Bilkent University

Computer Science

CS491

Senior Design Project

SummarEyes

2023 - 2024 Fall Semester

Project Specification Document

Project Team ID: T2315

Project Team Member Names:

Kürşad Güzelkaya (21902103)

Hazal Buluş (21903435)

Süleyman Gökhan Tekin (21902512)

Deniz Gökçen (22002850)

Burcu Kaplan (21902270)

Advisor: Uğur Güdükbay

Table of Contents

1. Introduction 4
    1.1 Description 5
    1.2 High Level System Architecture & Components of Proposed Solution 6
        1.2.1 Web App 6
        1.2.2 Online Meeting Service 7
        1.2.3 Speech-to-Text 7
        1.2.4 Distraction Detection 7
        1.2.5 Personalized Summary 7
        1.2.6 Speaker Feedback 8
    1.3 Constraints 8
        1.3.1. Implementation Constraints 8
        1.3.2. Economic Constraints 9
        1.3.3. Ethical Constraints 9
    1.4 Professional and Ethical Issues 10
2. Design Requirements 10
    2.1. Functional Requirements 10
        2.1.1 Registration: 10
        2.1.2 Login/Logout: 10
        2.1.3 Data Usage Permission: 10
        2.1.4 Create a Meeting: 11
        2.1.5 Join a Meeting: 11
        2.1.6 Turn On/Off Mic and Camera: 11
        2.1.7 Summarize Meeting: 11
        2.1.8 See Previous Summaries: 11
        2.1.9 See Previous Feedbacks: 12
        2.1.10 Download Summarization/Feedback: 12
        2.1.11 End Meeting: 12
    2.2. Non-Functional Requirements 12
        2.2.1. Usability 12
        2.2.2. Reliability 12
        2.2.3. Performance 13
        2.2.4. Supportability 13
        2.2.5. Scalability 13
3. Feasibility Discussions 14
    3.1. Market & Competitive Analysis 14
    3.2. Academic Analysis 14
4. References 15

1. Introduction

After the pandemic, remote working in companies and remote education have become popular since they provide time efficiency, comfortable work, and an educational environment at a lower cost. The crucial part of remote working and education is the online meeting which enables employees to collaborate and communicate effectively without needing physical presence. However, concentration is a significant problem in online meetings since it is hard to focus on a computer screen for a long time without getting distracted.

Recently, there has been interest in the development of several summary systems designed specifically for virtual meetings. Nevertheless, their efficacy frequently fails to satisfy the particular requirements of different consumers. These tools wind up taking up participants' important time instead of providing specific assistance, especially for individuals who have only missed small portions of the conference. The poor accuracy displayed by some systems when processing languages other than English is another significant issue. This language barrier makes these technologies even less useful and dependable, which presents difficulties for users looking for thorough and accurate meeting summaries in a variety of linguistic circumstances.

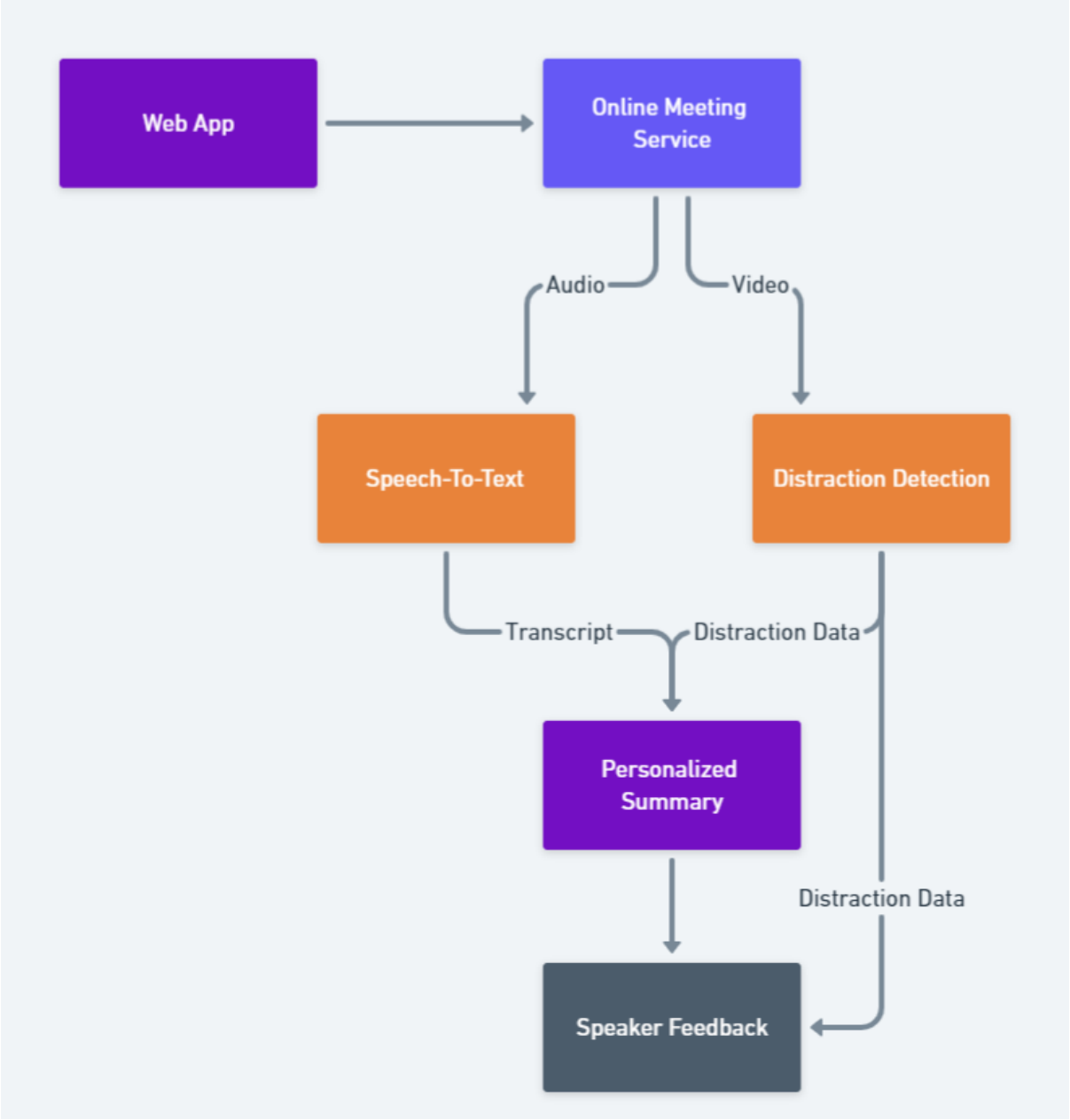
SummarEyes offers a cutting-edge solution for enhancing online meetings and lectures. Leveraging the power of machine learning and computer vision, it is aimed to detect and track instances when participants become distracted, ensuring that valuable content is not missed. The pretrained language model will generate comprehensive summaries with topic titles and timestamps for all the critical discussions during those distracted moments. These summaries are invaluable for attendees to catch up and for speakers to gain insights into audience engagement. SummaEyes will empower speakers to deliver more engaging and impactful presentations by providing feedback on when distractions occur most frequently. This innovation will revolutionize how companies and educational institutions approach online communication, making meetings and lectures more productive and engaging for everyone involved.

This report includes the proposed project plan, along with a full explanation of the application, limitations, ethical and professional concerns, and functional and non-functional needs.

1.1 Description

SummarEyes will be a Website Application. It will provide basic requirements for online meetings with the help of Agora SDK. Users will be able to create and attend online meetings. After the meeting is over, the user's video data will be analyzed to detect the moments the user gets distracted. The analysis consists of two parts. The first emotions of the user will be detected from their facial expression with the help FER library of Python. These emotion data will be converted to concentration data using a point-based system for emotions [1]. In the second part of the video analysis, the eye-tracking model that we will develop will find the moments the user is not looking screen for a considerable amount of time. Then, these two data will be merged to find the time slots that the user is distracted most. Audio of the meeting will be given to a speech-to-text tool which is Whisper AI, to get the transcript of the meeting. This transcript and distraction data will be given to Llama2, which is a pre-trained language model, to get a personalized summary of the meeting that contains all the critical discussions during those distracted moments. Gensim will be used to generate a general summary of the meeting for the people who couldn't attend the meeting. In addition to summaries, a report about most of the users who get distracted will be created using distraction data and sent to the meeting host.

1.2 High Level System Architecture & Components of Proposed Solution



1.2.1 Web App

The project will be a Web application which consists of a home page and user profiles. After signing up, users will be directed to the home page, where they can start or join an online meeting. On their profile pages, they can find the personalized

summaries of the meetings they joined as well as the general summary of the meetings. If they hosted any online meeting before, they will also find a feedback section in their profile which provides a report about the most common time ranges the other participants lost focus to the meeting they started. Django and React will be used to implement the backend and the frontend of the web application.

1.2.2 Online Meeting Service

The online meeting service of the application will enable users to have a new platform to conduct their online meetings. They will be able to start a new meeting or join an existing one by entering the meeting-ids. During the meeting, they will have options to turn on/off their camera or microphones, see the other participants, and leave the meeting. Agora SDK will be used to provide online meeting features.

1.2.3 Speech-to-Text

After a meeting ends, the record of the online meeting will be saved to the database and sent to the speech-to-text model to get the transcript of the meeting record. During the transcription process, the time when the sentences are received will also be saved in certain periods by using the same model to find the exact parts that the user missed. Whisper AI will be used to convert speech to text and produce a meeting transcript.

1.2.4 Distraction Detection

The saved record of the meeting will also be used for distraction detection. By using facial expression analysis and eye-tracking AI models, the application will find the time ranges when the user is distracted. Those saved time ranges will be used to find what exactly was talked about during that time range from the transcripted text. Python’s FER library will be used to detect emotions from facial expressions.

1.2.5 Personalized Summary

After getting the transcription of the meeting record and the times the user is distracted, we will check the saved times during the transcription and find the parts the user missed. Then, the missed parts because of the distraction will be sent to a pre-trained language model, and a personalized summary will be created. The general summary of the whole meeting will also be generated and sent to all of the participants of the meeting. Llama2 and Gensim will be used to summarize the meeting.

1.2.6 Speaker Feedback

The time information of all participants, when they are distracted, will be saved and analyzed to give feedback to the host. The feedback will include the time ranges when most speakers lost focus on the meeting. To create a fair and safe environment for the participants, the feedback will consist of anonymous data, and the host of the meeting won't be able to see when a particular user is not paying attention.

1.3 Constraints

1.3.1. Implementation Constraints

- To follow the development process of the project, Jira and Github technologies will be used.
- The project will be a Web application that uses Agora SDK to create the online meeting platform of the project.
- PostgreSQL will be used as the database management system.
- Django, a Python-based framework, will be used to build the website of the application.
- React library will be used for the front-end development of the project.
- To transcript the online meeting, a pre-trained AI model Whisper (by OpenAI) will be used as well as some other helper machine learning libraries.
- Python's FER library will be used to perform emotion detection of the users to help with distraction detection [2].

- Gensim, a machine learning library, will be used to generate the summaries of the transcripted meeting record.
- Llama2, a pre-trained language model, will be used to generate personalized summaries and topic predictions.

1.3.2. Economic Constraints

- To be able to provide online meeting features, we need to use Agora SDK, which gives 10,000 minutes of video call free each month and charges additional time with 3.99$/1000 minutes [3].
- We plan to use a pre-trained model of Whisper AI to convert speech-to-text, which is open-source and free. When the user count increases, using the premium version, which provides optimization with GPUs, will be an option with 6$ for 1000 minutes [4].
- Language model Llama2 is free [5].
- The libraries and frameworks we will use to implement our application are free.
- We plan to use Google Colab, which provides free GPUs for training the DL model.

1.3.3. Ethical Constraints

- Necessary consent will be taken from the user to record their video and analyze it to produce personalized summaries. After analyzing video data which will be stored privately, it will be deleted.
- We will inform users about the precision of the technology's evaluations and offer clarity on the methods used to identify distractions and produce summaries. Users will be aware of the system's limitations and possible mistakes.
- The gathered data —including video feeds and summaries— will be protected from abuse, breaches, and unwanted access. It is planned to use strong encryption techniques and safe storage mechanisms.

1.4 Professional and Ethical Issues

- Since the project involves an online meeting platform which requires access to the user's camera and microphone, informed consent should be taken from all users.
- The records of the conducted online meetings will be stored in the database until they are transcripted to text; hence, the conservation of the records is crucial to preserve data privacy.
- Providing transparency about the AI tools' usage to users is essential to keep them informed about what type of data they are providing to an AI model during distraction detection.
- The pre-trained AI models can be biased based on the data they were trained on, which may cause unfair results.

2. Design Requirements

2.1. Functional Requirements

2.1.1 Registration:

Users can register for an account on the platform. This process includes providing basic information and setting up login credentials. The registration process is designed to be straightforward yet secure.

2.1.2 Login/Logout:

Users can securely log in to access their personalized dashboard. The logout feature ensures that user sessions are securely ended, maintaining privacy and data security.

2.1.3 Data Usage Permission:

During registration or first login, users will be prompted to give consent for the use of their data, including camera feeds for analysis. This ensures compliance with data protection regulations and respects user privacy.

2.1.4 Create a Meeting:

Users have the capability to schedule new virtual meetings. This feature includes setting the date, time, and participants, along with customization options for meeting settings.

2.1.5 Join a Meeting:

Participants can join meetings through unique, secure links. The system ensures smooth entry into meetings, with features supporting different bandwidths and device types.

2.1.6 Turn On/Off Mic and Camera:

Users can control their microphone and camera during the meeting. This functionality provides flexibility, allowing participants to manage their audio and visual presence as needed. When the camera is on, the system can analyze eye contact and distraction rates to personalize the meeting summary.

2.1.7 Summarize Meeting:

The system can generate brief, accurate summaries of the meetings. For participants with their cameras on, the summarization will focus on the moments they lost focus, ensuring they catch up on missed information. If a participant turns off their camera, the system will provide a standard summary of the meeting.

2.1.8 See Previous Summaries:

Participants can view personalized meeting summaries generated based on their eye movements and distracted points. These summaries provide insights into their engagement during the meeting. Participants can also access default meeting summaries that offer a standard overview of the meeting's content. These summaries are not personalized and provide a general perspective on the meeting. Participants have the option to navigate through meeting summaries topic by topic. This functionality allows them to focus on specific areas of interest within the meeting content.

2.1.9 See Previous Feedbacks:

Users can view feedback reports that detail their performance as a host or speaker in previous meetings. These reports include information on speech quality and audience distraction rates, providing valuable insights for improvement.

2.1.10 Download Summarization/Feedback:

After a meeting, users can download the personalized summaries or feedback. This feature provides a convenient way to keep records of meetings for future reference, tailored to individual engagement levels.

2.1.11 End Meeting:

Hosts can end the meeting for all participants. This feature ensures that meetings are closed properly and all recording or summarization processes are concluded.

2.2. Non-Functional Requirements

2.2.1. Usability

The usability of the system is a primary concern, aiming to provide an effortless experience for all users, regardless of their technical background. The interface should be straightforward, guiding users through processes like registering, setting up meetings, and accessing unique features like personalized summaries. The layout and design should be clean and intuitive, minimizing confusion and potential errors. Instructions and help guides should be easily accessible to aid users in navigating through the platform. Additionally, the process for consenting to data and camera usage should be transparent and user-friendly, ensuring users feel comfortable and informed about how their data is being used**.**

2.2.2. Reliability

Reliability is a cornerstone of the system's design. It is vital that the system consistently provides accurate and timely meeting summaries, especially since these are personalized based on the participant's engagement levels. The reliability extends to the system's ability to process multiple languages and handle various

accents and dialects with high accuracy. It should be robust enough to handle technical variations in video and audio quality during meetings. The system should also have a strong uptime record, ensuring it is available whenever users need to schedule or join meetings, thereby fostering trust and dependability among the users.

2.2.3. Performance

Performance is critical, especially considering the system's functionalities. The system must be capable of processing data quickly and efficiently, providing meeting summaries shortly after the meeting's conclusion without any significant delays. The analysis of eye contact and distraction levels should be smooth and unobtrusive, not impacting the quality or flow of the meeting. The system should also be optimized for different network speeds to accommodate users with varying internet capabilities, ensuring a consistent experience across different environments.

2.2.4. Supportability

Supportability involves ensuring that the system is maintained and updated regularly, keeping up with the latest advancements in technology and user experience design. A dedicated support team should be in place to address any technical issues, user queries, or feedback. This team is responsible for regularly assessing system performance and user satisfaction, making necessary adjustments or improvements. The system should also be designed to allow for easy updates and maintenance without significant downtime or disruption to the service, thereby ensuring that it remains current and effective over time.

2.2.5. Scalability

The system needs to be scalable to manage an increasing number of users, meetings, and data processing requirements. As the user base grows and the frequency of meetings increases, the system should maintain its performance and reliability. This means having a robust infrastructure that can handle large data loads, more simultaneous meetings, and an expanding range of user demands. Scalability also involves the ability to introduce new features and expand the system’s capabilities as user needs evolve, ensuring the platform remains relevant and useful in a rapidly changing technological landscape.

3. Feasibility Discussions

3.1. Market & Competitive Analysis

As we delve into the virtual meeting platform market, we notice a distinct trend: key players like Zoom Meeting [6] and Microsoft Teams [7] are integrating meeting summarization features, but predominantly in English. This presents a market opportunity for our project, which aims to offer summarization in both English and Turkish. However, the feasibility of this bilingual approach is not without its challenges. Developing a system that can accurately process and summarize in two different languages involves complex linguistic algorithms and potentially significant investments in technology and resources. The question arises: Is it possible to develop such a system without compromising on accuracy and efficiency?

In addition to language capabilities, our project also proposes the innovative use of camera feeds to analyze participants' engagement levels for personalized summaries. This feature aims to provide summaries focused on the portions of the meeting where a participant’s attention may have lapsed. However, implementing this feature raises several questions regarding its technical and ethical feasibility. The analysis of facial expressions and distraction levels from video requires sophisticated AI technology, which is still in its nascent stages. Furthermore, there are privacy concerns with processing camera feeds, necessitating careful consideration of data protection and user consent. The fact that established platforms have not yet adopted such features indicates potential challenges in their practical application.

3.2. Academic Analysis

Academically, the integration of eye tracking and emotion detection systems in video meetings is a topic of growing interest. Current academic research in this field has made substantial progress in understanding human behavior through eye movements and facial expressions. These advancements have led to the development of sophisticated algorithms and models for accurately tracking eye movements and recognizing emotions. However, the transition from controlled

laboratory settings to the dynamic environment of video meetings presents certain challenges.

One key academic question is whether it is feasible to implement such systems for post-meeting analysis. While research has demonstrated the technical capabilities of tracking eye movements and detecting emotions, applying these technologies to recorded video meetings necessitates robust offline processing. This involves addressing challenges related to video file formats, storage, and computational resources. Academic discussions underscore that while offline analysis is attainable, optimizing these processes for various video formats and computational infrastructures encountered in video meetings is a significant consideration.

Furthermore, the academic discourse highlights privacy and ethical concerns when analyzing recorded videos. The post-meeting analysis of participants' eye movements and emotions raises questions about data privacy and consent. Researchers emphasize the need for stringent data protection measures and clear guidelines to ensure ethical usage. Additionally, there is ongoing academic debate regarding the accuracy and reliability of these systems when applied to recorded video, considering factors like variable video quality and diverse facial expressions that can impact performance.

4. References

[1] Meriem, Bouhlal & Benlahmar, Habib & Naji, Mohammed & Filali, Sanaa & Kaiss, Wijdane. (2022). Determine the Level of Concentration of Students in Real Time from their Facial Expressions. International Journal of Advanced Computer Science and Applications. 13. 10.14569/IJACSA.2022.0130119. Available: https://www.researchgate.net/publication/358334234_Determine_the_Level_of_Concentration_of_Students_in_Real_Time_from_their_Facial_Expressions. [Accessed: 16 Nov 2023]

[2] “fer,” PyPI, Jun. 08, 2023. Available: https://pypi.org/project/fer/. [Accessed: 16 Nov 2023]

[3] “Pricing”. Available: https://www.agora.io/en/pricing/. [Accessed: 16-Nov-2023].
[4] “Pricing”. Available: https://openai.com/pricing. [Accessed: 16-Nov-2023].
[5] “Introducing Llama 2”. Available: https://ai.meta.com/llama/. [Accessed: 16-Nov-2023].
[6] “Enabling Meeting Summary with AI Companion”, Zoom, Jul. 11. 2023. Available: https://support.zoom.com/hc/en/article?id=zm_kb&sysparm_article=KB0057960. [Accessed: 16-Nov-2023].
[7] ”Generate a Meeting Summary,” Microsoft, Sept. 1, 2023. Available: https://learn.microsoft.com/en-us/microsoft-sales-copilot/generate-meeting-summary. [Accessed: 16-Nov-2023].