CueMeIn - Product Requirements Document

About the Team

Team Name: Runtime Terror

Company: LogMeIn

Team Members:

Mason Corey (Lead)

Javier Jimenez (Scribe)

Juan Mendoza-Martinez

Leena Shankar

Armaan Shah

Introduction

What is the problem?

Nearly 1 in 4 people in the United States live with a disability that impacts major life activities. Certain disabilities - specifically intellectual or developmental ones - can cause limitations in cognitive function and abilities and even result in weakened communication skills. As the world shifts towards technologization and online interaction becomes more standard, communication is made more challenging when interacting through video chat, conference calling, and other similar online platforms. Individuals with certain cognitive disabilities can have difficulties connecting effectively over video chat, as communication protocols are different online and may be too subtle to pick up on. It is often difficult for

these individuals to read social cues or understand facial expressions and emotional content in an online environment.

Given the current COVID pandemic, the standard mode of communication - be it social, academic, or official - has almost completely shifted to virtual. It is very probable that virtual communication in all areas will continue long-term; consequently, it becomes even more urgent to make this transition easier for people with disabilities.

How is the problem addressed today?

As of right now, there are a few applications available that enhance communication for people with disabilities by means of improving social and emotional capabilities; however, there are no customizable tools that focus specifically on improving the video conferencing experience. Some of the existing available technologies include:

- SayWAT - A wearable technology that can assist and deliver feedback to individuals during face-to-face conversations to aid with communication issues
- Learn With Rufus - An iOS application that assists children and adults with autism in understanding, interpreting, and portraying emotions
- Affectiva SDK - An AI company that focuses on emotional recognition through the use of facial recognition, facial feature extraction, and emotional classification

Objective, Implementation, and Assumptions

What is our objective and who will benefit from it?

Our objective is to develop a unique tool for the Special Olympics - a non-profit organization that provides sports training and athletic competition for children and adults with intellectual disabilities - that enhances virtual communication for people with disabilities. This tool will be designed for people with intellectual and cognitive disabilities that prevent them from effectively recognizing and interpreting facial expressions during video conferencing. Autism and fragile X syndrome are examples of two such developmental disabilities that are commonly seen in Special Olympics athletes.

How will we implement our solution?

We intend to read the emotional state of the active 'dominant' speaker in a conference call and display that emotion on the screen as a readable popup, so people with cognitive disabilities can practice recognizing and interpreting emotions in real-time. This will be achieved by using the Twilio APIs to obtain the audio and video inputs of the current active speaker and passing these inputs to machine learning models - specifically, IBM Watson and Microsoft Computer Vision for video processing and Google NLP for natural language processing.

Within the front-end UI, there will be a toolbar that has links to various pages on the website: a home page, a create account page, and a login page. Once logged in, a control panel page will appear that gives the user access to a page that allows the user to host or join a video conferencing room. The video conferencing room will display the video streams of all participants at the top of the screen, with the user's own video stream appearing first, and will display the video stream of the current dominant speaker in the center of the room. Each

participant's video stream will have that participant's unique User ID available to view if a user hovers over that video stream. There will be a button labeled "Sentiment Analysis" directly below the dominant speaker's video stream, and clicking on this button sends a request to retrieve the dominant speaker's current emotion. The sentiment analysis output will then be displayed to the user as an emoticon in the top left corner of the dominant speaker's video stream. We also plan to implement a border around the dominant speaker's video screen which updates with a different corresponding color each time the dominant speaker's emotion changes.

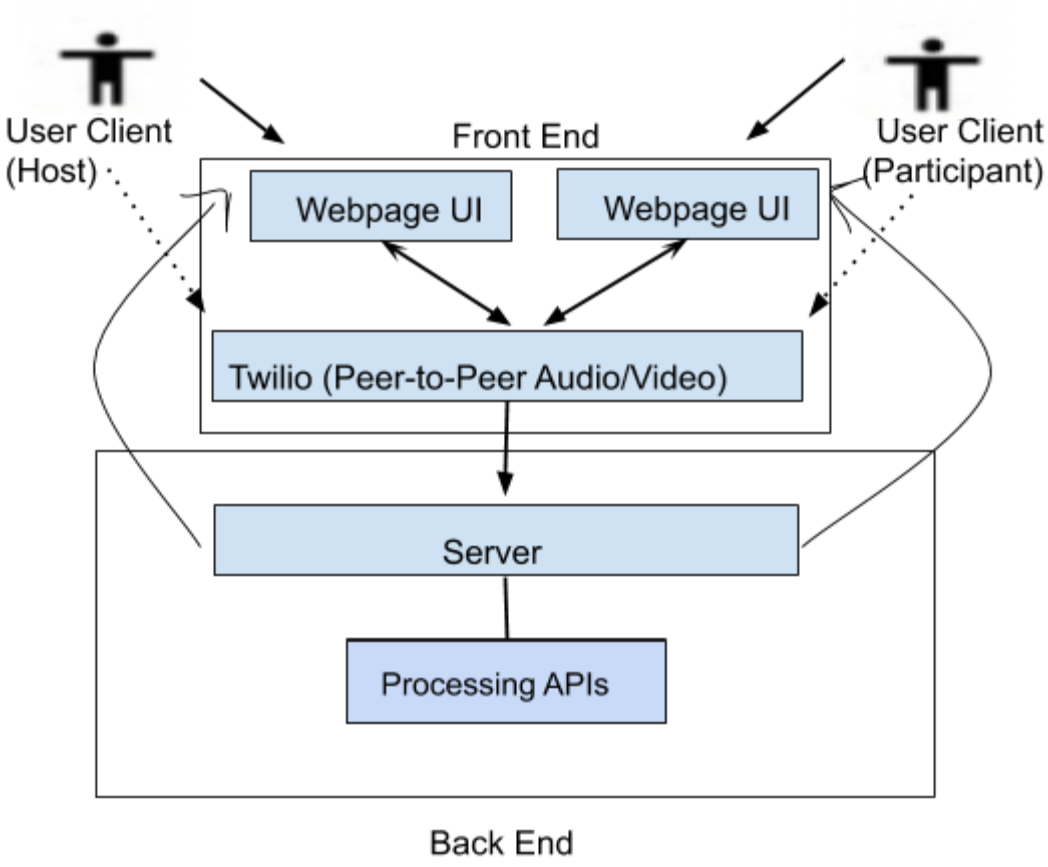
To host the backend, we will be utilizing an AWS server which will be connected to the Twilio REST API using NodeJS. We plan to use two local JSON files in dictionary formats to store input and output user information. The first file's dictionary will be storage for user information for when users create accounts, so that they may log in later. The second file's dictionary will contain live information about the current sentiment of each user in a video room, so that the Front End can easily query the Back End for the user's sentiment analysis without having to wait for the image and audio processing to be done. As such, this will necessitate an asynchronous workflow between the Front End and Back End, where the Front End will query the Back End's sentiment analysis JSON file on its own timer and the Back End will refresh these JSON values only for the Dominant Speaker (whose sentiment analysis will be displayed by the Front End) on its own separate timer. In terms of workflow for the processing APIs, they will have images and audio clips fed to them in regular intervals by the Twilio API on the Back End (which will monitor the Room's Dominant Speaker and regularly pass files to the sentiment analysis processors), which will then be processed by their respective APIs (Microsoft Vision and IBM Watson) and passed to a final analyzer, which will combine the results of the Image and Audio Processors

locally to get a final prediction of the emotion. This will then be saved into the JSON file for the appropriate user for use at a later time by the Front End.

What are the necessary assumptions to be made?

This application will only be available for conference calls hosted on a server, and will only be available as a web application. Each user will need to submit their name and unique room name to login and to host or join a video conferencing room.

System Architecture Overview



The diagram to the right depicts how the flow of data starts and ends. At the beginning, the user will be interfacing via the frontend, which will mainly consist of usage of the Twilio API for audio and video input/output. The user will have a login for the application that we will store in our database so they can sign in as well as speak to other users. Now interfacing with our backend, we will be taking the audio/video inputs and using multiple APIs for the purpose of video processing, natural language processing, and sentiment analysis. The outputs we take from these APIs will be passed to the frontend so the user can interact with or visualize what the other users are saying and feeling.

User Stories

As a user, I can visit the Home Page so that I can exit out of whatever page I'm currently on and have a base page to navigate from

Github links (issues):
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/1

- **Scenario 1:** The user clicks the Home Page from an ongoing video call
    - The user gets a prompt asking if they would like to leave the current page and if they click yes, get redirected to the Home Page
- **Scenario 2:** The user clicks the Home Page in any other scenario
    - Whatever forms the user already entered are cleared and the user gets redirected to the Home Page.

As a user, I can access the toolbar so I can navigate to different pages on the website

Github links (issues):
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/2

- **Scenario 1:** The user is logged into their account
    - The user can access the Home, Host a Room, Join a Room, and Logout pages via the toolbar (in that order).
- **Scenario 2:** The user is not logged into an account yet
    - The user can access the Home, Login, Create an Account, and the Join a Room pages via the toolbar (in that order). A user that hasn't been logged in cannot host a video call.

As a user, I can create an account so I can login and start a video call

Github links (issues):
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/5

- **Scenario 1:** There is no existing account associated with the email address provided
    - User can access the “create account” page via the toolbar
    - User can create a username and a password (minimum of 8 characters for a strong password)
    - User will be able to create an account successfully and fail creation if credentials are invalid
- **Scenario 2:** There is an existing account with the email address provided
    - User will encounter an error page that says an account already exists
    - User will be able to reset password to that existing account

As a user, I can log in to my existing CueMeIn account so that I can start a video call

Github links (issues):
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/4

- **Scenario 1:** The login credentials match the json file on our backend and the Twilio login succeeds
    - User can access their account
    - User will be able to host a room
- **Scenario 2:** The login credentials do not match the json file on our backend and the Twilio login fails
    - User will get an error message saying their credentials are invalid
    - User will be told to attempt login again

As a user, I can Host a Video Call from the Control Panel Page so I can start a Video Call

Git Issue: https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/3

- **Scenario 1:** User is prompted to the Host a Room Page after Login Credentials Succeed
    - Given the user login credentials are valid
    - After logging in through the Login page
    - Host a Room Page appears; Room and Name are required to join session
- **Scenario 2:** User is not prompted to the Host a Room Page after Login Credentials Fail
    - Given the user login credentials are not valid
    - After logging in through the Login page
    - Host a Room Page is not prompted
    - Error message is displayed

As a user, I can request to host a meeting so that I can communicate with my group via video call

Github Issue:
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/6

- **Scenario 1**: User clicks on “Host Meeting” to host a meeting
    - Given the user has created a Twilio account
    - Given “Host Meeting” is selected
    - Twilio API is called to create Meeting ID
    - Twilio REST API connects to server to host meeting
    - User is prompted to meeting room page
    - User can invite others to join with Meeting ID
- **Scenario 2**: User does not have a Twilio account set up
    - Error message is displayed

As a User, I can request to join a meeting so that I can communicate with my group via video call

Github Issue:
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/7

- **Scenario 1**: User successfully enters the meeting
    - The host or participant must have requested a meeting with one or more participants.
    - Given the meeting ID, user can join with the camera and audio on
    - Use the twilio API to make the connection/video feed.
- **Scenario 2**: User is not prompted to the meeting
    - Meeting ID was incorrectly given
    - Prompt a message that the ID is invalid
        - Error message is displayed

As a User, I can leave the meeting.

Git Issue: https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/8

- **Scenario 1**: User leaves the meeting as a host
    - The host must have requested a meeting with one or more participants.
    - The button to leave the meeting will send you back to the Host a Room Page
    - The participants will also be forced to leave the meeting and return to the Join a Room Page.
- **Scenario 2**: User leaves the meeting as a participant
    - The participant must have requested a meeting with one or more participants.
    - The button to leave the meeting will send you back to the Join a Room Page

As a user, I can see the video from the other participants, so that I can see their faces when they are talking

Github links (issues):
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/8

- **Scenario 1:** Video is successfully transmitted from another participant
    - Given the user and the other participant are in the same session
    - The user will be able to see the other participant's video stream in the window for that participant's video.
- **Scenario 2:** Video is not successfully transmitted from another participant
    - Given the user and the other participant are in the same session
    - The user will see a black screen with an error symbol where the participant's video stream should be

As a user, I receive the audio streams from the other participants, so that I can hear them when they are talking

Github links (issues):
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/8

- **Scenario 1:** Audio is successfully transmitted from another participant
    - Given the user and the other participant are in the same session
    - The user will be able to hear the other participant's audio, mixed with all the other audio inputs
- **Scenario 2:** Audio is not successfully transmitted from another participant
    - Given the user and the other participant are in the same session
    - The user will see an error symbol in the corner of the participant's video stream view that will have a mic symbol and an error (exclamation point) next to it

As a user, I can read the current emotion of the dominant speaker so I can figure out how he or she is feeling

Github links (issues):
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/9

- **Scenario 1:** Emotion of the dominant speaker is successfully read
    - Given there is a dominant speaker
    - Display the emotion on the screen where the Front End determines it should be (probably somewhere around the Dominant Speaker's video feed).
- **Scenario 2:** Emotion of the dominant speaker is not successfully read
    - Given there is a dominant speaker
    - Display an error message above the Dominant Speaker's video feed (i.e. emotion could not be read)
- **Scenario 3:** There is no dominant speaker
    - Given there is no dominant speaker
    - Do not display anything
- **Scenario 4:** The dominant speaker changes
    - Given the dominant speaker changes
    - Change the positioning of the emotion display

As a user, I can mute my audio to either block or allow other participants to listen to my audio tracks.

Github links (issues):
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/10

- **Scenario 1:** User is currently NOT muted
    - The user must be successful in joining a meeting.
    - The user can press the "mute" button to prevent other participants from hearing the user's audio tracks

- **Scenario 2:** User is currently muted
    - The user must be successful in joining a meeting
    - The user can press the “unmute” button to allow other participants to hear the user’s audio tracks

As a user, I can turn my camera on and off to either block or allow other participants to view my video tracks.

Github links (issues):
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/11

- **Scenario 1:** User’s video camera is currently on
    - The user must be successful in joining a meeting
    - The user can press the “video off” button to prevent other participants from viewing the user’s video tracks
- **Scenario 2:** User’s video camera is currently off
    - The user must be successful in joining a meeting
    - The user can press the “video on” button

As a user, I can either deafen the audio to block both other participants from hearing me and me from hearing other participants, or undeafen the audio to the opposite.

Github links (issues):
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/10

- **Scenario 1:** User’s audio is currently NOT deafened
    - The user must be first successful in joining a meeting
    - The user can press the “deafen” button to both block himself from hearing other participants’ audio tracks and block other participants from hearing the user’s audio tracks
- **Scenario 2:** User’s audio is currently deafened

    - The user must be first successful in joining a meeting
    - The user can press the “undeafen” button to allow himself to hear other participants’ audio tracks and allow other participants to hear the user’s audio tracks

As a user, I can press the “Sentiment Analysis” button to receive an updated visual cue to the dominant speaker’s sentiment analysis.

Github links (issues):
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/9

- **Scenario 1:** User presses button when there is no change or update in the dominant speaker’s emotion
    - Message will pop up alerting user that dominant speaker’s emotion has not changed
    - No changes will be made in the border color of dominant speaker
    - No changes will be made to the dominant speaker’s emoji
- **Scenario 2:** User presses button when there is a change or update in the dominant speaker’s emotion
    - Border color will update to new corresponding color for that emotion
    - Emoji will update to new corresponding emoji for that emotion

As a user, I can hover over a participant’s camera to view their User ID name.

Github links (issues):
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/8

- **Scenario 1:** User hovers over participant’s camera
    - Participant must have a valid name
    - Participant’s name will be displayed in a text container at the bottom of their video box
- **Scenario 2:** User does not hover over participant’s camera

    - Participant's name will not be visible

As a user, I can choose to have the dominant speaker sentiment analysis shown as a visual representation or audio response (for visually impaired).

Github links (issues):

https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/9

- **Scenario 1:** User hovers over to sentiment analysis display settings.
    - Participant hovers mouse over sentiment analysis button and clicks it
    - The sentiment analysis button prompts the options
- **Scenario 2:** User does not hover over display settings
    - By default, the visual aid will be prompted when the sentiment analysis is posted

As a user, I can hide my sentiment analysis to other users.

Github links (issues):

https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/8

- **Scenario 1:** User is entering meeting and chooses to show sentiment analysis
    - Participant clicks on button to display their emotion while not an active dominant speaker
- **Scenario 2:** User is entering meeting and chooses not to show sentiment analysis
    - Participant click button to not display the emotion while not an active speaker
    - Participant's emotion is read while dominant speaker

As a user in a call, I am able to toggle open a chatbox to send a message to the other user in the call

Github links (issues):
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/12

- **Scenario 1:** Chat is received successfully
    - Given that at least two users are in a session
    - When the user opens the chat box and types in a message
    - They will be able to send the message to the other user.
- **Scenario 2:** Chat is not received successfully
    - Given that at least two users are in a session
    - When the user opens the chat box and types in a message
    - When the user sends the message, a red exclamation point (error) will show up and the message will not be displayed

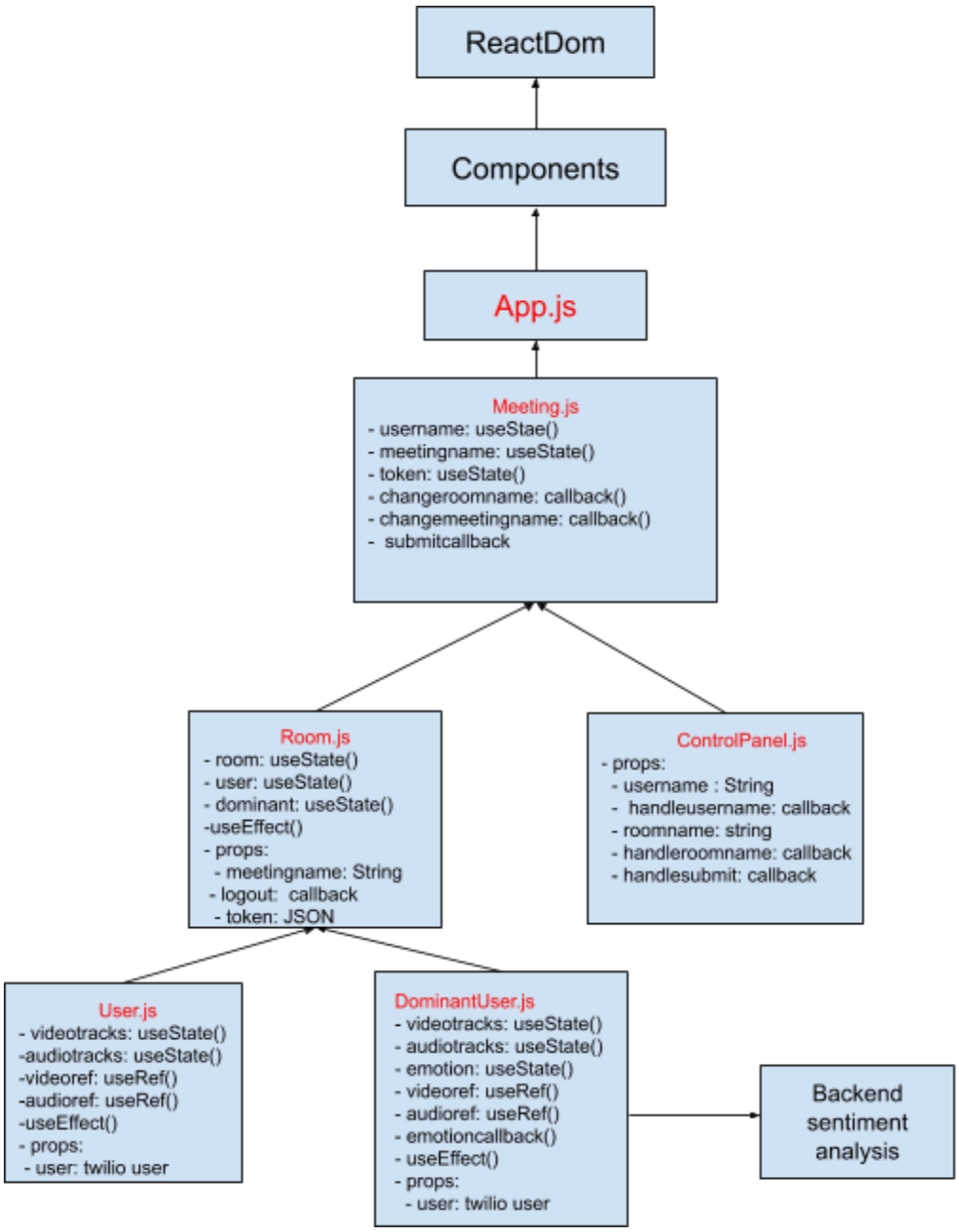
As a user in a call, I am able to request sentiment analysis for messages in the chat box to understand the emotion of the message

Github links (issues):
https://github.com/themason2011/CS189A-LogMeIn-Team-/issues/12

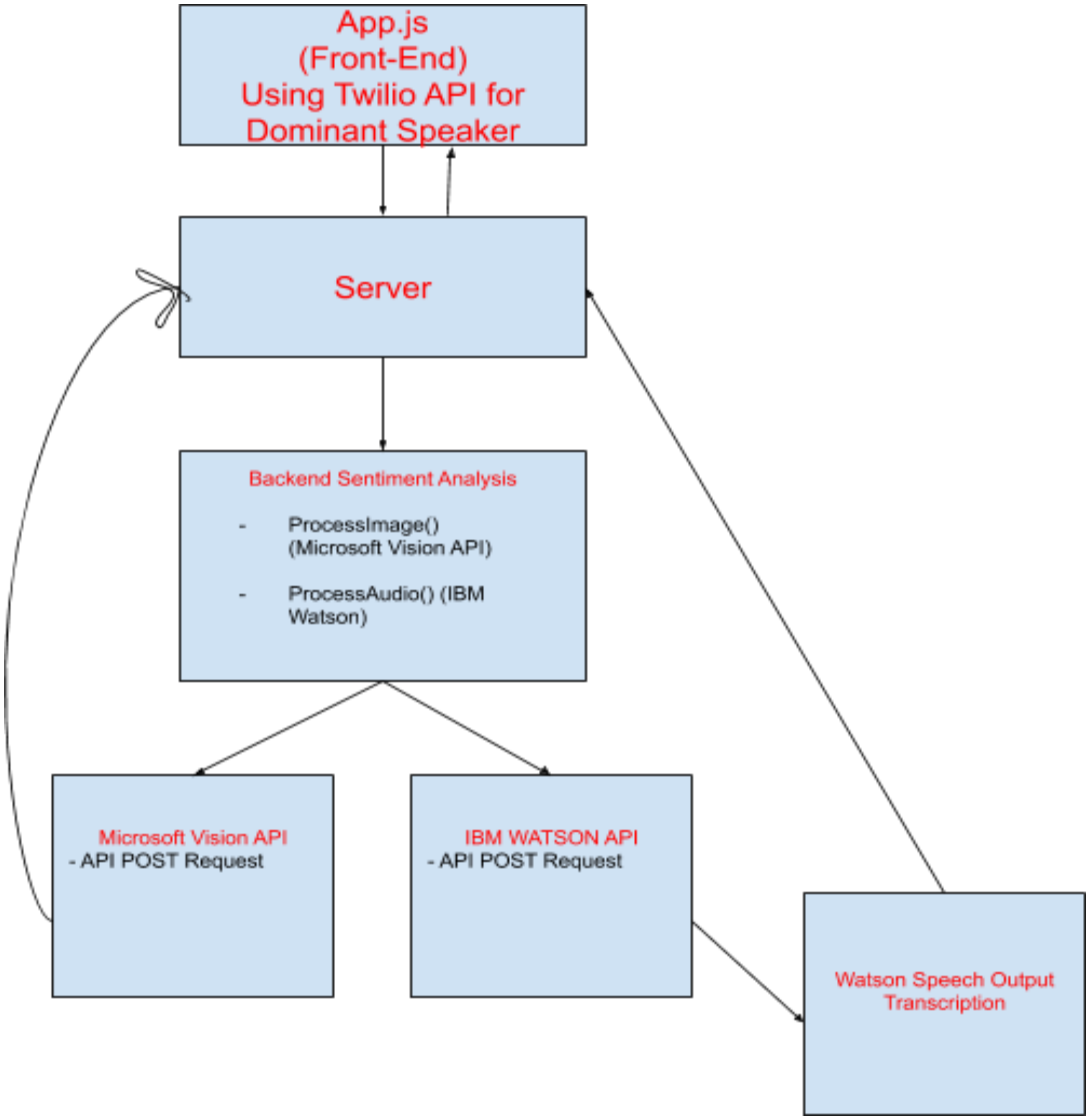
- **Scenario 1:** Sentiment Analysis for chat message is received successfully
    - Given user clicks the sentiment analysis button for a chat message
    - And the message has sentiment in it/is long enough to return a valid sentiment
    - Then the sentiment analysis result will show up next to the chat message
- **Scenario 2:** Sentiment Analysis for chat message is not received successfully
    - Given user clicks the sentiment analysis button for a chat message
    - And the message does not contain sentiment in it/is not long enough to return a valid or significant emotion
    - Then there will be a red exclamation point (error) where the sentiment analysis would normally be

Diagrams

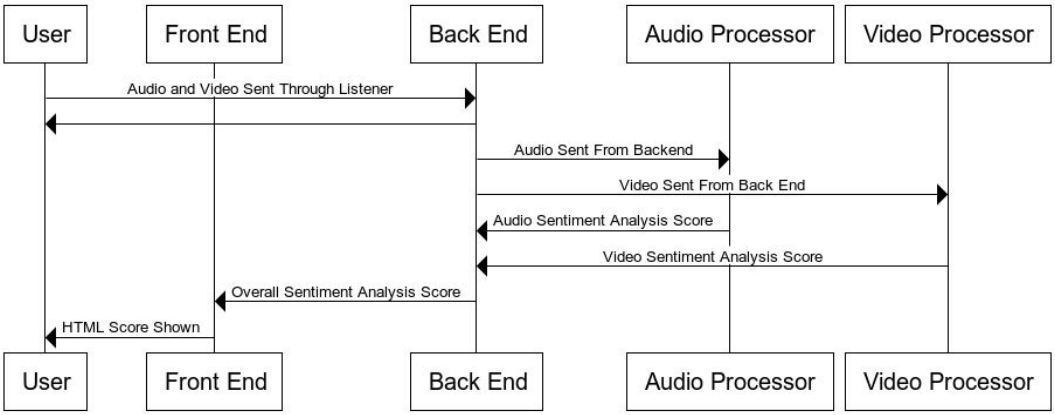
Front-end System Model



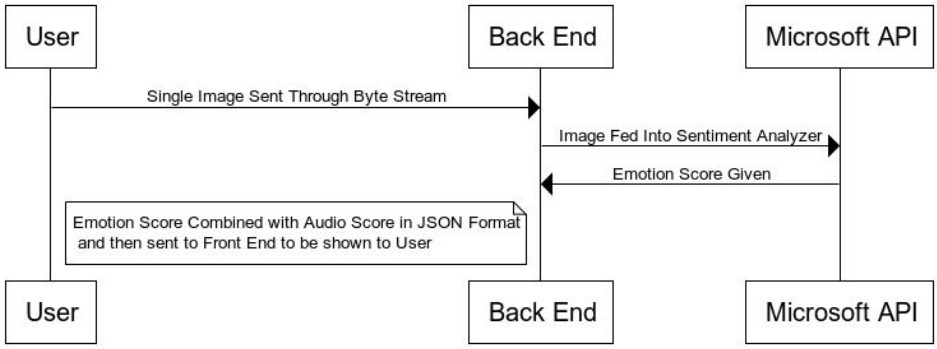
Back-end System Model



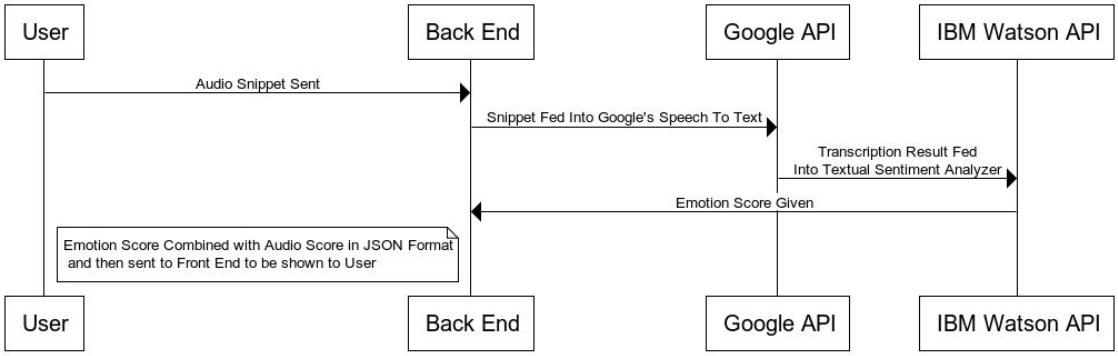
Sentiment Analysis Sequence Diagram



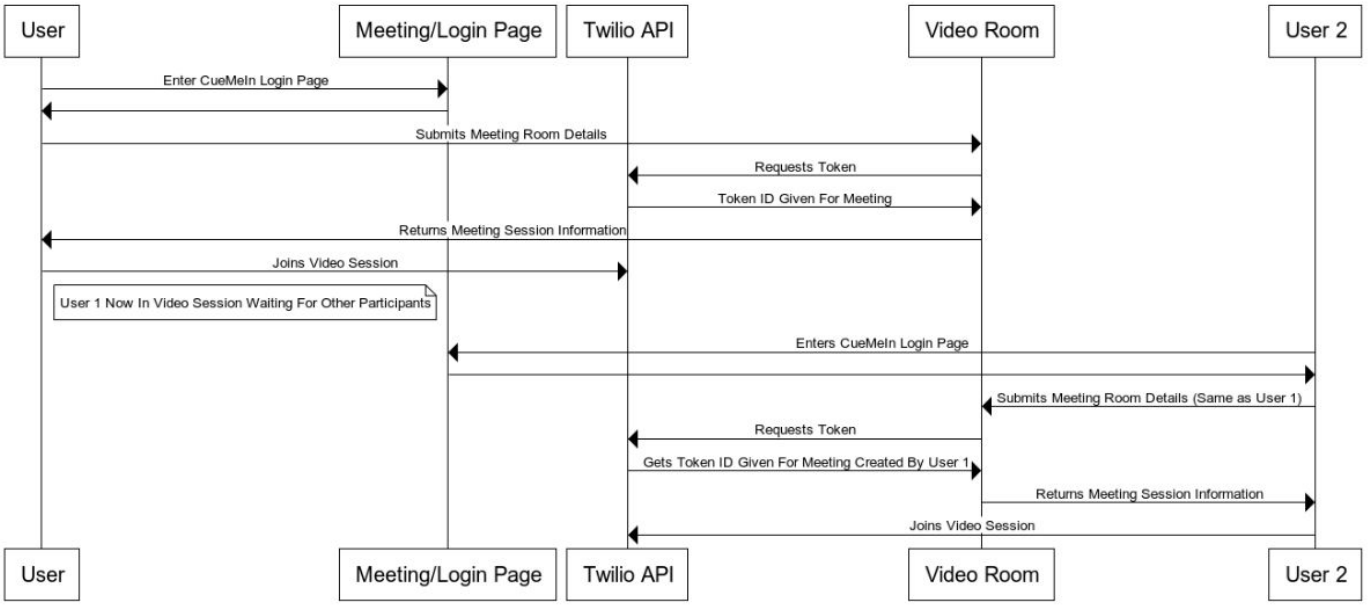
Video Sentiment Analysis Sequence Diagram



Audio Sentiment Analysis Sequence Diagram

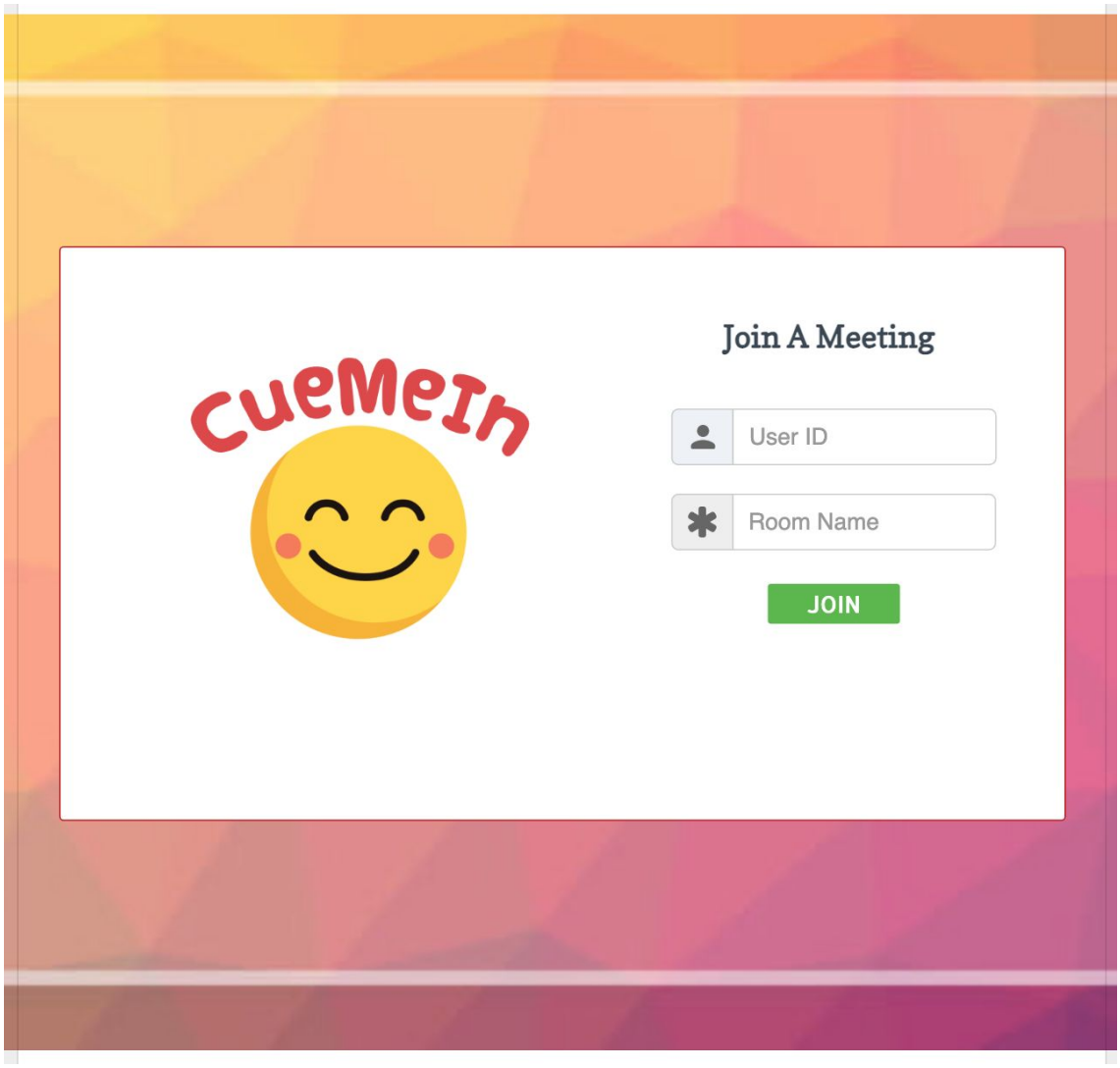


User Interaction Sequence Diagram



UI Mockups

Login Page Mockup



Video Page Mockup

Technologies Employed

- **Frontend:**

-

- **Backend:**
  -

  -

  - https://azure.microsoft.com/en-us/services/cognitive-services/#features

  - https://www.ibm.com/watson/products-services

- **Database:**
  -