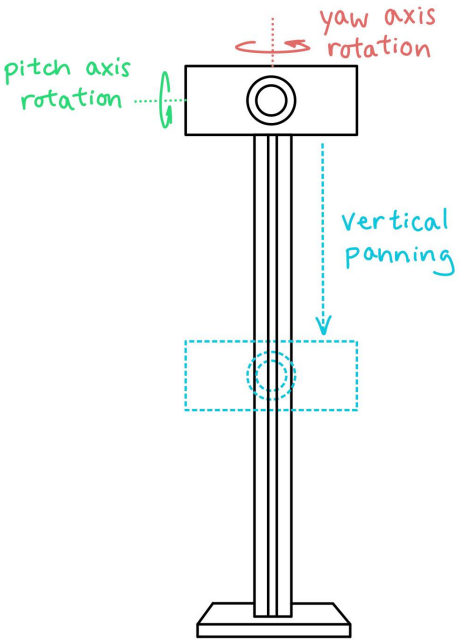
iContact

Team A3: Anna Li, Edward Lucero, Heather Baker

Project Proposal

- How can we better immerse the remote viewer into a video call?
- **The solution: An agile camera that keeps the focus on you by physically adjusting the camera**
- A bi-camera mechanism on a motorized tripod
  - Utilizes CV and audio detection to locate and physically reposition the camera to focus on the current speaker
  - Can listen for commands to preset/remember a camera angle that can be invoked at a later time during a meeting



Existing Solutions & Use Cases

- What makes it unique
  - Rotates on pitch and yaw axes to capture the best angles
    - Spring 2020 Capstone: COMOVO
  - Can raise/lower vertically to adjust to the speaker's height
  - Cheaper than existing products with 360° view
    - Meeting Owl ~$1,000, Polycom ~$5,000
- Use cases
  - Video calls (individual or conference)
  - Remote classes
    - Education, fitness, cooking, etc.

Metrics & Requirements

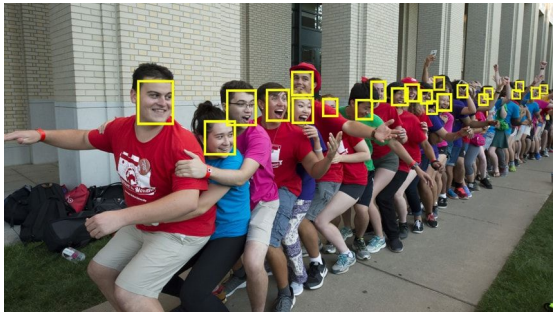
Functionality	Requirements
Viewing	Compatible with any conferencing software 1080p @30fps
Working range	360-degree field of view 3ft vertical panning range 10ft microphone pickup range 10ft person detection radius
Algorithm accuracy	90% centering accuracy (distance between center of head and frame) 90% speaker identification accuracy (based on speaker centered in frame) 90% verbal command comprehension 95% preset position alignment (how close to the preset view the motors can return)
Speed	<1s motor control for camera adjustment <1s audio input processing latency <1s video input processing latency

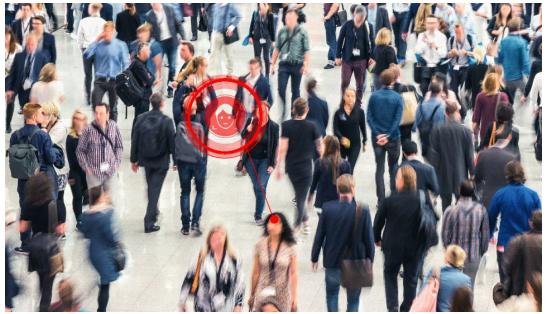
Testing

Functionality	Tests
Viewing	Run with Zoom, Webex, and Google Hangouts
Working range	Stationary or moving speaker around the room at various distances and angles from iContact
Algorithm accuracy	Stationary speakers converse back and forth (identification accuracy) Subject moving while continuing to talk (centering accuracy) Subject presets camera position and invokes using verbal commands
Speed	Stationary speakers rapidly conversing back and forth

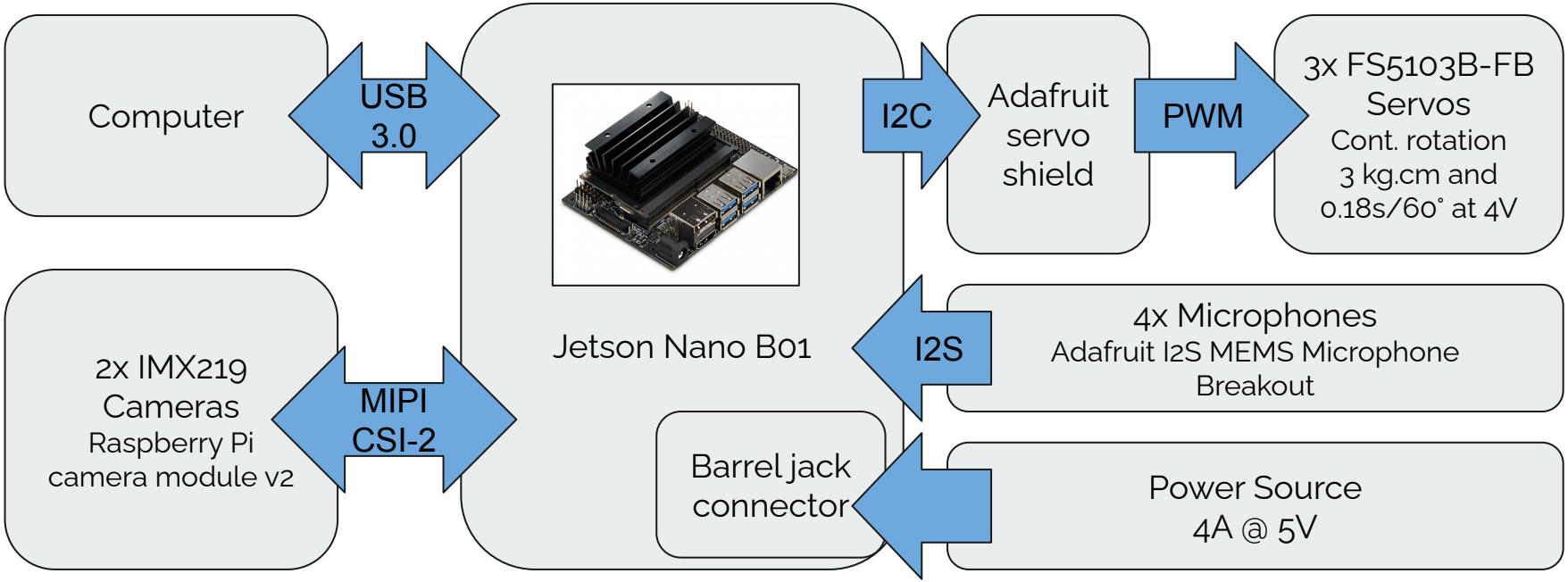
Challenges

- Processing audio/visual input and moving cameras quickly to keep up with conversation
- Moving the cameras smoothly and quietly
- Pinpointing a speaker when there are multiple sound sources
- Centering on speaker with stationary/moving objects in frame
- Where to point the cameras when there are no heads in view
    - Will need to be able to identify torsos (→ pan up) and use audio detection to locate a talking head





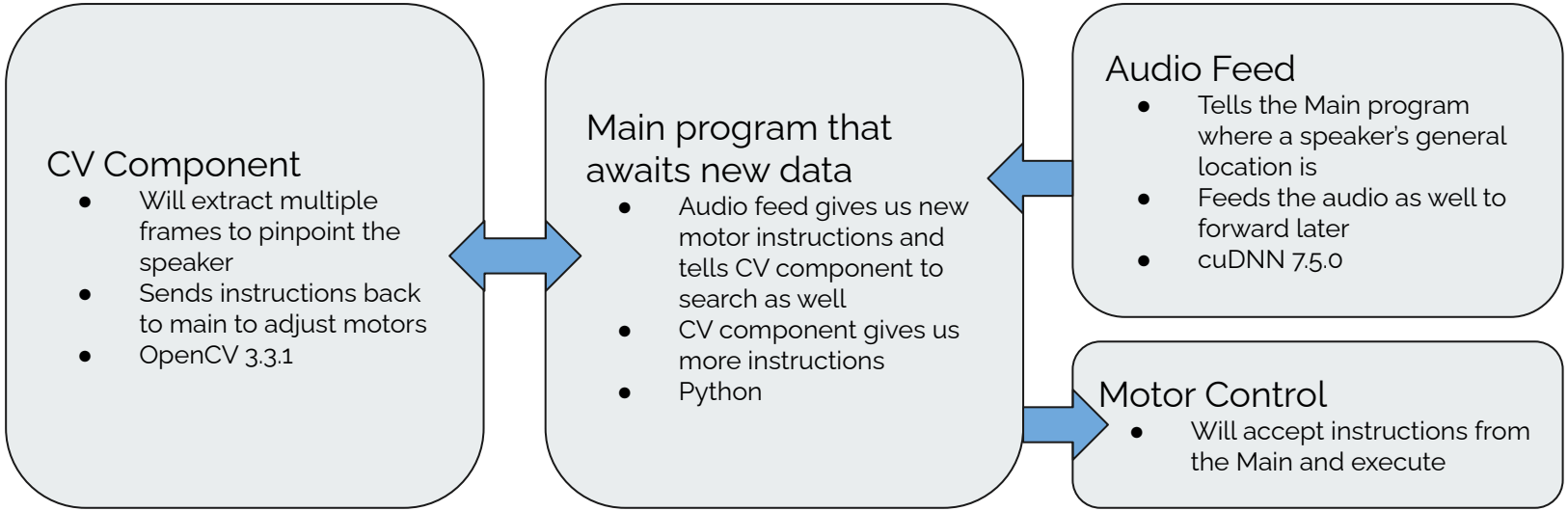
Hardware Design



Spec Comparisons

	Jetson Nano B01 with 4 GB RAM	Raspberry Pi Model 4 with 4 GB RAM
Video	2 MIPI CSI-2 DPHY lanes	1 MIPI CSI-2 DPHY lanes
USB	4 USB 3.0	2 USB 3.0, 2 USB 2.0
GPIO	40 pin GPIO	40 pin GPIO
Video Decoder	H.264 up to 1080p240	H.264 up to 1080p60
Audio	2xI2S, (Can wire as stereo for 4 mics)	USB or I2C with ADCs
CPU	CPU ARM A57	CPU ARM A72 (1 generation newer)
Motor	4xI2C, 1xPWM	6xI2C, 2xPWM
Total cost	$224.70	$220.65

Software Design

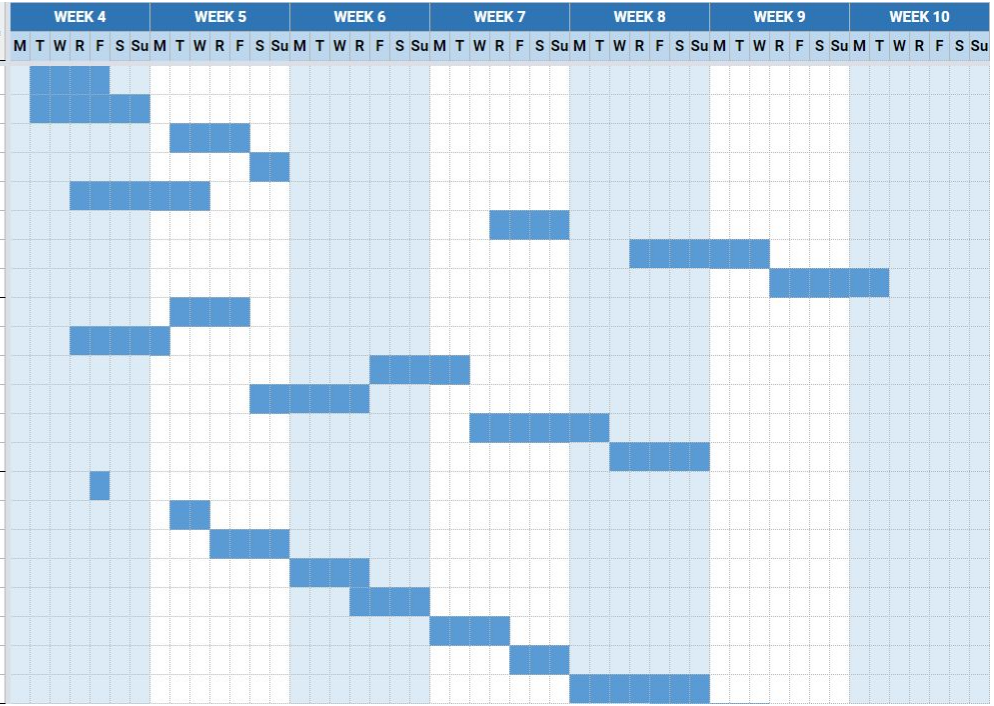


Division of Labor

Area	Task	Anna	Edward	Heather
Signals	Audio processing	✔		
Software	Video processing (CV)		✔	✔
Hardware	Motor control			✔
	Camera input to Jetson		✔	✔
	Microphone input to Jetson	✔		✔
	Transmitting audio feed to computer	✔		
	Transmitting video feed to computer		✔	

Schedule

Phase	Task	Assigned to	Start Date	End Date	Number of Days	Completion
Physical Product	Purchase Hardware Components	All	9/22/20	9/25/20	4	0%
	Review Datasheets for Components	All	9/22/20	9/27/20	6	0%
	Verify multiple motor control	Heather	9/29/20	10/2/20	4	0%
	Test timing for rotations	Heather	10/3/20	10/4/20	2	0%
	Design model for product	All	9/24/20	9/30/20	7	0%
	Assemble first draft of physical model	All	10/15/20	10/18/20	4	0%
	Make alterations to original design after tests	All	10/22/20	10/28/20	7	0%
	SLACK	All	10/29/20	11/3/20	5	0%
Audio	Verify stable multi mic connection	Anna	9/29/20	10/2/20	4	0%
	Research algo to detect speaker's general location	Anna	9/24/20	9/28/20	5	0%
	Research algo to detect voice commands	Anna	10/9/20	10/13/20	5	0%
	Implement algo to detect speaker's general location	Anna	10/3/20	10/8/20	6	0%
	Implement algo to detect voice commands	Anna	10/14/20	10/20/20	7	0%
	SLACK	All	10/21/20	10/25/20	5	0%
Computer Vision	Meet with CV Professor	Edward	9/25/20	9/25/20	1	0%
	Basic setup and installation for Jetson	All	9/29/20	9/30/20	2	0%
	Create a simple videofeed for a computer	Heather	10/2/20	10/4/20	3	0%
	Verify multi camera communication	Heather	10/5/20	10/8/20	4	0%
	Extract single frames	Edward	10/8/20	10/11/20	4	0%
	Facial and Body detection	Edward	10/12/20	10/15/20	4	0%
	Zooming in on a portion of the frame	Edward	10/16/20	10/18/20	3	0%
	SLACK	All	10/19/20	10/25/20	7	0%



Schedule: Continued

Phase	Task	Assigned to	Start Date	End Date	Number of Days	Completion
Integration	Integrate the video passthrough	Edward	10/23/20	10/28/20	6	0%
	Integrate audio into the passthrough	Anna	10/23/20	10/28/20	6	0%
	Integrate motor movement with presets	A/H	10/29/20	11/2/20	4	0%
	Integrate motor movement with CV component	E/H	10/29/20	11/2/20	4	0%
	SLACK	All	11/3/20	11/9/20	7	0%
Testing	Test Latency of the system	All	11/12/20	11/17/20	6	0%
	Optimize Latency	All	11/17/20	11/22/20	6	0%
	SLACK	All	11/23/20	11/29/20	7	0%
Course Logistics	Project Proposal	All	9/14/20	9/21/20	8	100%
	Design Presentation	All	10/12/20	10/14/20	3	0%
	Demo 1	All	11/9/20	11/11/20	3	0%
	Demo 2	All	11/30/20	12/2/20	3	0%
	Final Presentation	All	12/3/20	12/9/20	7	0%
	Final Report	All	12/8/20	12/13/20	6	0%