FACULTY OF SCIENCE

SCHOOL OF COMPUTER SCIENCE

2020-2021

A

PROJECT REPORT

ON

AI BASED AUDIO DETECTION [SYSTEM]

BY

IN PARTIAL FULFILLMENT OF

BACHELOR OF SCIENCE (COMPUTER SCIENCE)

MIT WORLD PEACE UNIVERSITY

SCHOOL OF COMPUTER SCIENCE

Certificate

This is to certify that, **GAURAV SURYAWANSHI, SHIVAM SATAV, OMKAR HUNDEKARI &AVISHKAR PACHPUTE** students of B.Sc.(Computer Science) Trimester has successfully / partially completed **MINI PROJECT** at **MITWPU** in partial fulfilment of B.Sc. Computer Science under MIT World Peace University, for the academic year 2020-2021.

_______________ __________________

Internal Guide Head of the School

Date: _______________

External Examiners:

1. ______________________________ 2. ______________________________

ACKNOWLEDGEMENT:

DECLARATION:

Pune **Signature of the Candidate**

Date:

INDEX

INTRODUCTION

Existing System

- Detecting Audio with highly inaccurate outputs.
- Device hears and recognizes sound with Reality AI.
- To build public awareness about the risks of AI-powered speech synthesis, a few months ago we shared example of the technology with the public. Using a proprietary speech synthesis model, they built called Real Talk, engineers at the company recreated the voice of the popular podcaster Joe Rogan.
- Example, the world's first AI-powered Cyber Crime reported on earlier this September. Using speech synthesis technology, thieves were able to convince an energy executive into thinking he was on the phone with his parent company's CEO, tricking him into wiring over $250,000 into their account.

Problem Definition: Need of Computerization

Inaccuracies can be reduced, we can't promise 100% accurate results but in this proposed system, we are trying to achieve close results. As machine learning practitioners, we have the capabilities to do this, and can help mitigate a real-world problem with drastic consequences.

To differentiate between real and fake audio, the detector uses visual representations of audio clips called spectrograms, which are also used to train speech synthesis models. You can read more about how spectrograms help create synthesized audio in our technical post on speech synthesis.

PROPOSED SYSTEM

Proposed System

- Current system will be using online internet connectivity to generate the audio samples.
- Current system will not have any kind of upload option to use pre-recorded models.

Objectives

- To achieve the difference as the output of voice recognition, for eg if a class of 20 students have their audio samples we can differentiate each person based upon their audio
- To minimize the inaccuracies and achieve close to 90-95% precise results.

User Requirements

- Internet Connection (stable Broadband or 4G)
- Internet browser (Google Chrome 11 /Mozilla Firefox 10)
- His/Her Login credentials

Operating Environment – Software & Hardware

Software :

- Chrome 65 or greater
- Google Audio Engine
- HTML, CSS, JS (Frontend)
- Python (Backend)

Hardware :

- Processor – Dual Core
- Hard Disk – 15 GB
- Memory – 4 GB RAM
- Display: 1.33" TFT or any SD Display Screen

ANALYSIS & DESIGN

Module List

This system is having 3 Modules:

1. Feeding the audio sample
2. Recognising the audio
3. Displaying the output

Description:

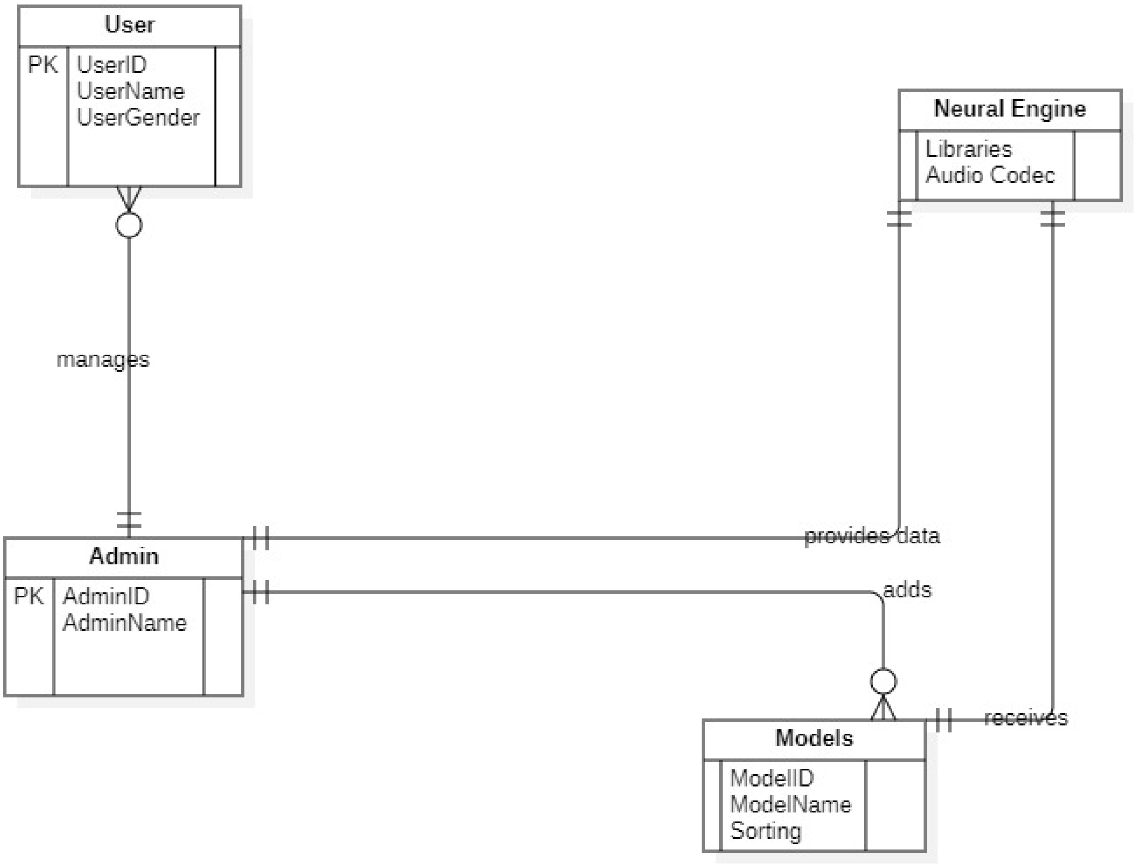
1. Feeding the audio sample:
- Here, the input is taken directly from the user mic or pre-recorded
audio samples from the user.
- Added audio is then listed in the system.

2. Recognising the audio:
- The audio is recognized using Neural Engine.
- The audio is sent directly to the backend call of Google API.
- The audio is then processed into the final format of wav or mp3.

3. Displaying the output:
- System will analyze the audio data from the user class and display
the output.

SYSTEM DIAGRAMS

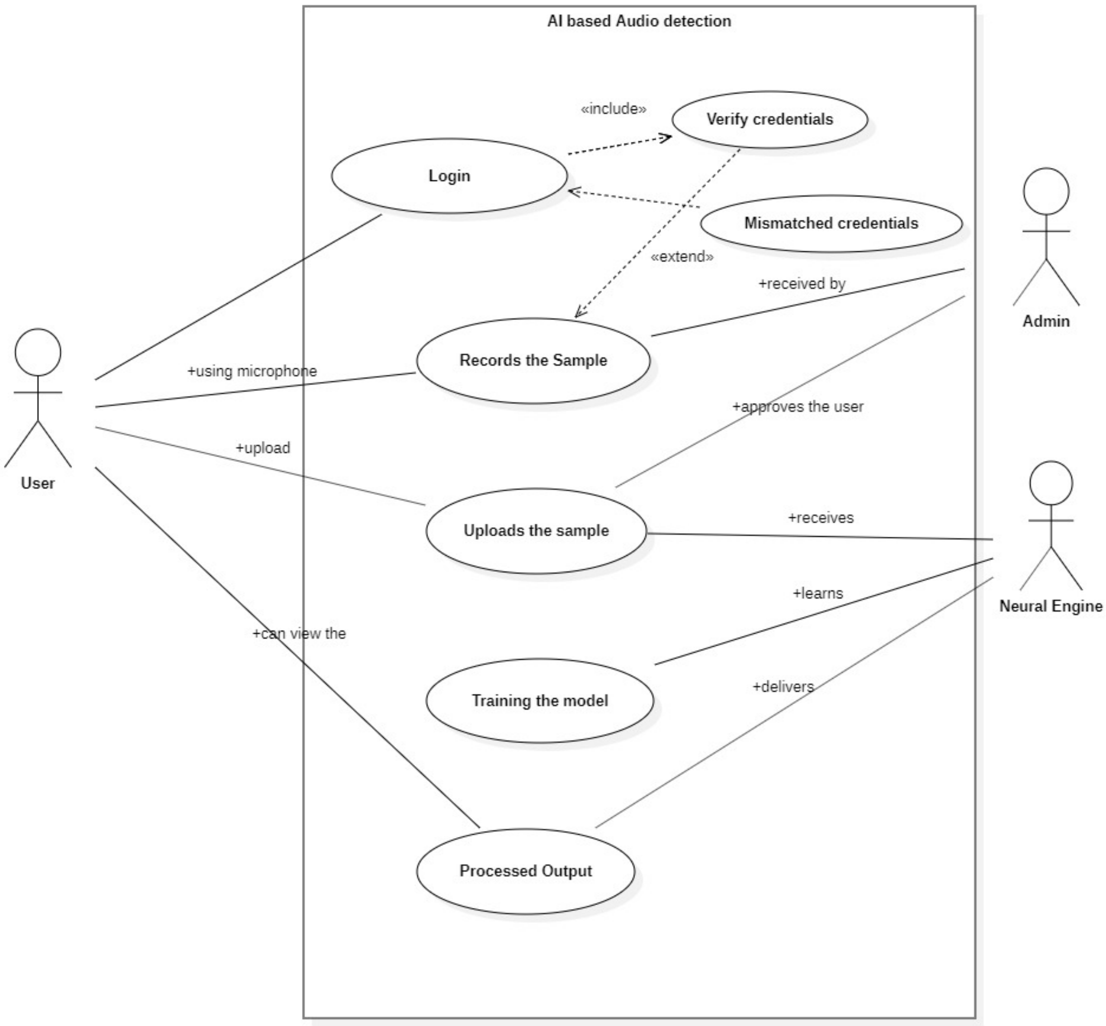
- ## ER Diagram



(Entity-Relationship Diagram)

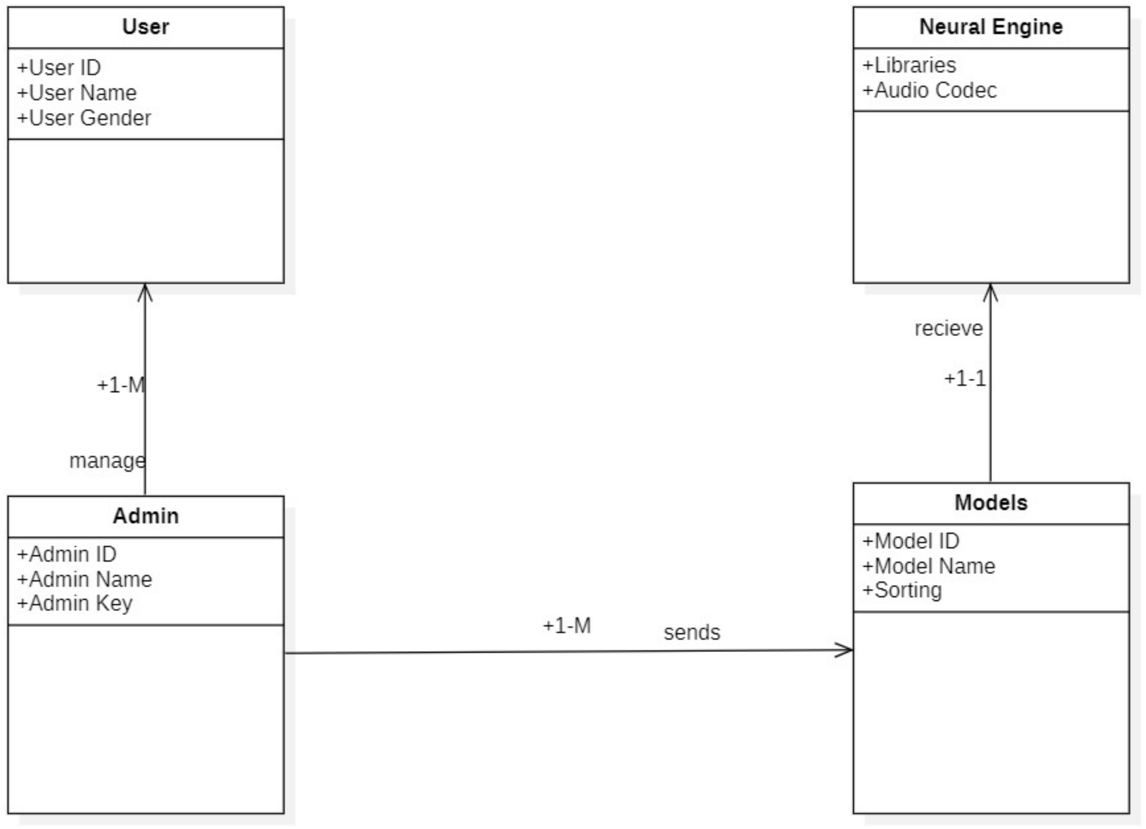
- **UML Diagrams**

Use-Case Diagram



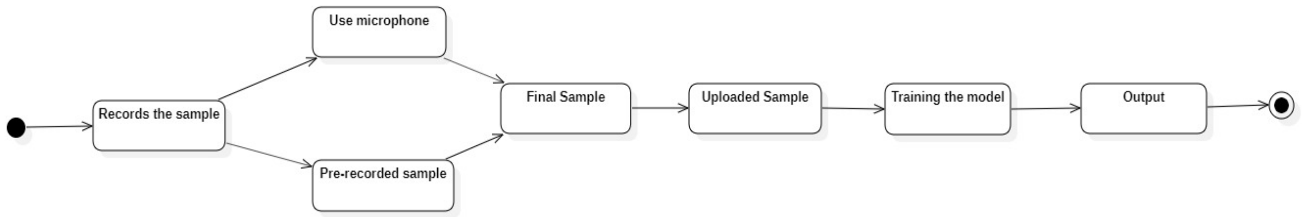
(Use-Case Diagram)

Class Diagram



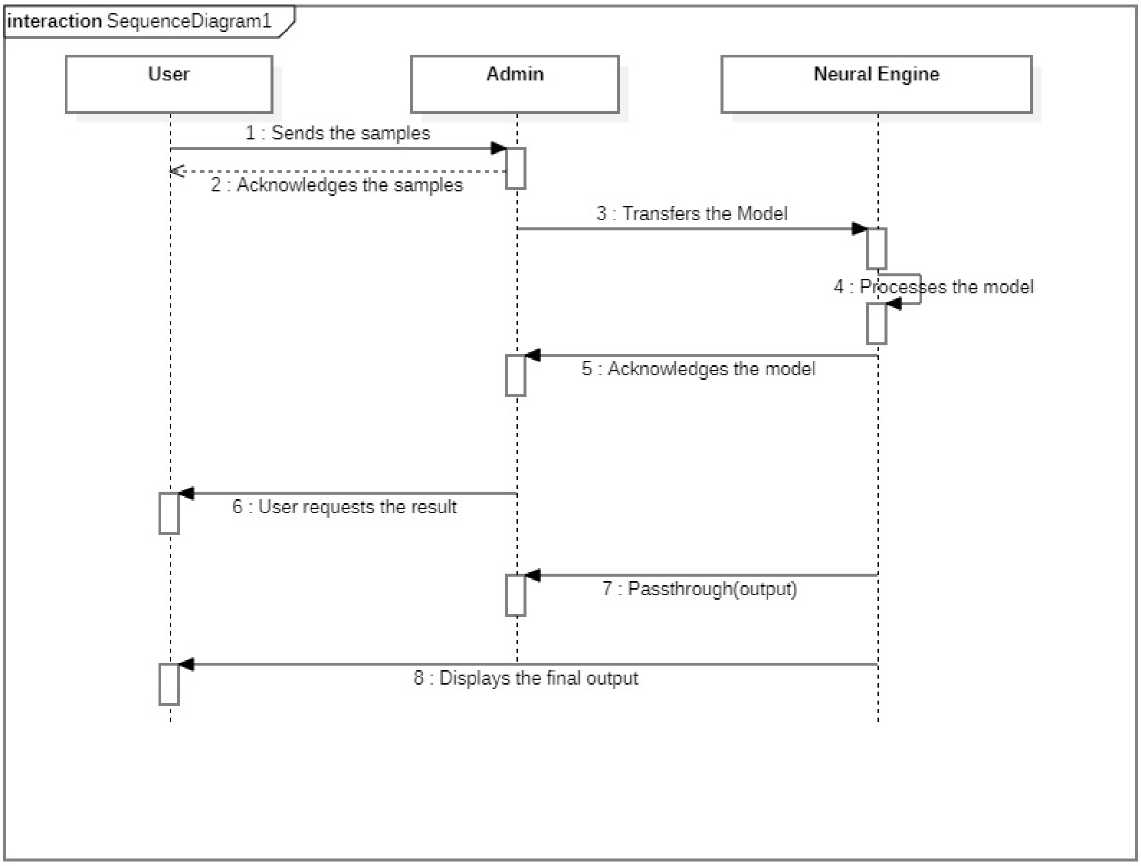
(Class Diagram)

State-chart Diagram



(State-chart Diagram)

Sequence Diagram



(Sequence Diagram)

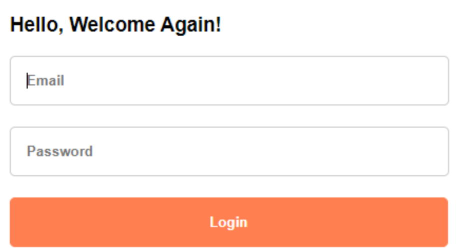
DATA DICTIONARY (TABLE)

Field Name	Data Type	Field Length	Constraint	Description
User_id	Int	10	Primary Key	User id,Auto generated
User_name	Varchar	30	Not Null	Name of User
Gender	Varchar	10	Not Null	Gender of an User
User_password	varchar	30	Not Null	Login Password for User
User_mail_id	Varchar	30	Not Null	Any email_id
Admin_id	Int	10	Primary Key	Admin id,Auto generated
Admin_name	Varchar	30	Not Null	Name of Admin
Admin_password	Varchar	30	Not Null	Login Password for Admin
Admin_mail_id	Varchar	30	Not Null	Any email_id
Codec_id	Int	10	Not Null	Audio sample id

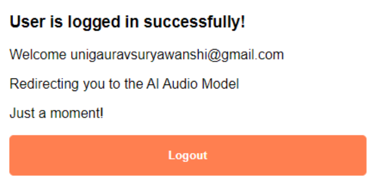
Above, is the data dictionary of our system.

SCREENSHOTS OF SYSTEM DESIGN (USER INTERFACE)

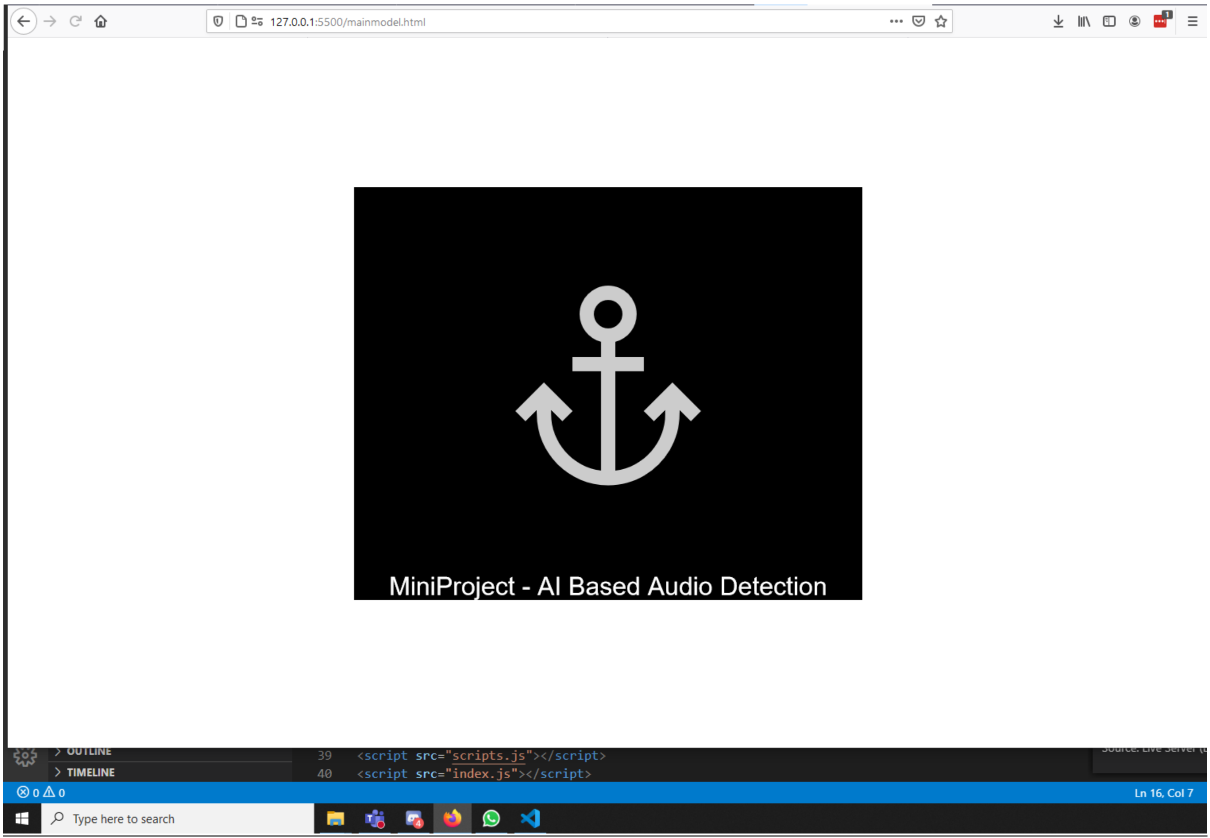
<u>Screen 1 (Login)</u>



<u>Screen 2 (Redirection Screen)</u>



Screen 3 (Output Screen)



USER MANUAL

1. Version History

Version	Time	Modify content
2.0	2020.12.12	Future Release
1.0	2020.08.08	Initial Release

2. Software & Hardware Resources Used :

Software Requirements :

- Chrome 65 or greater
- Google Audio Engine
- HTML, CSS, JS (Frontend)
- Python (Backend)

Hardware Requirements :

- Processor – Dual Core
- Hard Disk – 15 GB
- Memory – 4 GB RAM
- Display: 1.33" TFT or any SD Display Screen

3. Instructions for using the system :

I - The user logins with the provided credentials to him by the admin.	
II – The user can select which model he/she wants to determine voices.	
III – Allow access to your microphone and close the permission pop-up.	
IV – Now using the microphone, the user can speak directly and as per the sampling output will be shown.	

4. Q&A

- Q: How can I update the software ?

A: Please go to the about section, select direct update option from the web-based application.

- Q: What should I pay attention to, while using the application?

A: Check for epoch count and variation of audio samples, if you are working with a new audio model.

- Q: Can I use this software in offline mode ?

A: Unfortunately, the modules can only be used with internet connectivity for now, but we are planning for offline deployment in upcoming updates, stay tuned !

CONCLUSION

Future Enhancements

In the Version 2.0, which will be available in the near future, we are planning to implement following features:

1.C# Modules to implement the source code as a Windows based application.

2.Making online payloads available without any internet connection.

3.Fixing bugs and adding functionalities with small patches and updates.

Disadvantages

Drawbacks in the existing system

1.User requires an internet connectivity to interact and receive output from the system.

2.44k Frequency microphone is needed to record audio samples.

3.No option to upload pre-made audio samples by the user as it requires manual approval from the admin.

Final conclusion

From the following, we can derive that the model is still in development phase and will be available as a fully-fledged web-based application to the public.

Also, the current prototype works with a 90% accuracy in recognizing the audio once the model is pre-fetched and made available in the system.