ELECTRICAL &
COMPUTER ENGINEERING
UNIVERSITY OF MICHIGAN

Gesture-Controlled Speaker

EECS 452: Digital Signal Processing Design Lab – Winter 2022

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

Motivation

Have you ever been frustrated trying to control your Alexa when it was playing music too loudly? To tackle this problem, we designed and built a gesture-controlled speaker. This allows a user to make a hand gesture and the system will adjust the speaker settings accordingly. We are excited by the possibility of gesture control to work as a control mechanism for other smart home devices.

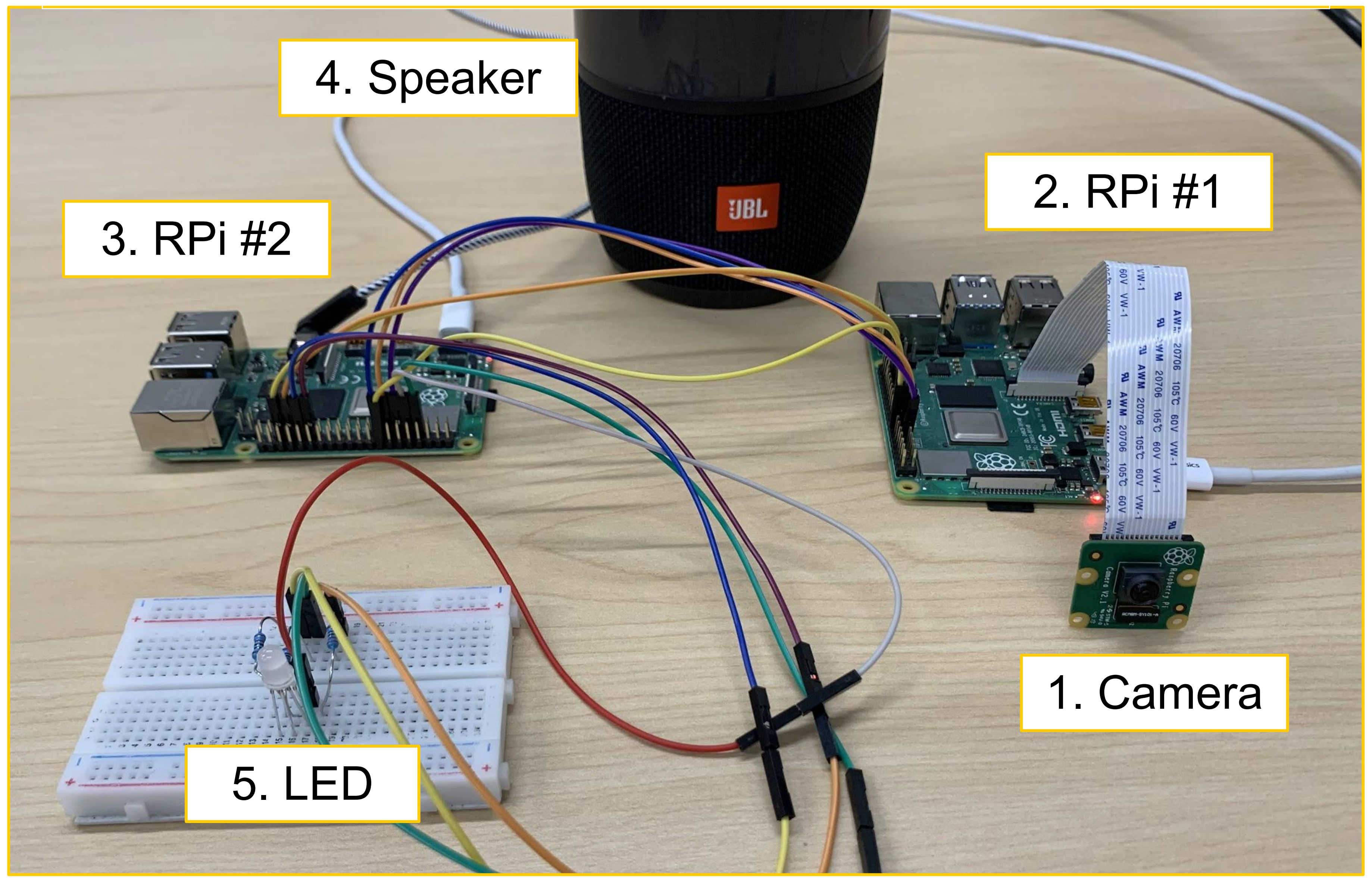
Algorithm

We use an Raspberry Pi (RPi) camera to monitor for user gestures. The following steps are taken for each video frame captured by the camera

1. Capture frame
2. Run MediaPipe hand localizer on captured image
3. Collect 20 hand feature coordinates
4. Run features through logistic regression gesture classifier
5. Return best gesture guess
6. Execute designated action on speaker and LED

System Architecture

The numbered diagram below illustrates the flow of data through the system. We use two RPi's, one in charge of image acquisition and processing and the other in charge of speaker and LED control. This helps increase the rate at which the system can process new images.

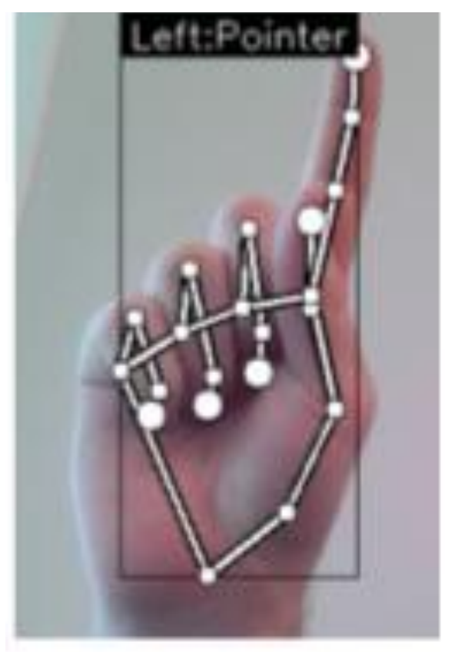


Differentiation

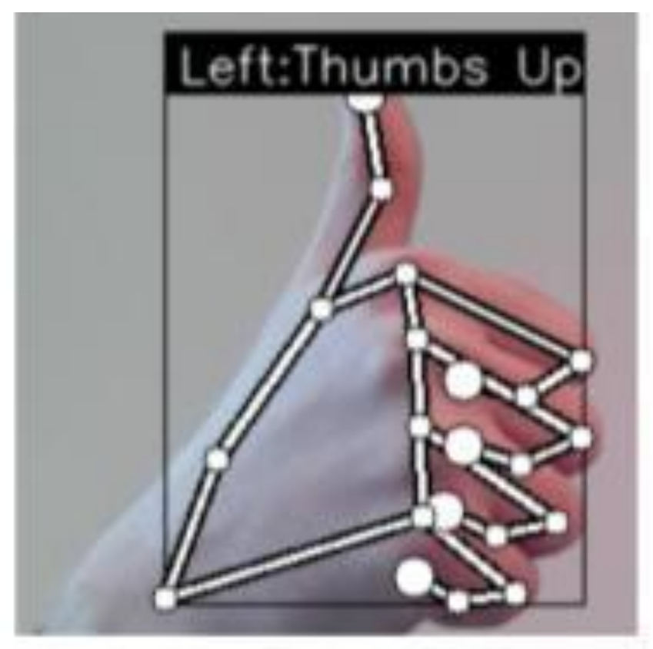
As opposed to most speakers, which require either a button to be pressed or a command to be spoken, this speaker responds to simple hand gestures. Functionality can be enhanced by training the speaker to respond to additional gestures. The system lets the user know that their command was successful via an LED that turns different colors. The user can select songs from a preloaded set of mp3 files.

Results

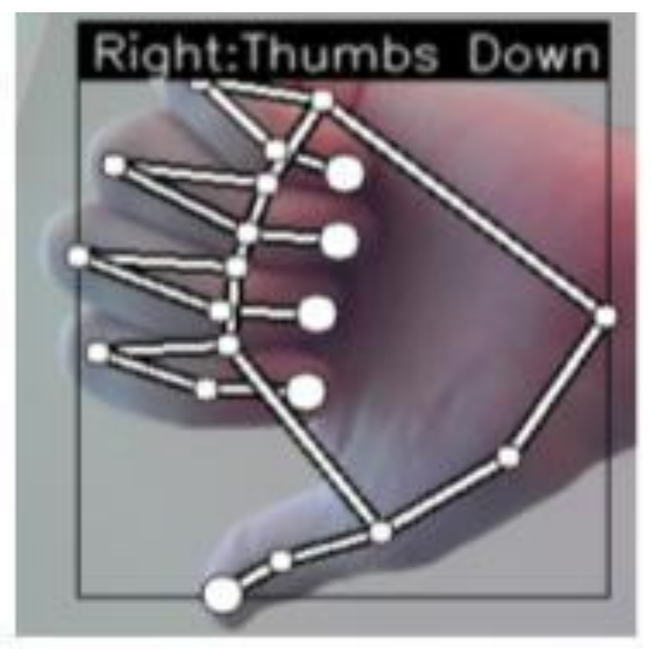
The final product identifies the following five gestures with 92% accuracy.



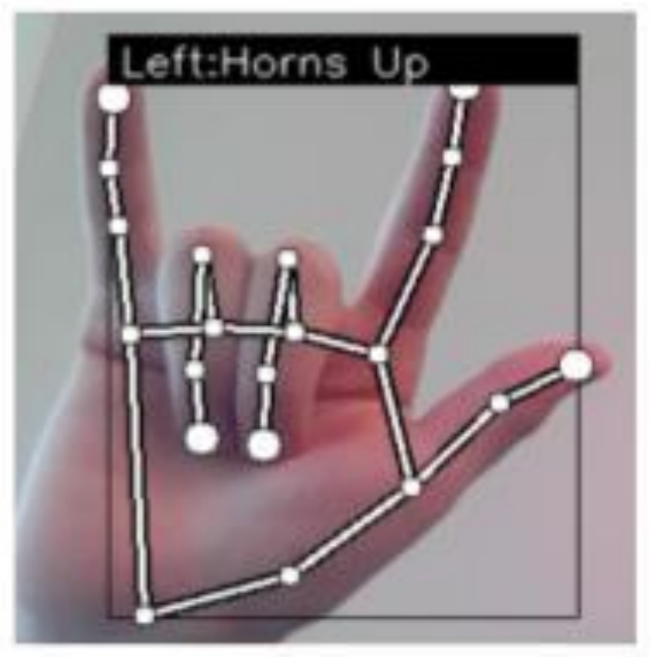
Play/Pause



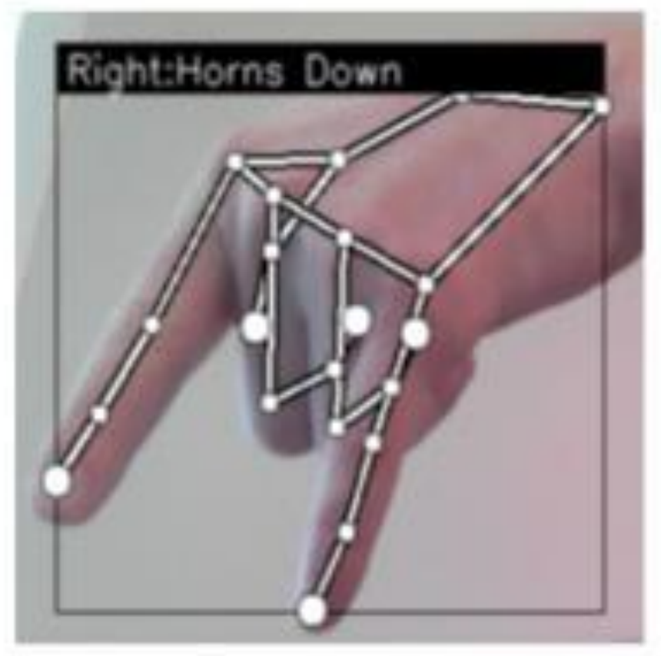
Volume Up



Volume Down



Skip Song



Go Back a Song

The gesture classification only relies on a set of 20 key points, which can be seen in the images above. This allows the system to operate well in a multitude of different lighting conditions. And because the signal processing and actuation tasks are split between two different processors, there is no drop in music quality while the system is processing a new command.

Acknowledgements

We would like to thank Professor Alfred Hero, Dr. Ren Wang, and Mr. Marion Anderson for their advice and guidance and the EECS department for funding.