COMSATS University Islamabad (CUI)

Project Proposal
(SCOPE DOCUMENT)

For

Speech2face

Version 1.0

Assignment-01
BCS-4-B

By

Shahzaneer Ahmed **CIIT/SP21-BCS-087/ISB**
Shayan Zameer **CIIT/ SP21-BCS-088/ISB**

Supervisor
Mr. Tehseen Riaz Abbasi

(Date of Submission: 6th November, 2022)

Bachelor of Science in Computer Science (2020-2025)

SCOPE DOCUMENT REVSION HISTORY

No.	Comment	Action

Supervisor Signature:

Date:

Contents

Abstract 1-iv
1. Introduction 1-iv
2. Problem Statement 2-iv
3. Problem Solution for the Proposed System 3-v
   3.1 Objectives 3-v
4. Vision Statement 4-v
5. Related System Analysis/Literature Review 5-v
   5.1 Table 1: Similar Systems like Speech2Face 5-vi
6. Advantages/Benefits of Proposed System 6-vii
7. Project Scope 7-vii
8. Modules 8-viii
   8.1 Module 1: Profile Management 8-viii
   8.2 Module 2: Place Voice Record 8-viii
   8.3 Module 3: Sound to Face Vector Model 8-ix
   8.4 Module 4: Face-Vector to Face-Image Model 8-ix
   8.5 Module 5: Image View Customization 8-ix
   8.6 Module 6: Features Enhancer 8-ix
   8.7 Module 7: Insight Panel 8-ix
   8.8 Module 8: Feedback Panel 8-ix
   8.9 Module 9: Help and Support 8-ix
9. System Limitations/Constraints 9-x
10. Software Process and Design Methodology 10-x
11. Data Gathering Approach 11-x
12. Learning Concepts 12-x
13. Tools and Technologies 13-xi
   Table 2: Tools and Technologies for the Speech2face. 13-xi
14. Project Stakeholders and Roles 14-xii
   Table 3: Stake Holders of Speech2Face 14-xii
15. Module based Work Division Among Team Members 15-xiii
   Table 4: Team Member Work Division the Targeted Project. 15-xiii
16. WBS and Gantt Chart 16-xiii
17. Mockups 17-xvii
18. Conclusion 18-xviii

19. References 19-xviii
19.1 Related System: 19-xviii
19.1.1 https://www.github.com/topics/sound-classification. 19-xviii
19.1.2 https://speech2face.github.io/ 19-xviii
19.1.3 Google Facenet 19-xviii
19.2 YouTube Resources: 19-xviii
19.2.1 https://www.youtube.com/watch?v=aKYlSIs3UDY&t=334s 19-xviii
20. Plagiarism Report 20-xviii
21. Scope Document Work Division 21-xviii

Project Category: (Select all the major domains of proposed project)

☐ **B-**Web Application/Web Application based Information System
☐ **C-**Problem Solving and Artificial Intelligence
☐ **E-**Smartphone Application
☐ **H-**Image Processing

Abstract

Speech2face is a web and mobile application-based software that is mainly built to recognize the face general structure, ethnicity and gender with the audio waves. Currently there is no full fledge software in the world that helps in this regard although there are deep learning libraries on which much work is done in the past. It was nearly impossible to predict the appearance of some person with his/her voice. But Artificial Intelligence has made it.
It will automate the attendance system and reinvigorate the authentication system. Moreover, it will prove itself as an asset to the security and intelligence agencies by recognizing criminals faces with the voice notes and eventually help them resolving complex cases.
Speech2face can also be used as a general-purpose software for recognizing the individuals in old audios and images where their appearance is not clear.

1. Introduction

Recognizing the facial features of a person based on their audio notes by a human being who don't actually know the person make no sense in real life. There is no such way to do something like this. Every prediction will be a wild guess with zero or no reliability at all. But with the advent in the Technology, Artificial Intelligence has made it possible to do so. The Machine Learning and Deep Learning Models can be trained on data of all type of people with some mathematical perspectives of voice that use clustering algorithms and categorize the people in different domains. By doing so, they generate predicted images.
The images will not be 100% real. But they can give a lot of insights of the data which can help us identify the person.

2. Problem Statement

There are many loop holes in the user authentications used in the security systems including banks, critical profile accounts and many more places where authenticity should be the first and foremost priority. Every other person who has the credentials of the account of that specific user can log into the accounts specified merely for the former. Secondly In most of the criminal scenes it is observed that the faces are not recognized and hence the criminals are not capture by the security forces. There is always a hindrance in the recognizing of the culprits involved in the crime scene. Mostly the voice can be seen but the faces are under the veil and thus cannot be seen directly. Market has no such software that provides such facilities at the same place as utilities.

3. Problem Solution for the Proposed System

Many Application users and Agencies have requested systems that are helpful, efficient, and reliable in user authentication. People have asked to ensure the dependability and security of critical accounts. Speech2face will provide an acceptable solution to these concerns. Security and intelligence agencies face many complications in identifying criminals as in most cases their faces are not revealed by the camera. They cannot give their visual representation to the security forces to capture them and mark them wanted. Speech2face will help us in this regard. Moreover, the old audio and video notes can be used to find insights about the person's basic information.

3.1 Objectives

O-1: Authenticate the user login system by 50%
O-2: Assist Intelligence Agencies and security forces to identify the criminals by 60%
O-3: help to retrieve the images with audio/video by 50%

4. Vision Statement

For users (generic, intelligence agencies) who want to authenticate their user systems and find a visual representation (facial view) of voices depending upon different factors, the Speech2face is an internet-based and smartphone-enabled application that will help in authenticating the users and finding real-time insights about voices unlike the data analyzed models which are not in practical usage form, the users who will use our speech2face system will be able to transform the audios into images and gain generic information about the person whose voice is under observation.

5. Related System Analysis/Literature Review

Following are some existing apps related to our project:

Sound Classification:

Audio Classification is the practice of examining audio recordings. The project classifies the sound into different categories e.g., environmental sound, the voice of speech, music sound, etc. Several AI and Data Science applications including chatbots, automatic voice translators, text to speech software can be created through this technology.

The link to their website is https://www.github.com/topics/sound-classification.

Traffic Signs Classification:

We have many kinds of traffic signs which we see in our daily life such as turning left or right, No passing heavy vehicles, indicating speed, etc. Traffic Signs classification is the process of figuring out which class a traffic sign belongs to.

The link to their website is https://github.com/topics/traffic-sign-recognition.

Speech2Face:

Speech2Face is all about guessing a person's Face using voice. In this project, we are trying to reconstruct the facial image of a person from a short audio recording of that person speaking.

The link to their website is https://speech2face.github.io/.

Facenet:

Facenet is a project developed by three researchers at Google. Facenet basically takes a face and turns it into a vector of 128 values. The project is designed to produce an embedding from a face of a person.

Table ***1*** provides the weakness of these apps and proposed solutions that would be done in this project.

5.1 Table 1: Similar Systems like Speech2Face

Application Name	Weakness	Proposed Project Solution
·Sound Classification	· There are too many sounds so There is a possibility that the system can't get sound and displays results according to to need.	· The project deals with speech recognition and converts facial images. The risk is low as compared to related projects.
Facenet	. There are a number of areas that are still left to be explored and how different ages and races play a role in face recognition.	·

·Speech2Face	· No Mobile application, only available as a website.	· The application will be available on the web as well as a mobile application.

6. Advantages/Benefits of Proposed System

- ✓ Privacy of a person is there because the resemblance is apparently not 100% right.
- ✓ This application helps the investigation team to get an idea of what the criminals look like by using their voices.
- ✓ We evaluate and numerically quantify how and in what matter reconstruction from audio resembles the true face of images of the speaker.

7. Project Scope

Speech2Face will be a Research and development-based Product with the main functionality to convert the voice into vector form and the vector form to image form and thus assisting in providing insights about the details of the person whose voice is under observation. It will be developed using different technologies and Deep Learning and machine learning techniques will be used. The Image generated can be modified in accordance with the users' perspective

General user: They can retrieve their images from their voice notes.

Security Agencies: It Can help them identify the criminals and speed up the process of resolving complex cases.

. A bird's eye view of the Project is given below:

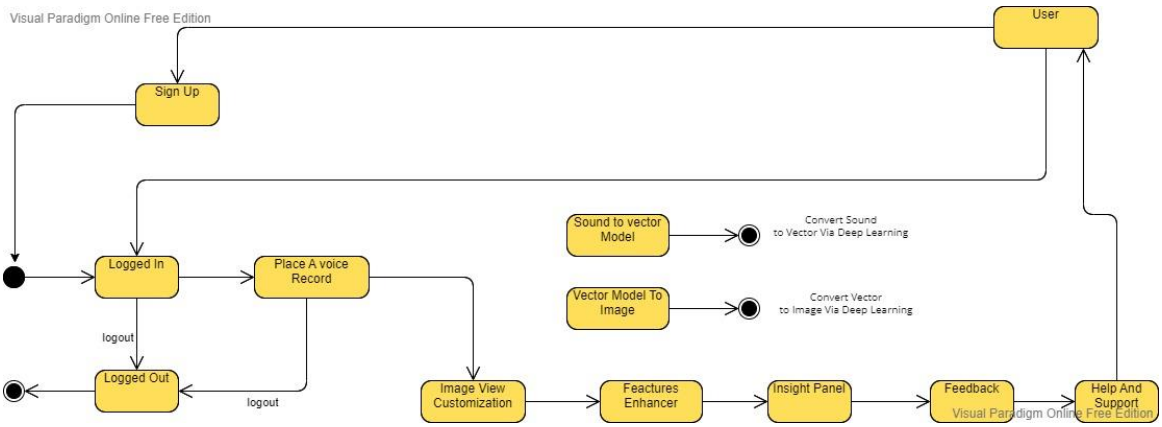


Figure 1: Context Diagram of Speech2Face

8. Modules

8.1 Module 1: Profile Management

MF-1.1 Sign Up
MF-1.2 Login
MF-1.3 Sign in Via Phone
MF-1.4 Sign in as a Guest
MF-1.5 Sign in Via Voice
MF-1.6 Update Profile Information
MF-1.7 Delete Profile
MF-1.8 Logout

8.2 Module 2: Place Voice Record

MF-2.1 Record Voice
MF-2.2 Upload Existing Voice
MF-2.3 Upload Existing Video to fetch Voice
MF-2.4 Update Voice
MF-2.5 Update Video
MF-2.6 Delete Voice
MF-2.7 Delete Video

8.3 Module 3: Sound to Face Vector Model

MF-3.1 Sound to vector Modeling via Deep Learning
MF-3.2 Generate Vector Model

8.4 Module 4: Face-Vector to Face-Image Model

MF-4.1 Vector to Image Modeling via Deep Learning
MF-4.2 Generate Image Model

8.5 Module 5: Image View Customization

MF-5.1 Brightness Control
MF-5.2 Saturation Management
MF-5.3 Skin Color Management
MF-5.4 Filters

8.6 Module 6: Features Enhancer

MF-6.1 Face Shape Enhancement
MF-6.2 Nose Enhancement
MF-6.3 Eyebrow Enhancement
MF-6.4 Beard Maker
MF-6.5 Eye Enhancement

8.7 Module 7: Insight Panel

MF-7.1 View Report
MF-7.2 Download Report
MF-7.3 Share on Socials

8.8 Module 8: Feedback Panel

MF-8.1 Rate Result
MF-8.2 Feedback in terms of words
MF-8.3 System Lagging Checks

8.9 Module 9: Help and Support

MF-9.1 Chat with AI Bot
MF-9.2 Contact Support Team
MF-9.3 Change Bots-Language

MF-9.4 View Bot's Query History

9. System Limitations/Constraints

Following are the limitations of our proposed system:

9.1 The system cannot predict the image 100% right.

9.2 The system is unable to guess some voices if it consists of the type on which the data is not trained. It is the limitation of AI.

9.3 The System can be accessed over the internet only.

10. Software Process and Design Methodology

The software process model which we will be following for this project is **Iterative Process Model.** Our application has finite number of functionalities and most of the requirements are not surely known as the project is not very common in market and it is research and development based. Therefore, the most suitable process model we could select is iterative process model so that we may go back to the requirement or design phase when the need arrives.

The design methodology we will be using is **Object Oriented Programming Approach** because it increases the reusability of the code, reduces the complexity of the code, and it would be easier for the team members to work together without any confusion as it is the most followed programming paradigm in the market.

Therefore, we are using Dart for Flutter (Mobile-end) and JavaScript (Web-end) for Web which follows object-oriented approach. Also, it is easier to describe the code using Unified Modeling language diagrams.

11. Data Gathering Approach

The data gathering for requirement analysis adopted for the proposed system includes surveys and interviews with the general users, intelligence agencies, and security forces.
However, the data set which was trained was obtained from the Open Source Kaggle platform which consists of voice interpretation insights of different peoples of the world with their corresponding facial features.

12. Learning Concepts

The concepts that we would be learning are

Concept-1: UI/UX Designing via Figma:
firstly, we will develop the UI of the system for both web and mobile interfaces.

Concept-2: Flutter for Mobile App Development:
For Mobile-end, we are using Flutter SDK. We will be learning to change the UI in Flutter code using Dart.

Concept-3: HTML, CSS, and JS for Web Development:
For Web-end, we are going to use vanilla HTML for web structuring, CSS for design, and JS for client-side scripting

Concept-4: Firebase for user Authentication and Data Storage:
We will also implement backend using firebase for authentication and data storage using firestore database and firebase Auth package.

Concept-5: Data Analysis using Numpy, Pandas, and Matplotlib:
For the initial Data analysis of the data set we will use the power of python packages for statistical computations and visualization of the data.

Concept-6: Machine Learning and Deep Learning using Tensorflow and Keras:
The data set will be trained by using machine learning techniques using Python libraries for Data modeling.

13. Tools and Technologies

The tools and technologies that we will be using are provided in the ***Table 2*** below. The main IDE tool to develop the code will be **Microsoft Visual Studio Code 2022** on which different Technologies like **Flutter, HTML-CSS** and **Java Script** will be implemented. **Figma** will be used for mockup creation. **Microsoft Word** and **PowerPoint** will be used for documentation and presentation. **Firebase** Firestore will be used as the backend database.

Table 2: Tools and Technologies for the Speech2face.

	Tools	Version	Rationale
Tools	Visual Studio Code	2022	IDE
	MS Word	2021	Documentation
	MS Power Point	2021	Presentation
	Figma	2022	Mockups Creation
	Flutter	3.3	SDK

Technologies	Git	4.3	VCS
	Netlify		Deployment
	Technology	**Version**	**Rationale**
	Dart	3.5	Client-side Scripting
	Firebase	5	Ready Made Backend
	Html	5	Web Structuring
	CSS	3	Web Design
	Tailwind CSS	3.0	Web Design
	Bootstrap	5	Built-in Web Components
	JavaScript	ECMA Script 2017	Event Based Programming language
	Python	3.8	ML/DL Modeling

14. Project Stakeholders and Roles

The stakeholders of the project comprise of both the Software Engineers working on the project from the development and documentation side. The project will be developed in the supervision of **Mr. Tehseen Riaz Abbasi** and for the evaluation of the project, there will be a Final Evaluation Project Committee. The key roles of the stakeholders are:

- **Shahzaneer Ahmed:** Responsible for Development, Documentation, and Design.
- **Shayan Zameer:** Responsible for Development, Design and Testing.

Table 3: Stake Holders of Speech2Face

Project Sponsor	RaviCom Solutions Private Limited.
Stakeholder	• Shahzaneer Ahmed • Shayan Zameer • Project Supervisor Name: Mr. Tehseen Riaz Abbasi. • Final Year Project Committee: Evaluation of project.

15. Module based Work Division Among Team Members

Table 4: Team Member Work Division the Targeted Project.

Name	Registration Number	Responsibility/ Modules
• **Shahzaneer Ahmed**	• **SP21-BCS-087**	• **(Module1-Module3-Module6-Module7)** Mobile App, Documentation, Backend, ML/DL
• **Shayan Zameer**	• **SP21-BCS-088**	• **(Module2-Module4-Module5-Module8)** Web A, Documentation, Artificial Intelligence, Deign

16. WBS and Gantt Chart

Table 5: WBS for software development.

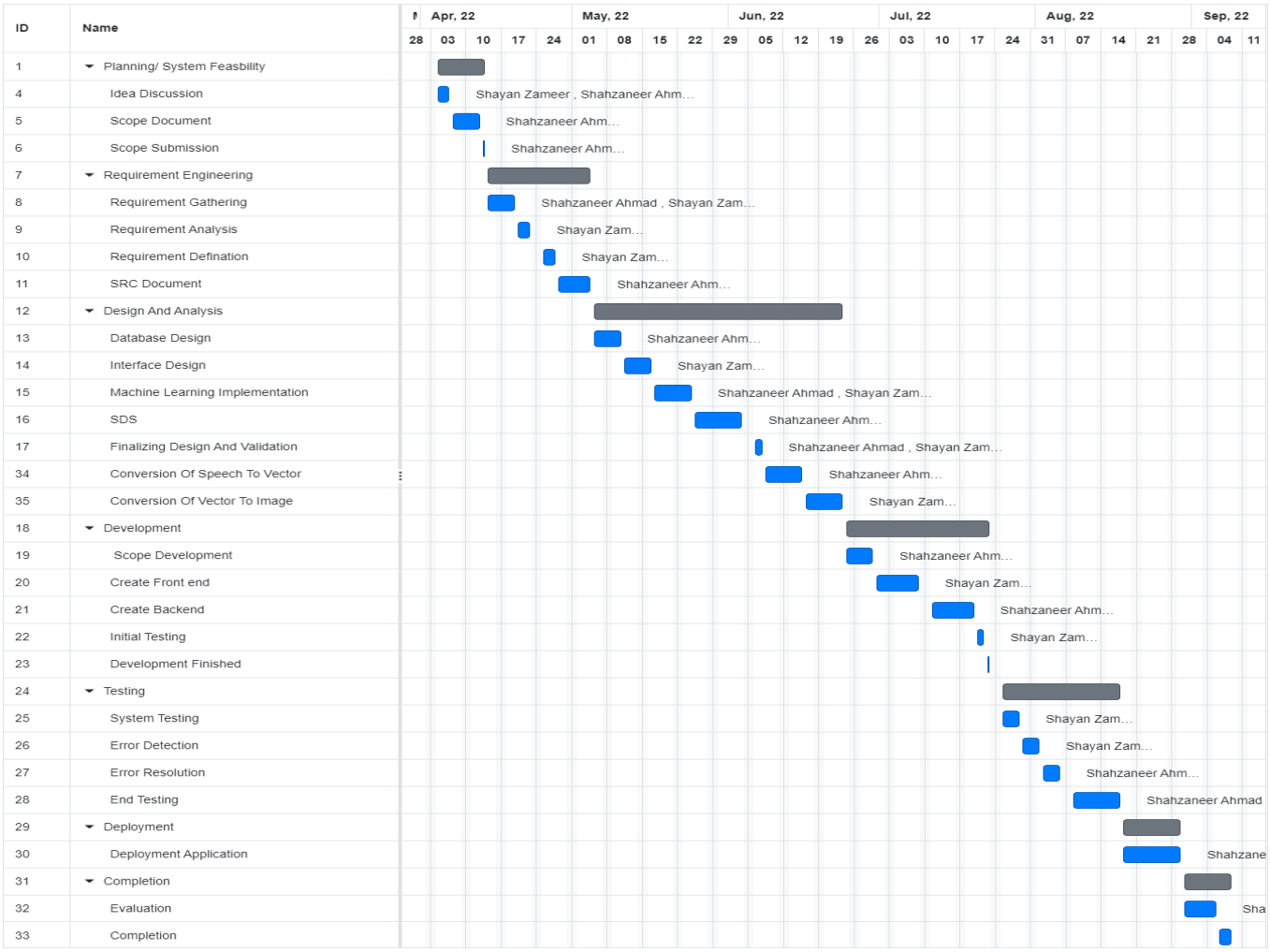
ID		Task	Duration	Resources
1	**Planning/System Feasibility**		**8 d**	**Shahzaneer , Shayan**
2		Idea Discussion	4 d	Shahzaneer, Shayan
3		Scope Document	3 d	Shahzaneer
4		Scope Submission	1 d	Shahzaneer
5	**Requirement Engineering**		**22 d**	**Shahzaneer , Shayan**
6		Requirement Gathering	7 d	Shahzaneer , Shayan

7		Requirement Analysis	3 d	Shayan
8		Requirement Definition	5 d	Shayan
9		SRC Document	7 d	Shahzaneer
10	**Design And Analysis**		**45 d**	**Shahzaneer , Shayan**
11		Database Design	12 d	Shahzaneer
12		Interface Design	7 d	Shayan
13		Machine Learning Implementation	5 d	Shahzaneer , Shayan
14		SDS	1 d	Shahzaneer
15		Finalizing Design And Validation	8 d	Shahzaneer
16		Conversion Of Speech To Vector	6 d	Shahzaneer
17		Features Enhancer	6 d	Shayan
18	**Development**		**15 d**	**Shahzaneer , Shayan**

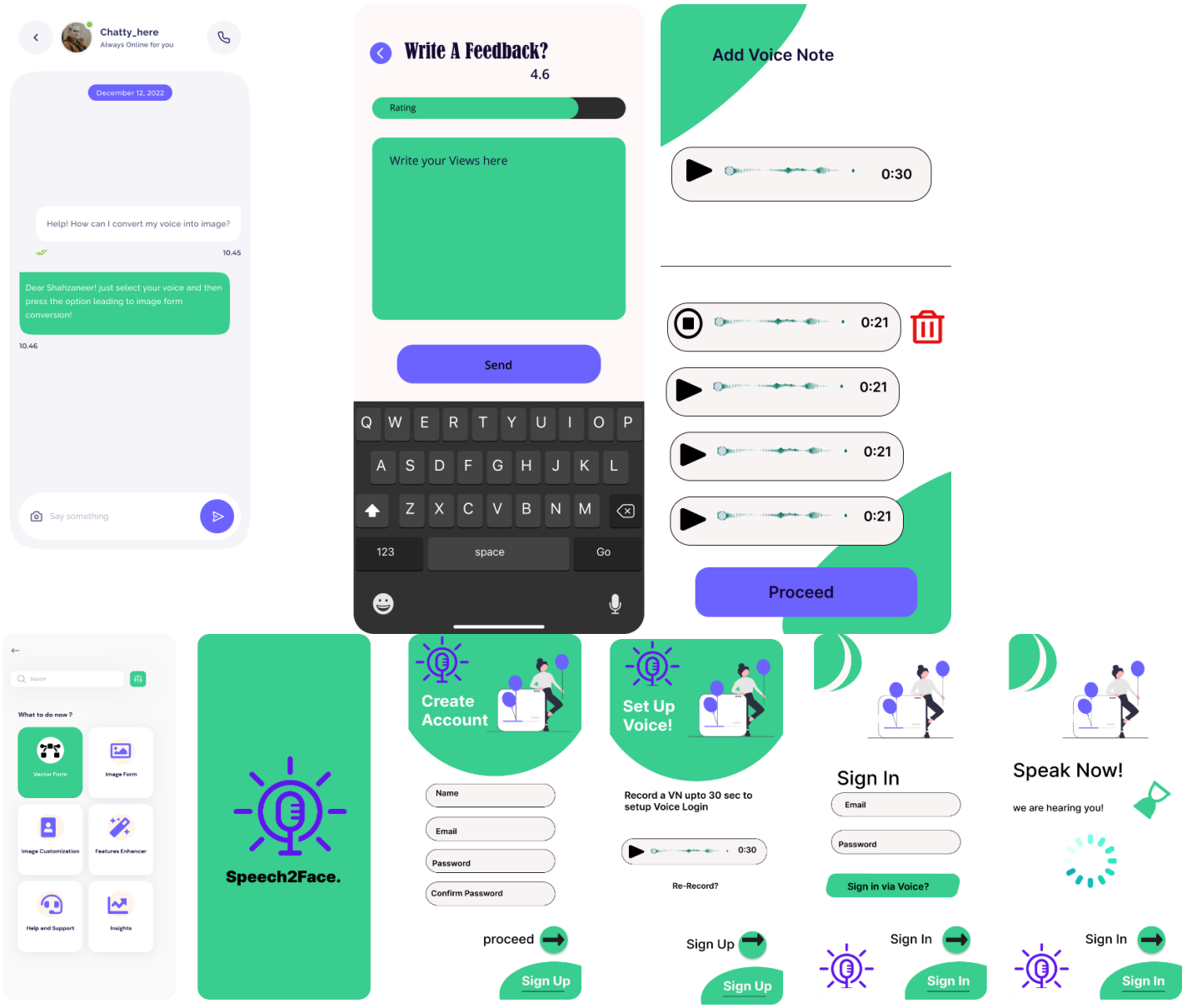
19		Scope Development	1 d	Shahzaneer
20		Create Frontend	15 d	Shayan
21		Create Backend	6 d	Shahzaneer
22		Initial Testing	1 d	Shayan
23		Development Finished		
24		**Testing**	**8d**	**Shahzaneer,Shayan**
25		System Testing	2d	Shayan
26		Error Detection	1d	Shayan
27		Error Resolution	3d	Shahzaneer
28		Error Evaluation	2d	Shahzaneer
29		**Deployment**	**1d**	**Shahzaneer**
29		Deployment Application	1d	**Shahzaneer**
30		**Completion**	**3d**	**Shayan, Shahzaneer**

31		Evaluation	3d	Shayan, Shahzaneer
32		Completion		

Figure 2: Gant Chart for software development



17. Mockups


Chatty_here
Always Online for you
December 12, 2022
Help! How can I convert my voice into image?
10.45
Dear Shahzaneer! just select your voice and then press the option leading to image form conversion!
10.46
Say something
Write A Feedback?
4.6
Rating
Write your Views here
Send
Add Voice Note
0:30
0:21
0:21
0:21
0:21
Proceed
Speech2Face.
Create Account
Name
Email
Password
Confirm Password
proceed
Sign Up
Set Up Voice!
Record a VN upto 30 sec to setup Voice Login
0:30
Re-Record?
Sign Up
Sign In
Email
Password
Sign in via Voice?
Sign In
Speak Now!
we are hearing you!
Sign In

18. Conclusion

The Speech2Face's ultimate objective is to help users to authenticate their logins and sign-ups. It helps to retrieve images based on voices firstly in vector form and then by converting them into image form. It also helps intelligence and security agencies to identify the culprits and criminals on the basis of their voices. It won't generate real images but is very helpful in generating specific insights.

Nonetheless, this project will serve as a solid foundation for getting acuities via speech and it will give an amazing experience of authentications for the users.

19. References

19.1 Related System:

19.1.1 https://www.github.com/topics/sound-classification.

19.1.2 https://speech2face.github.io/

19.1.3 Google Facenet

19.2 YouTube Resources:

19.2.1 https://www.youtube.com/watch?v=aKYlSIs3UDY&t=334s

20. Plagiarism Report

- N/A

21. Scope Document Work Division

Table 6: Work Division

Shahzaneer Ahmed (SP21-BCS-087)	Shayan Zameer (SP21-BCS-088)

1. Abstract 2. Introduction 3. Problem Statement 4. Problem Solution for the proposed system and Objectives 5. Vision Statement 6. Modules 7. System Limitations/Constraints 8. WBS and Gantt Chart 9. Conclusion 10. Final Format Preparation of Scope Document (Word .docx file)	1. Related System Analysis/Literature Review 2. Advantages/Benefits of Proposed System 3. Project Scope 4. Software Process and Design Methodology 5. Tools and Technologies 6. Project Stakeholders and Roles 7. Work Division 8. Concepts 9. Mockups 10. Presentation (PowerPoint .ppt file)