1.5 MB SRAM and 64 MB HyperRAM



## GCC-Phat TDOA Algorithm



## Full Setup

