iContact

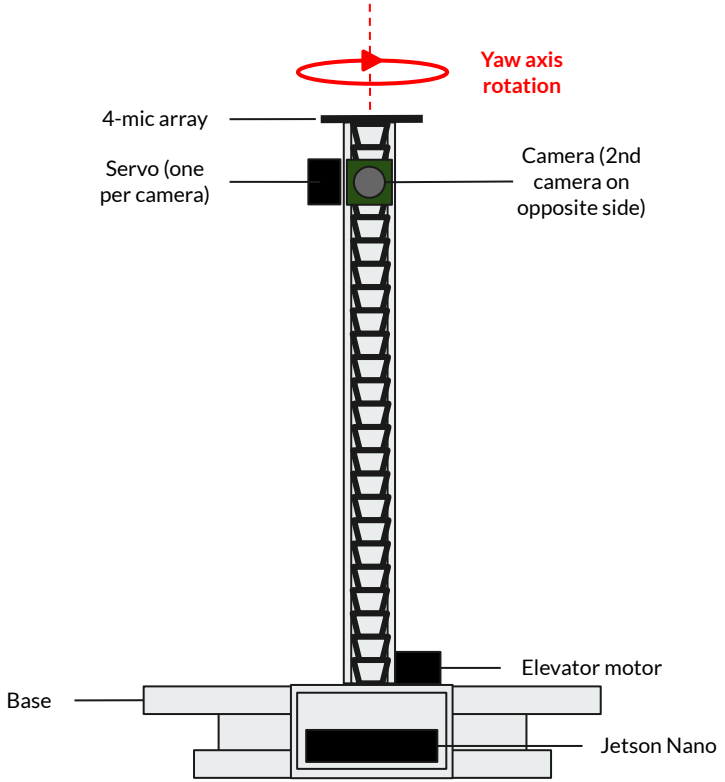
Team A3: Edward Lucero, Heather Baker, Anna Li

Application Area

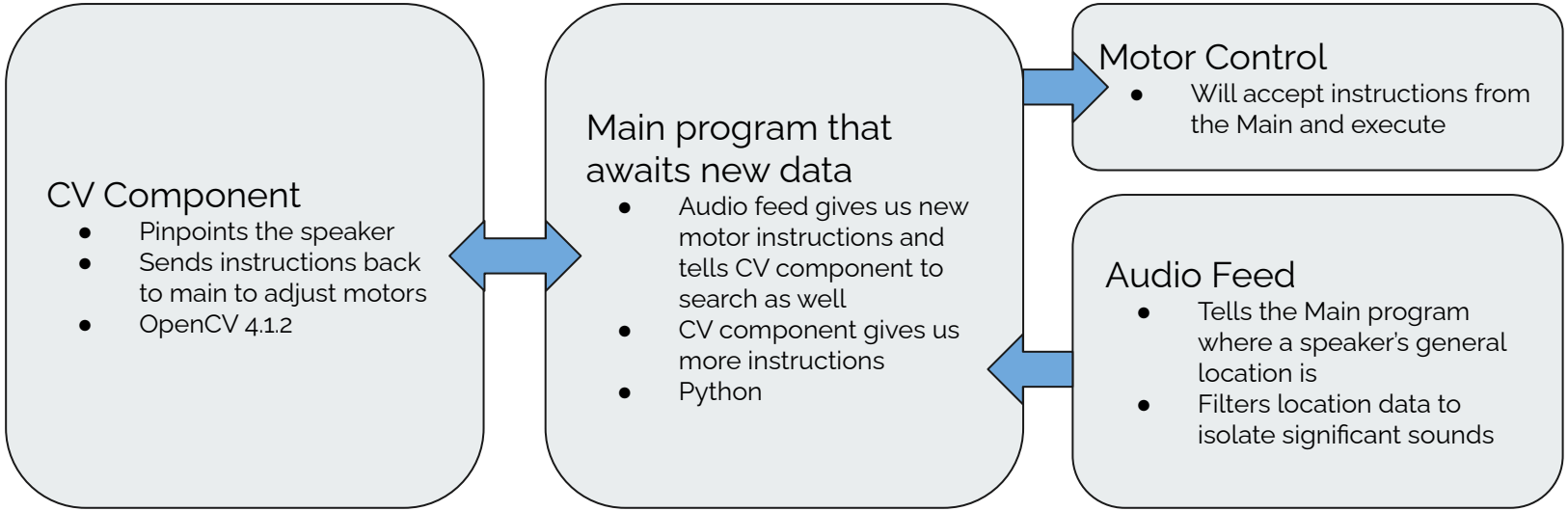
- Video calls have become an indispensable part of our daily lives
  - Classes
  - Virtual hangouts with friends
  - Work/internships
- Even before COVID, video calls were becoming essential
  - Conference calls in the workplace
  - Keeping in touch with friends and family
- Video calls have become more crucial, but have not evolved much
- **How can we better immerse the remote viewer into a video call?**
  - **Our solution: An agile camera that keeps the focus on you**

Solution Approach

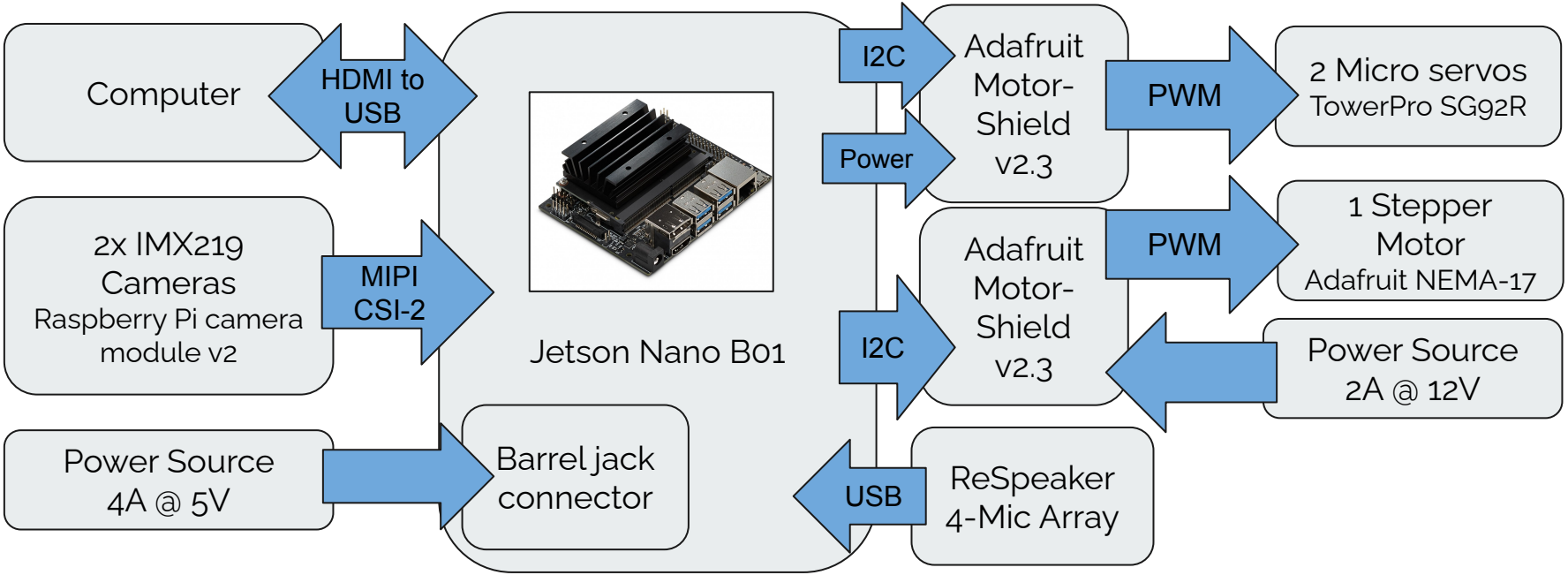
- A bi-camera mechanism on a motorized tripod
  - Can raise/lower
  - Rotates side-to-side
  - Utilizes audio detection and CV to locate and physically reposition the camera to focus on the current speaker



Software System Specification

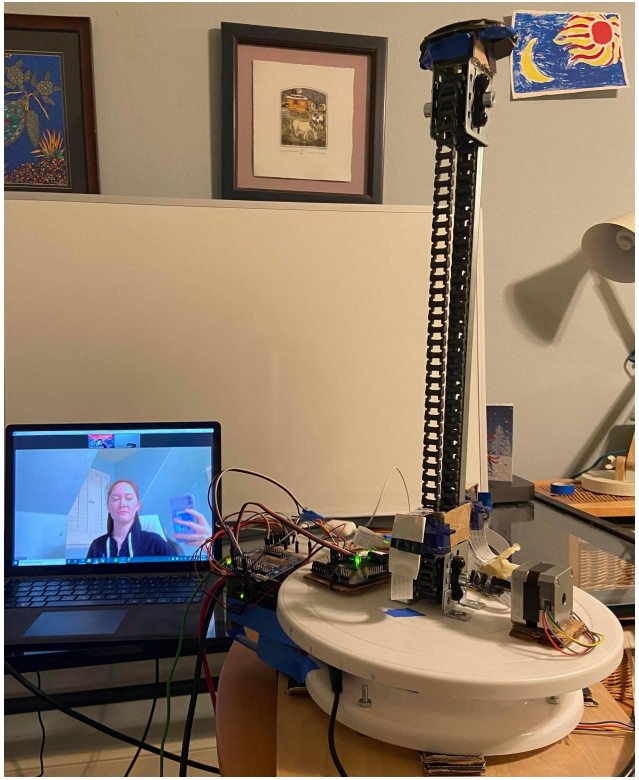


Hardware System Specification



Complete Solution

- We will first show the iContact's camera feed with:
  - A singular speaker starting in frame, then moving side-to-side, standing up and sitting down
  - Two speakers in the room sitting at different positions relative to the device, exchanging conversation
  - Two speakers in the room standing at different positions relative to the device and moving occasionally, exchanging conversation
- Then we will show the iContact itself with a speaker moving around in front of it to show the mechanics



Complete Solution: Video Demo

https://youtu.be/WYAHrr29xNo

Metrics & Validation

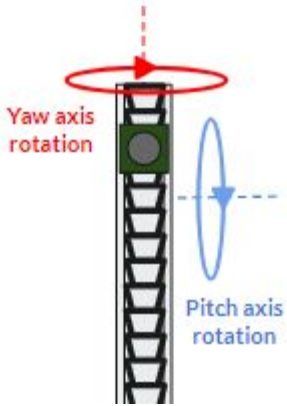
Functionality	Requirements	Testing
Viewing	Compatible with any conferencing software 1080p @30fps	Run with Zoom, Webex, and Google Hangouts
Working range	360-degree field of view 1ft vertical panning range 10ft acoustic location range 10ft person detection radius	Stationary or moving speaker around the room at various distances and angles from iContact, speaking between 50-65dB
Algorithm accuracy	90% centering accuracy 90% speaker identification accuracy	Stationary speakers converse back and forth (identification accuracy) Subject moving while continuing to talk (centering accuracy)
Speed	<1s motor control for camera adjustment <1s audio input processing latency <1s video input processing latency	Stationary speakers conversing back and forth, taking turns speaking one sentence at a time

Results

Functionality	Requirements	Results	
Viewing	Compatible with any conferencing software 1080p @30fps	✓ ✗	Worked with Zoom, WebEx, Google Hangouts 1080p @21fps
Working range	360-degree field of view 1ft vertical panning range 10ft acoustic location range 10ft person detection radius	✓ ✓ ✓ ✓	Can turn to any direction Full range of motion up/down 1ft shaft 100% accuracy up to 4ft; 90% up to 6ft 100% accuracy up to 10ft
Algorithm accuracy	90% centering accuracy 90% speaker identification accuracy	✗ ✓	Horizontal: 70-90%; vertical: 90% 100% with 2 speakers; 90% with 3 speakers
Speed	<1s motor control for camera adjustment <1s audio input processing latency <1s video input processing latency	✓ ✓ ✓	0.3957s 0.6451s 0.2749s

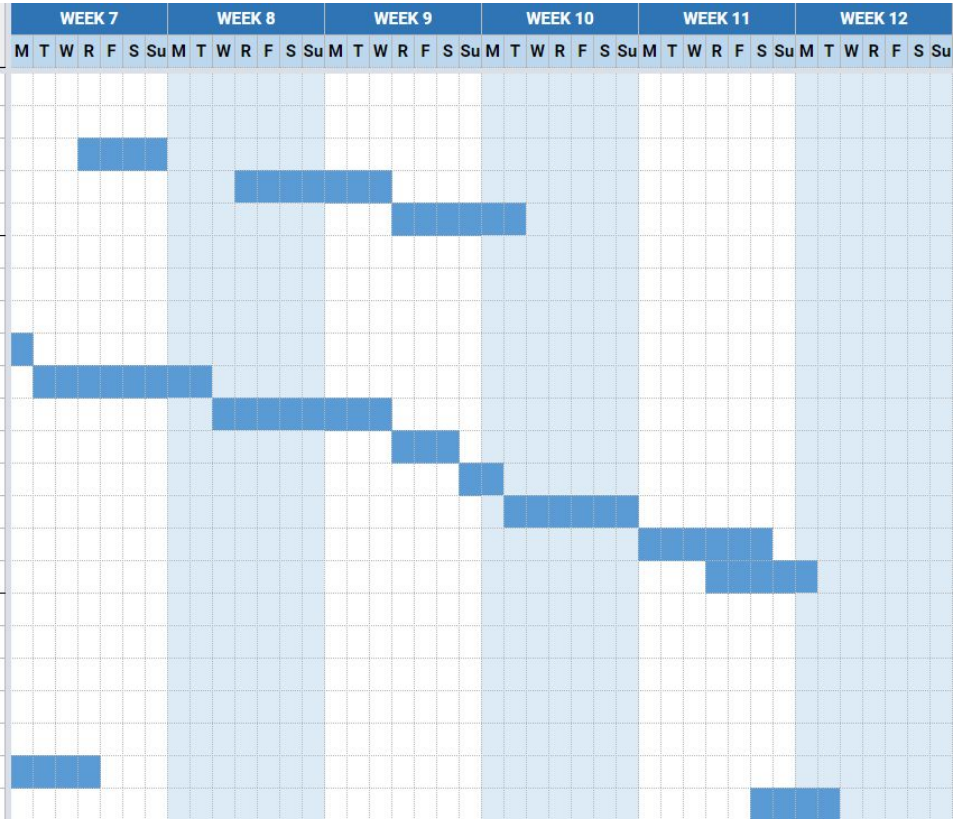
Trade Offs

- Increasing wait time after movement to gain more accurate audio readings and facial recognition
- Cropping frames at the bounds to simplify cable management
- Ignoring repeated audio measurements to reduce camera jittering
- Minimizing OpenCV processing to maintain higher framerates
- Removed pitch axis rotation because Servos were repurposed for yaw rotation



Schedule

Phase	Task	Assigned to
Physical Product	Test timing for rotations	Heather
	Design model for product	All
	Assemble first draft of physical model	All
	Make alterations to original design after tests	All
	SLACK	All
Audio	Verify stable multi mic connection	Anna
	Research algo to detect speaker's general location	Anna
	Build microphone array	Anna
	Write/test acoustic location algorithm for 2-mic array	Anna
	Assemble/test 4-mic array	Anna
	Assemble/test I2C 3-mic array	Anna
	Optimize sampling	Anna
	Research high-sampling mics	Anna
	Assemble/test USB 2-mic array	Anna
	Assemble/test USB 3-mic array	Anna
	Assemble/test ReSpeaker 4-mic array	
Computer Vision	Meet with CV Professor	Edward
	Basic setup and installation for Jetson	All
	Create a simple videofeed for a computer	Heather
	Verify multi camera communication	Heather
	Extract single frames	Edward
	Facial and Body detection	Edward
	Zooming in on a portion of the frame	Edward



Schedule

Phase	Task	Assigned to
Integration	Integrate audio into the passthrough	Anna
	Integrate motor movement with presets	A/H
	Integrate motor movement with CV component	E/H
	Add in the additional Camera for CV	E/H
	SLACK	All
Testing	Test Latency of the system	All
	Optimize Latency	All
	SLACK	All
Course Logistics	Project Proposal	All
	Design Presentation	All
	Demo 1	All
	Demo 2	All
	Final Presentation	All
	Final Report	All