Final Presentation
Gesture-Controlled Speaker

Isabel Holtan, Dale Kercorian, Renee Li,
Kevin Pere, Eli Smith, Teddy Withey

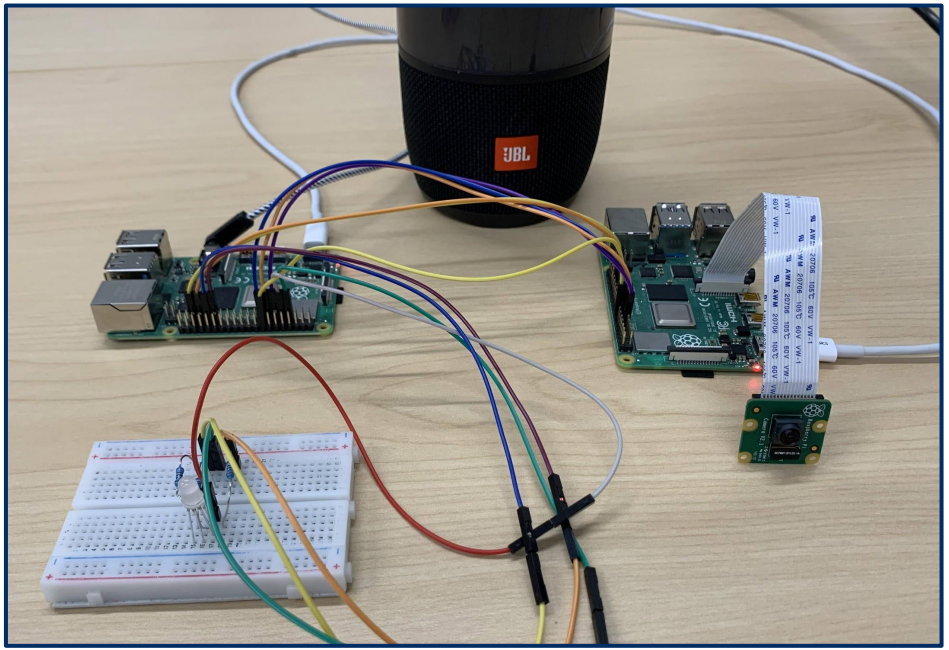
Agenda

1. Introduction
2. Motivation
3. Work Breakdown
4. System Architecture
    a. Data acquisition
    b. Data processing
    c. Decision and actuation
5. Demo
6. Design Expo
7. Future Work

Introduction

Speaker that responds to five hand gestures

- Raise & Lower Volume
- Play & Pause
- Skip Song
- Go Back a Song



Motivation

Hygiene

- No physical contact
- Possibility of social distancing

Accessibility

- Voice control can be difficult to use with an accent or speech impediment

Adaptability

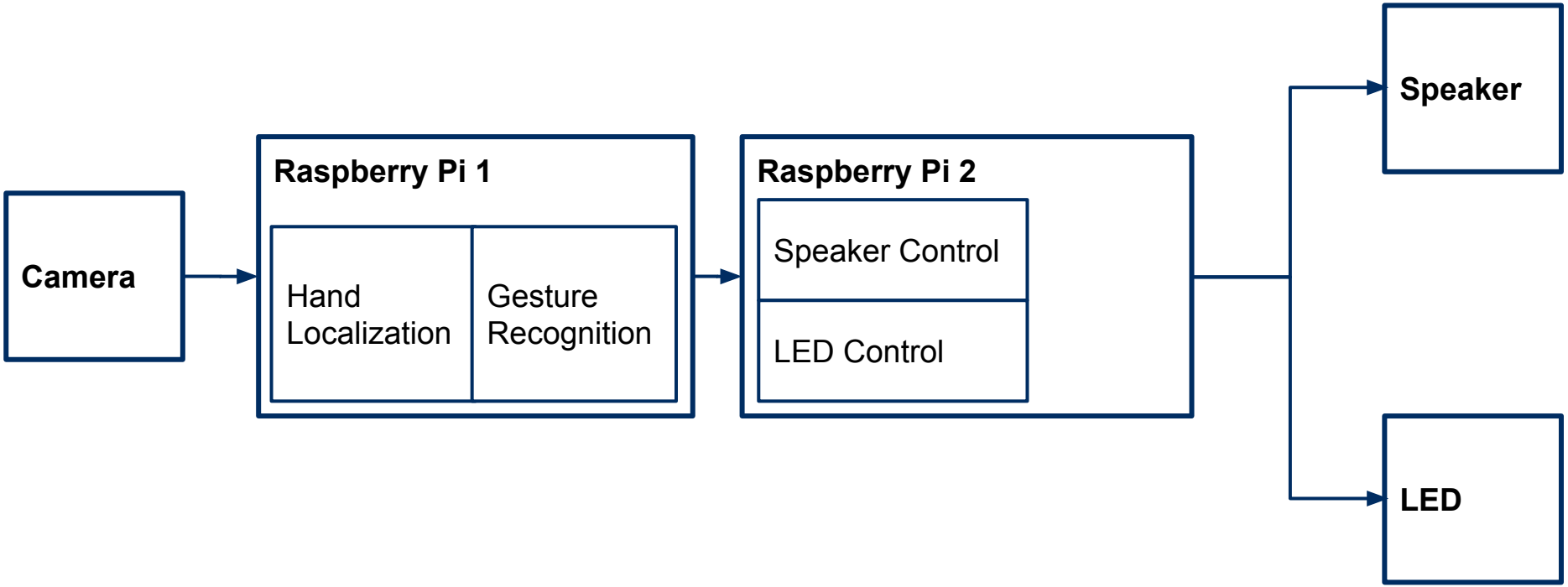
- Technology can be applied to other products

Work Breakdown

	Isabel	Dale	Renee	Kevin	Eli	Theo
Team Lead					X	
System Integration	X	X	X	X	X	X
IR/other sensor selection and integration		X		X		X
Image feature extraction	X		X		X	
Machine learning model selection and training			X		X	
Speaker integration	X	X		X		X
Lighting system integration				X		
Music selection		X		X		

System Architecture

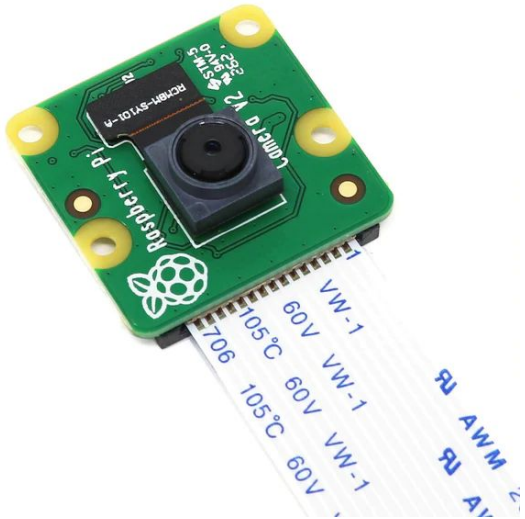
Three main phases: data acquisition, data processing, decision and actuation



Data Acquisition

Raspberry Pi Camera

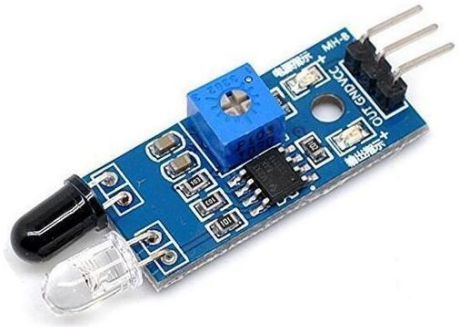
- Used to send image data to RPi to be processed
- Camera continuously captures images
  - Sent to gesture decision making RPi (RPi 1)
- Could run up to 30fps



Data Acquisition

IR Sensor

- Was going to be used to limit processing/active time of system
- Decided not to include in final design due to unreliability
  - Sensors either didn't have a sufficient range or their data transmission was inconsistent
- Inconsistencies could be due to lack of ground-ground between RPis
  - By the time this issue was discovered, we found the IR was unnecessary anyway



Data Processing - Goal

Hand Localization

- Given a camera frame, find and isolate relevant hands
- Return hand data

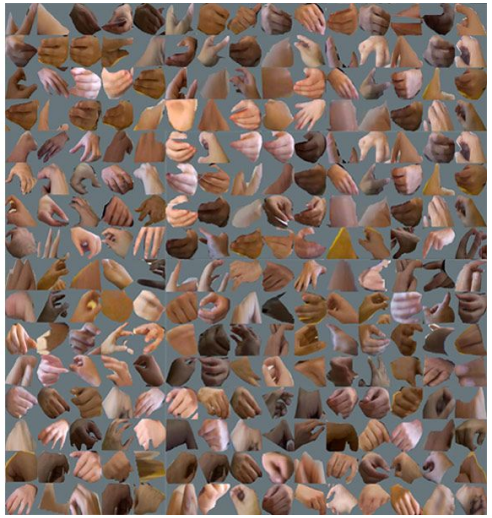
Gesture Classification

- Given hand data, determine which hand gesture is being made
- Return identified gesture

Hand Localization - Process

Tensorflow + transfer learning

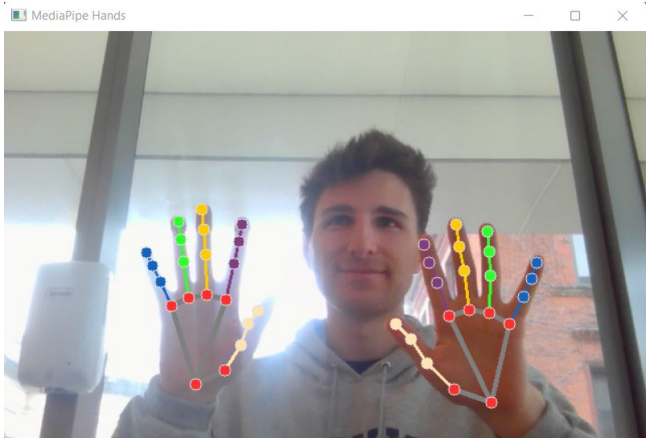
- Slow training process
- Required lots of data
- Low fps

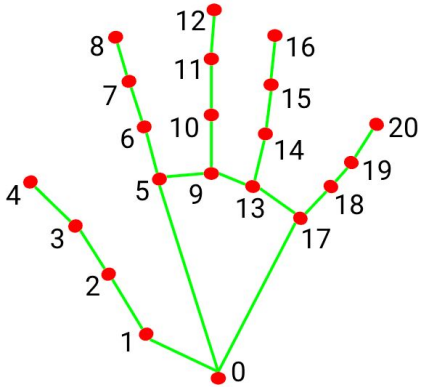


Hand Localization - Result

MediaPipe (Google’s open-source ML pipelines)

- Palm detection model operates on full image, returns a hand bounding box
- Hand landmark model operates on cropped image, returns 21 hand keypoints
- Fastest option





0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

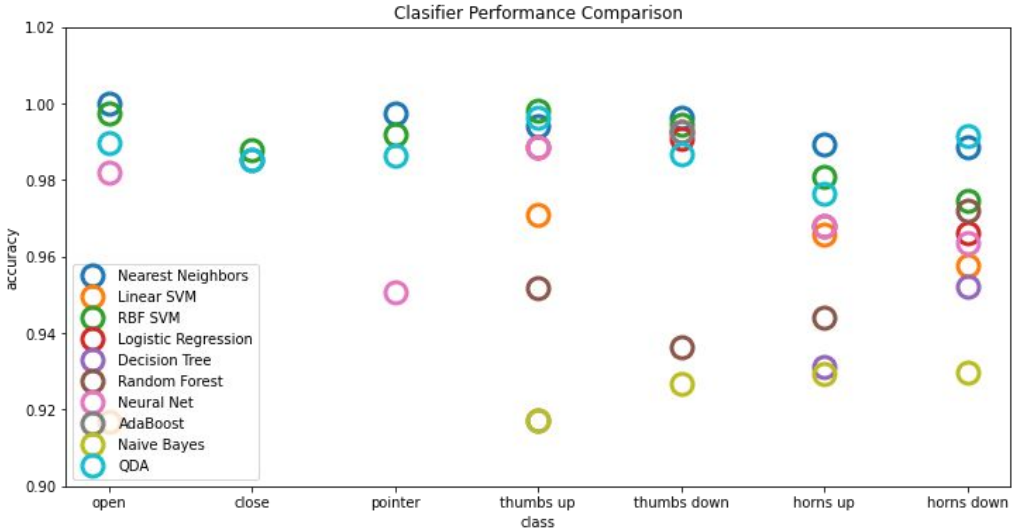
Gesture Classification - Process

Classification based on images

- CNN is slow, heavyweight
- Implementation with 10 gestures: ~50% misclassification

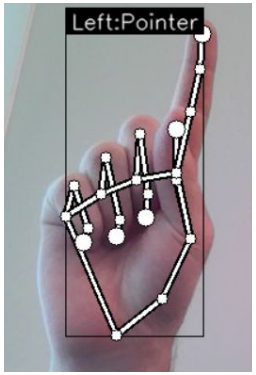
Classification based on keypoints

- Fast, accurate
- Lighting invariant

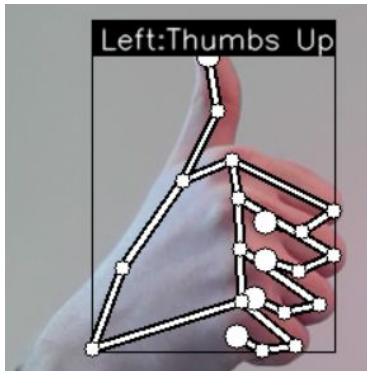


Gesture Classification - Results

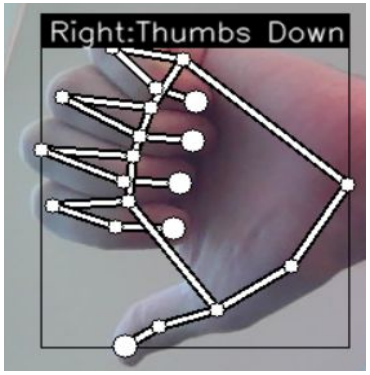
Detect and respond to five gestures with ~92% accuracy



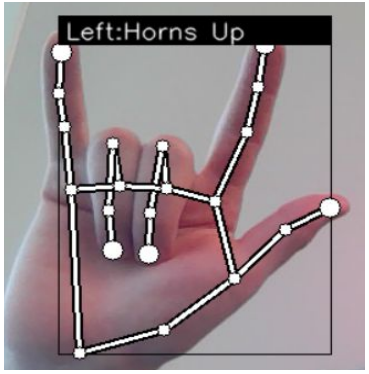
Play/Pause



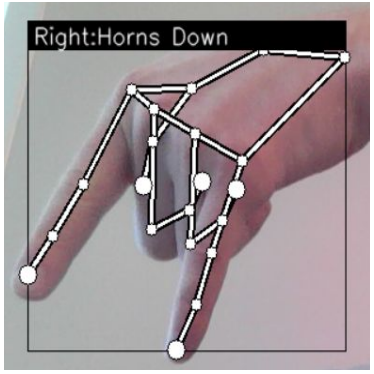
Volume Up



Volume Down



Skip Song



Go Back a Song

Decision and Actuation - Goal

Command Speaker

- Assign each gesture a number, which corresponds to a speaker function
- Gesture number passed to speaker program, perform function

LED Feedback

- Light up an LED to let user know that their gesture was detected
- Each gesture would correspond to a different color on the LED

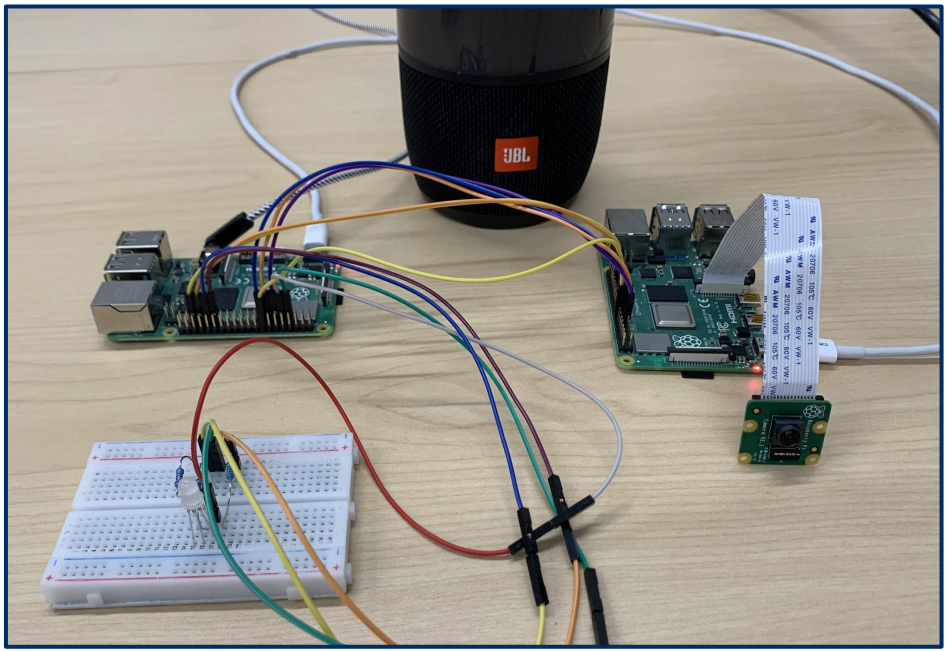
Decision and Actuation - Process

VLC Media Player with AUX speaker

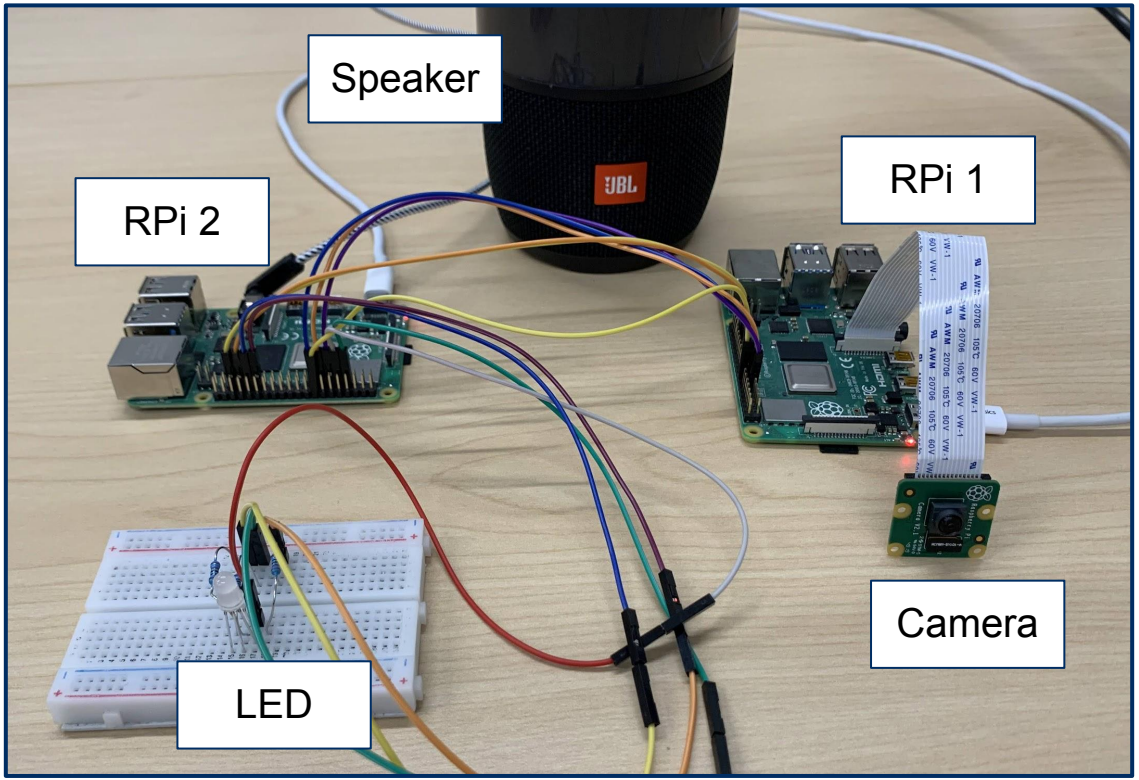
- Pause/play, set volume functions for playlist of pre-downloaded songs

Double RPi Setup

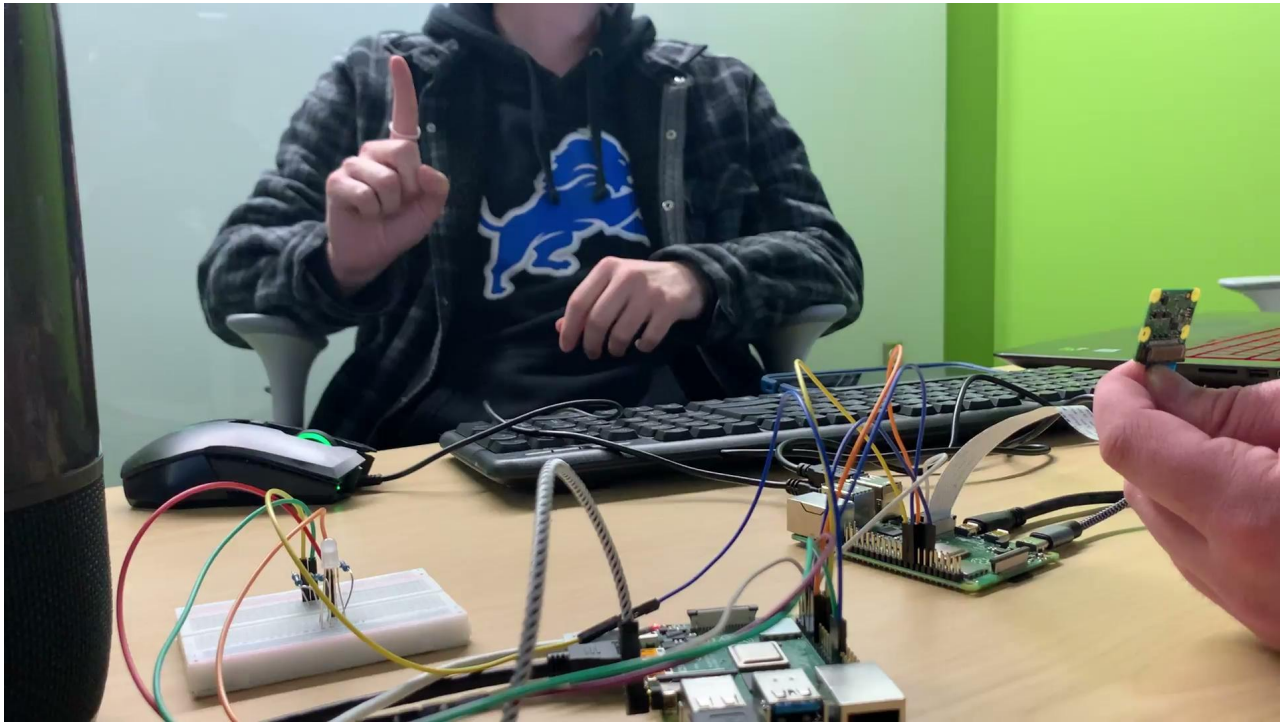
- One RPi does not have the power to perform gesture detection and actuation
- Send detected gesture to second RPi via GPIO



Decision and Actuation - Result



Demo



Design Expo

Future Work

- Second sensor that is more robust than original IR sensor
- Second LED to show user if there hand is being fully detected
- Code starts up with Raspberry Pi
- Encasing around entire system

Contact

Isabel Holtan

iholtan@umich.edu

Kevin Pere

kpere@umich.edu

Dale Kercorian

dkercori@umich.edu

Eli Smith

erssmith@umich.edu

Renee Li

reneeli@umich.edu

Teddy Withey

witheyth@umich.edu