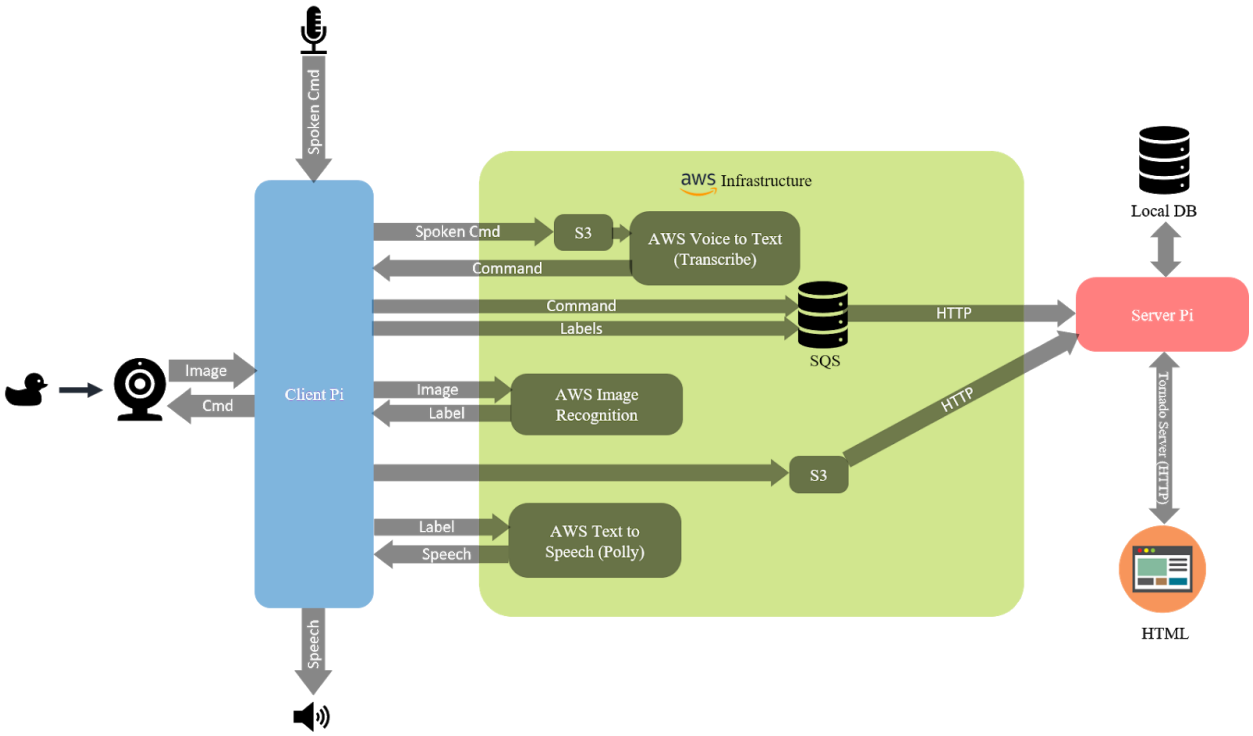
Project Assessment Report

Team Members

Souvik De
Devansh Mittal

Final System/Architecture Diagram and Statement



Project Deviation Statement

1. Added a hotword detection system that keeps listening for the hotword “start” and only when the user says that, the system starts.
2. The voice command is first uploaded to Amazon S3 from where AWS Transcribe transcribes the audio file.
3. Instead of sending the detected voice command and the detected label directly from AWS Transcribe and AWS Image Rekognition respectively to AWS SQS, they are first sent back to the client pi from where they are sent to AWS SQS.
4. The image captured is sent to the server pi via AWS S3.
5. Added a table on the client web page which highlights the last 10 identified objects and whether they were correctly identified or not.

6. Added another table on the client web page which highlights the last 10 voice commands whether they were correct or wrong.

Third-Party Code Used Statement

- Snowboy Hotword Detection Toolkit - https://snowboy.kitt.ai/, https://pimylifeup.com/raspberry-pi-snowboy/
- PyAudio to capture audio - https://makersportal.com/blog/2018/8/23/recording-audio-on-the-raspberry-pi-with-python-and-a-usb-microphone
- Pygame to output audio on speaker - https://raspberrypi.stackexchange.com/questions/7088/playing-audio-files-with-python
- Tornado Web Server - https://os.mbed.com/cookbook/Websockets-Server
- AWS S3 - https://boto3.amazonaws.com/v1/documentation/api/latest/guide/s3-examples.html
- AWS Transcribe - ttps://sunjackson.github.io/2018/08/10/73b0c6a36db932d181f93b01831e3425/, https://docs.aws.amazon.com/code-samples/latest/catalog/python-transcribe-GettingStarted.py.html
- AWS Polly - https://medium.com/@julsimon/johnny-pi-i-am-your-father-part-4-adding-cloud-based-vision-8830c2676113
- AWS SQS - https://boto3.amazonaws.com/v1/documentation/api/latest/guide/sqs.html
- AWS Image Rekognition - https://boto3.amazonaws.com/v1/documentation/api/latest/reference/services/rekognition.html

Project Observation Statement

1. AWS transcribe was used to transcribe the voice commands. It was expected that this process would occur instantly but it takes a minute to return the transcribed word.
2. Images taken in low light were identified correctly. This was better than expected.
3. AWS Transcribe sometimes does not correctly transcribe the spoken command. This might be due to an accent issue.