StenoPhone

Team D6: Cambrea Earley, Ellen Seeser, and Mitchell Yang

Use Case

- Increase of distributed workplaces
- Web conference tech like zoom designed for individual participants
- Our project: a team meeting solution
  - Microphone/speaker hardware and web application
  - Audio conferencing
  - Automatic transcript generation featuring speaker identification

- ECE areas
  - Software (networking, web server, machine learning integration)
  - Signals (audio processing)

Requirements: Audio Conferencing

- Setting Specifications: Conference room
- Audio transmission latency: “mouth to ear delay” from mic to second speaker less than 150 ms
- Audio quality: fewer than 5% of audio packets should be lost
- Battery life: two hours of continuous use

Requirements: Transcripting

- Transcript latency: average word delay of less than three seconds
- Transcript accuracy: word error rate of less than 25%
    - Word errors: insertion, deletion, substitution of words
- Speaker identification accuracy: identification error of less than 25%
- Transcript formatting error for multi-room meetings: 5%
    - Formatting errors: incorrect chronology, missing speaker identification

Key Challenges

- Networking: streaming audio with metadata
    - Risks: Latency, dropped packets
    - Mitigations: UDP connection
- Overlapping audio from multiple users in one room
    - Risks: Can affect transcription
    - Mitigations: Protocol for speaker selection, beamforming
- Identification of moving speakers
    - Risks: speaker movement near the beginning of the meeting when the ML model has very little voice data for the speaker
    - Mitigations: merge ML findings with direction-of-arrival

Solution Approach: Hardware

- Microphone array: ReSpeaker Mic Array v2.0 with attached Speaker
- Local networking and processing: Raspberry Pi 3
- AWS server with both UDP and HTTP/TCP Ports
- Lithium Ion Battery

Solution Approach: Software and ML

Audio Processing: Audacity python scripting

Website: Django, SQLite database

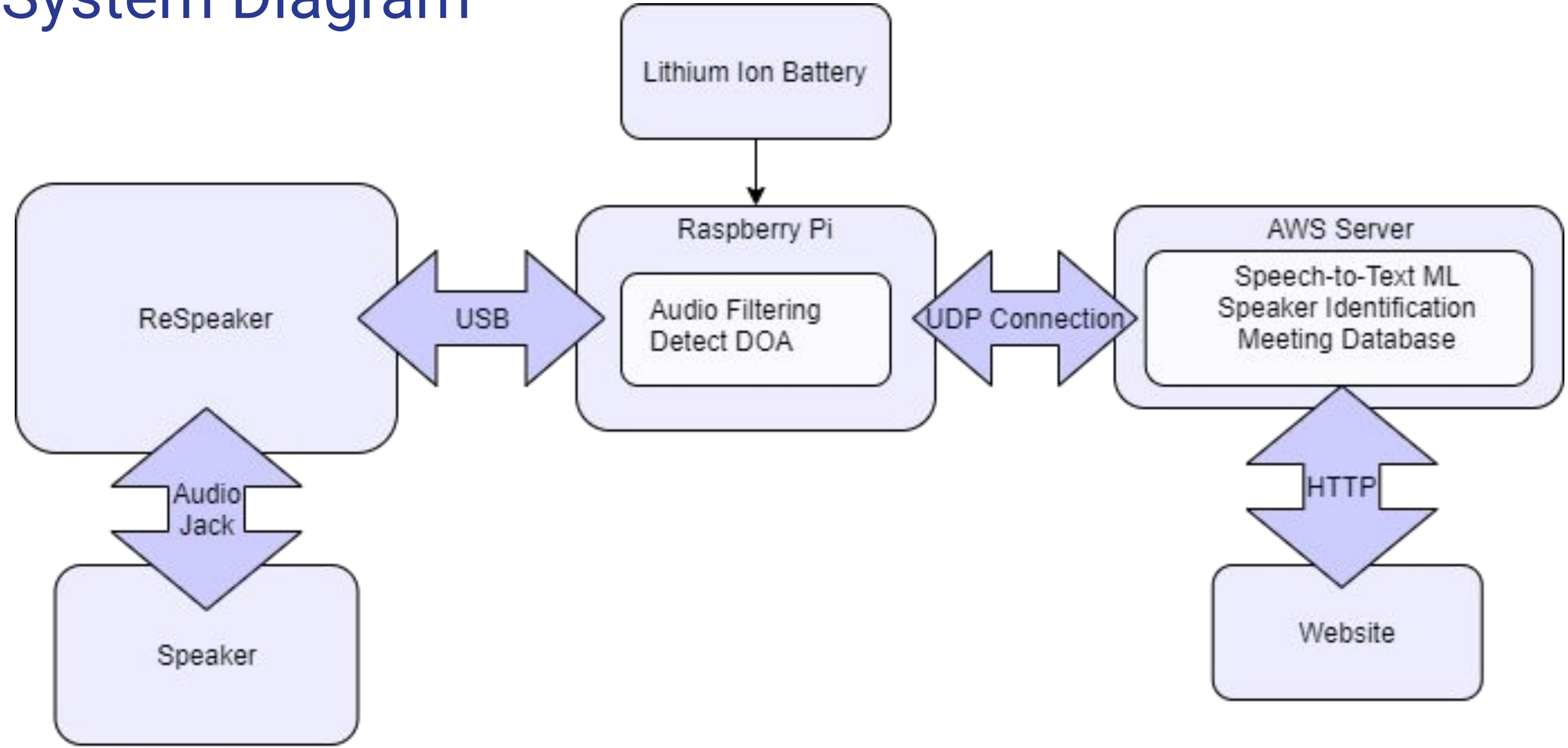
Speaker Identification & Diarization:

Solution	Advertised DER
pyannote audio	25% (2020)
pyBK	12% (2016)
Hitachi Speech EEND	15% (2019)
BUTSpeechFIT VBx	22% (2020)

Speech to Text Solution:

Solution	Advertised WER
Mozilla DeepSpeech	7.5% (2019)
CMU PocketSphinx	10% (2006)
PyPi SpeechRecognition	Variety of engines included

System Diagram



Testing, Verification, and Metrics

Requirement	Metric	Test
Audio Transmission Latency	Mouth-to-Ear Latency (ms) < 150 ms	Capture timestamp of sound snippet input and final sound output after routing through server
Audio Quality	Dropped packets (%) < 5%	Count original and final number of packets after transmitting an audio stream
Battery Life	Hours of continuous use > 2hrs	Run device under heavy load for a set time to find battery usage

Testing, Verification, and Metrics

Transcript Latency	Average Word Delay (s) < 3s	Capture timestamp of audio captured by mic and timestamp of packet arrival in browser
Transcript Accuracy	Word Error Rate (%) < 25%	Check transcript for word error (substitution, deletion, and insertion) after speaking a known text
Speaker Identification Accuracy	Speaker Identification Error (%) < 25%	Check transcript for identification error after conducting a conversation with speaker switches
Formatting Accuracy (chronology and speaker ID tags)	Formatting Error Rate (%) < 5%	Check transcript for formatting error instances after conducting a conversation with known contents

Tasks and Division of Labor

- Cambrea
  - Hardware setup and integration
  - Networking
  - Audio streaming
- Mitchell
  - Audio processing
  - Web app infrastructure
  - Database
- Ellen
  - Machine learning integration
  - Speaker identification system
  - Meeting setup flow

Schedule

			Week 3(Feb 15-21)			Week 4(Feb 22-28)			Week 5(March 1-7)			Week 6(March 8-14)			Week 7(March 15-21)			Week 8(March 22-28)			Week 9(March 29-April 4)			Week 10(April 5-11)			Week 11(April 12-18)			Week 12(April 19-25)			Week 13(April 26-May 2)			Week 14(May 3-5)	
Task	Team Member	Status	M	W	F	M	W	F	M	W	F	M	W	F	M	W	F	M	W	F	M	W	F	M	W	F	M	W	F	M	W	F	M	W	F	M	W
Phase 0: Design	all	progress																																			
Initial research	all	done	all	all	all																																
Project Proposal Presentation	all	progress	all	all	all																																
Phase 1: System Setup	all	todo																																			
Network Design: RPi to AWS	cne	todo				cne	cne	cne																													
Network Design: AWS to Website	cne	todo						cne	cne	cne																											
AWS Database + Django Initialization	mfy	todo				mfy	mfy	mfy																													
AWS EC2 Setup	mfy + cne	todo				mfy	mfy	mfy	cne	cne	cne																										
Speech to Text ML Initial Version	eseeser	todo				eseeser	eseeser	eseeser																													
Design Presentation & Report	all	todo				all	all	all	all	all	all																										
Backend Pre-ML Processing	eseeser	todo						eseeser	eseeser	eseeser																											
Test & Slack Time	all	todo										all																									
Phase 2: Backend	all	todo																																			
Database Creation and Integration	mfy	todo						mfy	mfy	mfy	mfy	mfy	mfy																								
Website Creation, Including Backend	mfy	todo						mfy	mfy	mfy	mfy	mfy	mfy																								
Non-Moving Speaker ID (no ML)	eseeser	todo								eseeser	eseeser	eseeser	eseeser																								
Server connection at AWS	cne	todo										cne	cne	cne																							
Server connection at RaspberryPi	cne	todo											cne	cne	cne	cne	cne																				
Speech to Text ML Integration	eseeser	todo											eseeser	eseeser	eseeser																						
Audacity Filter Design + Audio Process	mfy + eseeser	todo								mfy	mfy	mfy	mfy	mfy	eseeser	eseeser	eseeser																				
Hardware Setup and Integration	cne	todo													cne	cne	cne	cne																			
Respeaker Config on Raspberry Pi	cne	todo														cne	cne	cne																			
Mic Initialization Backend	mfy	todo															mfy	mfy	mfy	mfy																	
Testing & Slack Time	all	todo																all	all	all																	
Phase 3: Frontend	all	todo																																			
Transcript Support for Multi-mic Meetin	eseeser	todo																eseeser	eseeser	eseeser																	
Full On-Website Meeting Setup	eseeser	todo																eseeser	eseeser	eseeser	eseeser	eseeser															
Simultaneous Speaker Handling	mfy	todo																		mfy	mfy	mfy	mfy	mfy	mfy	mfy											
ML + Hardware Integration	cne	todo																			cne	cne	cne														
Moving Speaker ID (ML)	eseeser	todo																			eseeser	eseeser	eseeser	eseeser	eseeser												
Transcript Web Streaming	cne	todo																				cne	cne	cne													
Testing & Slack Time	all	todo																						all	all	all											
Phase 4: Touchup	all	todo																																			
Web CSS	all	todo																								all	all	all	all								
Integration and Complete System Testi	all	todo																			all	all	all	all	all	all	all	all	all	all							
Slack Time	all	todo																												all	all	all	all	all			
Final Report and Presentation	all	todo																									all	all	all	all	all	all	all	all	all	all	