

Develop Communication using Virtual Reality and Machine Learning

Deepak Chouhan¹, T Aryan Raju², Dr. Rajesh S. Prasad³

^{1,2}Student, Dept. of Computer Science & Engineering MIT Art, Design and Technology University, Loni Kalbhori Maharashtra Pune, India

³Professor, Dept. of Computer Science & Engineering MIT Art, Design and Technology University, Loni Kalbhori Maharashtra Pune, India

Abstract—Nowadays, a lack of communication skills is a prevalent issue. Around 75% of the world's population lacks basic communication skills. As a result, they are unable to give a good speech, provide a good presentation, or perform well in interviews. According to surveys, most people want to enhance their communication abilities but don't have somebody to practise with or listen to them. That's where our solution comes in. By raising a user's vocabulary, word knowledge, pronunciations, and fluency, AI-buddy aims to improve a user's command and control over the language.

Keywords—3DVR, VR, Communication skills, Machine Learning, Deep Learning, Personality development.

I. INTRODUCTION

A. Purpose of this project

This project is a website-based language learning and personality development platform which focuses on overall improvement on the user's ability to speak with confidence in public and also get to know the language in depth. It provides a competitive platform where users develop their communication skill by practicing with our Machine Learning model and can experience a stage like experience standing a home. users can monitor their progress, work to improve, learn new words etc. This increases the user's confidence in their ability to face an interview, speak in front of a group, or speak confidently without stumbling. We additionally give important review material from the best sources with the goal that your insight never gets stale and you continue to learn. Goal and Vision

This system aims to provide users an environment where they can play and learn to deliver speech/presentation in public. This system consists of various features, like new words and tongue twister to exercise their tongue and 3D VR environment to give them an emerging experience of stage at home with the help of VR Headset and provide them an AI helper to help them with their speech.

II. LITERATURE REVIEW

A Study of Virtual Reality

In 2017, a paper was published on virtual reality. It gives an in-depth study of the Virtual reality, it talks about everything required to implement Virtual reality in technology. But the only limitation is that it talks about implementing the virtual reality only in PC environment

not in the mobile environment, which is necessary in current era.

Virtual reality (VR) is a strong and intelligent innovation that changes our life dissimilar to some other. Computer generated reality, which can likewise be named as vivid media, is the craft of mimicking an actual presence for the crowd in places both genuine and nonexistent. It as a rule includes two faculties specifically sight and sound. The key property that recognized VR from all past media types is "presence".

Virtual reality in education: The promise, progress, and challenge

This paper was published in year 2020, it talks about benefits and implementation of Virtual reality in modern education system, how it can change the education forever, and the challenges which would come in implementing VR based learning.

Virtual Reality (vr) is an instrument that numerous language teachers have considered for use in their study halls. Be that as it may, since vr is a constantly advancing innovation with rapidly evolving elements, applications, and instructive affordances, it is hard for educators to see precisely what the innovation is and what it can bring to their study halls. VR is a clever innovation for some, however its predominance at amusement parks and attractions have started somewhat to decrease its persona. Regardless the improving probability that understudies have sooner or later encountered a type of vr conceivably decreasing the oddity factor, vr in training has reliably been viewed as persuading and invigorating for understudies.

III. REQUIREMENT SPECIFICATION

A. User characteristics

There is only one class of user that interacts with the website and i.e., individuals who connect with our web application since no outsider applications are being utilized in this project. The user interacts with the web application using a web browser. The website provides different elements and features for the user to interact with. The web application mainly targets the people who want to improve on their basic understanding in English language. As a result, it is primarily directed at students and job seekers aged 8 to 30. The web application's secondary target are people above the age of 30 who may not have the professional requirement to improve their strength over the language but are looking for a

platform to improve their speaking skills for various reasons like fear to speak, improve basic interaction fluency etc. The user needs to have the basic knowledge of how a web application works and a little knowledge of English.

B. Functional Requirements

Functional requirements 1

Actor: User

Input: String and base64 encoded audio

Description: The API processes base64 encoded audio into an audio file and then processes that audio file into text and compares it with the string to find accuracy. The audio is also passed through an emotion detection algorithm in the API and The emotional value, accuracy status, audio converted to text, and the string are all returned by the API.

Functional requirement 2

Actor: User

Input: VR

Description: User can put on a VR cardboard on his/her mobile and practice speaking in a 3D VR environment

C. Dependencies

There are two main web frameworks used in the construction of the web application: -

1) Node.JS:- A java Script based web framework used for making the backend of the website

2) Django:- A python-based web framework use to make the main API for the website

Apart from the frameworks the main front-end tools used for the website are HTML, CSS and JS

There are a bunch of libraries also used in the website mainly for tasks related to machine learning, natural language processing, data manipulation like: -

- NumPy
- TensorFlow
- Django-rest framework
- Pandas
- Scikit learn
- Speech recognition
- Librosa
- Matplotlib
- Keras

D. Hardware requirements

CLIENT SIDE

To run the web application user needs to have a functional device like mobile\desktop\laptop\tablet etc.

with an internet connection and an upgraded web browser like

Desktop browser:- Safari, Chrome, Firefox, Opera etc.

Mobile browser:- Android, Chrome mobile, iOS safari etc.

An additional feature requires a cardboard VR but even without that user can take care of all other requirements of the website.

SERVER SIDE

- Any operating system
- At least 2 GB of RAM
- Python compiler installed
- Node and express and Django installed
- Python libraries installed mentioned in dependencies

IV. PROPOSED SYSTEM

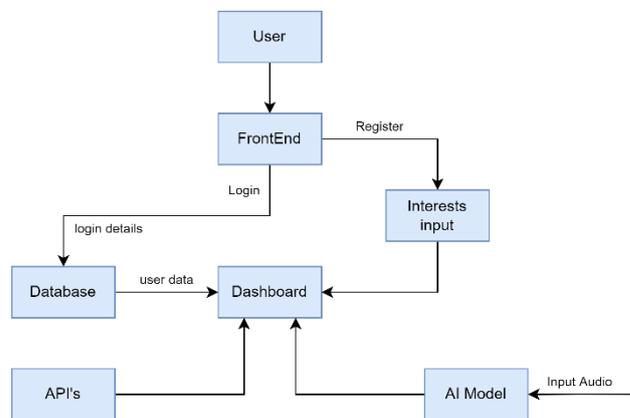


Fig 1. Working of the website

The above diagram shows the brief working of our website, this includes Frontend, Backend, database, API, and machine learning models.

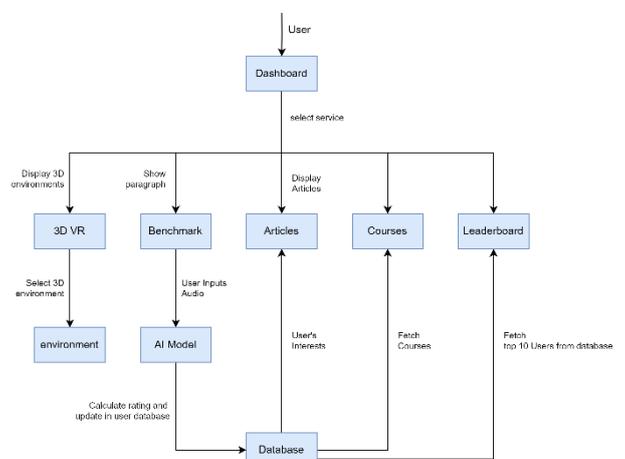


Fig 2. In-depth working of the server

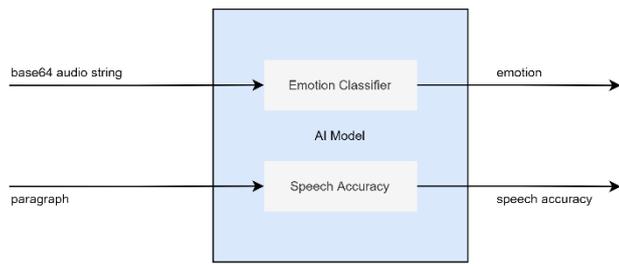


Fig 3. Speech Analyzer machine learning model

The Machine learning model consist of 2 features which are

1. Emotion Classifier
2. Speech Accuracy analyzer

A. Emotion Classifier

The purpose of this model is to classify emotion from the audio file given by user. this model is created using Deep neural network, we use Sequential model API to create neural network of our model to predict emotion in audio file. The Sequential model API is a way of creating deep learning models where an instance of the Sequential class is created and model layers are created and added to it.

The emotion classifier divides the feeling into five groups.

- Calm
- Happy
- Sad
- Angry
- Fearful

The steps involved for predicting the emotion from audio:-

- The Model takes audio file as an input and with the help of the python library "Librosa".
- We extract the features, then we reshape the features by [1, 40, 1], so that it matches with
- The machine learning model input type.
- After reshaping it, we feed the data to our machine learning model.
- In return the model gives us the array of number which we need to decode the help of "Label Encoder".
- It has a function named "inverse transform" that was created expressly for this task. We'll be able to decode it once we've decrypted it.
- Once we have decoded the array of number, we will get the exact emotion of Audio speech given by user.

B. Speech Accuracy analyzer

The accuracy detecting algorithm used in the website is called cosine similarity methodology. This is a very technically efficient method of determining how similar

two sentences are. In our website the user passes the audio file in which he/she is reciting the sentence that is being shown to him/her. The audio is then converted to base64 format and sent to the API via backend for processing. The speech accuracy analyzer detects the similarity of the cosine of two sentences, in this case the accuracy of the speech to the original sentence.

Cosine similarity in short refers to finding the cosine of the angle formed between 2 non-zero vectors. It is a very accurate method to find out in the scale of 0-1 how much similar the two non-zero vectors are.

Steps involved for finding accuracy

- Firstly, the API receives the audio in base64 format and the string.
- The base64 format is then converted to audio and saved temporarily in a file in .mkv format.
- The speech to text library named Speech Recognition converts the mkv formatted file into text and saves it in a variable named 'text'.
- The text and sting are taken in a single array variable named 'sentences. Example : sentences = [text, string], Where let text = "Hello I am dcdepakdc", And string = "Hello I am dcdepakdc".
- The two sentences are then converted to vector format using different functions from premade libraries like CountVectorizer and Vectorizer and is then stored in an a 2d array variable named array
- The vector arrays are then passed through the cosine similarity function which makes a 2d matrix which symbolizes the correlation between 2 vectors.
- From the correlation we can find the similarity of both the paragraphs

C. Interest based articles

For user to develop the habit of reading, we decided to recommend user articles based on his liking, if user like "business" then we will recommend articles on business. We are using an Api called "gnews". Using this Api, we will fetch the articles based on users liking, which will contribute in developing reading habit of user.

Steps Involved: -

- Fetch the interests from user object data.
- Create a new URL with query as a user interest, then send the request to Api using axios library.
- The Api will return the Articles as our response, then we will post the data to frontend for displaying.
- Display the data to web page

D. Scheduled rank update

In order to keep the leaderboard fresh with the top user, we are using scheduled rank update system, which will update the rank after specific interval of time and keep the leaderboard fresh with top users. We fetch the all the user from the database, then we sort all the user according to their rating in ascending order. Then we loop though all the user and give them the rank.

V. END USER INTERACTION

This project provides all of its user's lot of features to improve their communication skill.

A. 3D VR

User can use our 3D VR feature to get the experience of realistic stage or any other public hall but just standing home. The only thing user need is a VR headset, then he/she can navigate to 3D VR page from the dashboard and click on any type of environment he/she desire. It will be presented in front of him/her with just one click.

B. AI Critique

This feature allow user to get rating on his/her speech. To use this feature user can navigate to Benchmark page from the dashboard. Then a paragraph will be shown to user, user needs to record his/her voice reading that paragraph using the Audio recorder in his/her device. Once recorded, user needs to upload that file to the website. After the file has been uploaded to the website, the AI model will do the preprocessing of the Audio file and return a score on the speech recorded by the user. And it will be reflected to his/her rating.

C. New Words and Tongue twisters

There are times when you get stuck using some words. That is what this feature solves, this feature help you exercise your tongue. To use this feature, user can navigate to dashboard, their user can find both new word and tongue twister, he/she can practice that tongue twister or if he/she need more then click on the more button down there.

D. Courses and Articles

This feature allow user to learn new skill and topic which user is interested in with the help of articles and courses. User will be show articles based on his interests using the gnews Api, and also the top courses from Udemy database will be shown to user which he/she can enroll in to learn.

The courses are divided in different categories

- Leadership
- Communication skill
- Time Management
- Social Skill
- Public Speaking
- Emotional intelligence

VI. RESULTS

This sections show the result and accuracy different machine learning model in this project.

A. Emotion classifier

The Emotion classifier is the machine learning model, in which we pass the audio file as a parameter and the emotions are classified based on it. The accuracy data is

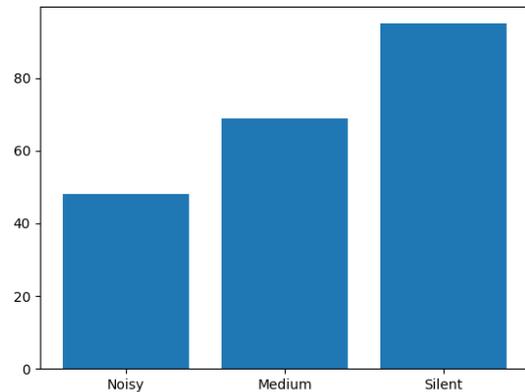
provided in Table 1. It displays the accuracy of several emotions that were identified using a sample dataset.

Table 1. Accuracy of different emotions

Emotions	Test size	Accuracy
Angry	50	79%
Calm	50	81%
Happy	50	78%
Fearful	50	68%
Sad	50	76%

B. Speech Accuracy analyzer

The Speech accuracy detector, on the other hand, compares the real paragraph with the speech provided by the user. We put this model to the test in a variety of environments, including noisy and silent environments. This model has the highest accuracy of 98%.



In a noisy environment, speech accuracy can be as low as 20%, while the average accuracy was around 44%. The accuracy ranged from 60% to 80% in a semi-noisy environment with low noise, and it might reach 98 percent in a completely calm environment.

VII. FUTURE IMPROVEMENTS

As this is the initial stage of this project, their are some are in which it is lacking. Some of the possible improvements which can be done in this project are:-

A. AI Critique enhancements

Currently there are only two parameters on which we are rating the user speech. But in future AI critique can have more enhancements, where more parameters will be considered such as Filler word and the overall voice modulation etc. when judging your speech. This will allow us to rate the user with more accuracy.

B. Freely Speaking feature

For now, we display a paragraph in front of user, which he/she need to read and record his/her voice and upload it to the website and the rating is give based on that

speech. However, in the future, we may be able to create a feature that allows a user to talk freely rather than reading a random paragraph and receive real-time feedback.

C. 3D VR Interactive Environments

More Interactive 3D environment can be added with virtual audience, the virtual audience will react when a user is delivering his/her speech. Which overall will enhance the experience of users and help them develop their communication skill quickly.

VIII. SUMMARY

This project is primarily focused on improving the user's command over their communication skill and the English by improving various factors like vocabulary, word knowledge, fluency, confidence and grammatical knowledge. Focuses not only on improving one's language, but also on developing the overall personality of the user by improving factors such as leadership skills, emotional intelligence, social skills, time management, and public speaking.

IX. ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our Mini Project guide Dr. Rajesh S. Prasad for their able guidance and support in completing our project. We would also like to extend our gratitude to the Principle of MIT-School of Engineering Dr. Kishore Ravande and our Dean and Computer Science Department Head Prof Dr. Rajneesh Kaur Sachdeo for giving us such an auspicious opportunity to convert our idea into a working project. We came to know about a lot of different new things both theoretical and practical by working on this mini-project. It not only imparted knowledge but also the made us understand the value of working together in a group. At last, we would like to thank each and every faculty member, co-students who helped us directly or indirectly to make our idea into a working project.

X. REFERENCES

- [1] Mathivanan, Kundalakesi & T, Swathi & B, Ashapriya & R, Sruthi. (2017). A Study of Virtual Reality. International Journal of Trend in Research and Development. 4. 2394-9333.
- [2] Lege, Ryan & Bonner, Euan. (2020). Virtual reality in education: The promise, progress, and challenge. JALT CALL Journal. 16. 167-180. 10.29140/jaltcall.v16n3.388.
- [3] Shaikh, Naziya & Deshmukh, Ratnadeep. (2016). Speech Recognition System – A Review. IOSR Journal of Computer Engineering. 18. 01-09. 10.9790/0661-1804020109.
- [4] Hiebert, E.H., Scott, J.A., Castañeda, R., & Spichtig, A.N. (2019). An Analysis of the Features of Words That Influence Vocabulary Difficulty. Education Sciences.

[5] Mazuryk, Tomasz & Gervautz, Michael. (1999). Virtual Reality - History, Applications, Technology and Future.

[6] Sharman, Josh. (2021). Virtual reality in education. British Journal of Nursing. 30. 1270-1270. 10.12968/bjon.2021.30.22.1270.

[7] Klačková, Ivana & Kuric, Ivan & Zajacko, Ivan & Tlach, V & Więcek, Dariusz. (2021). Virtual reality in Industry. IOP Conference Series: Materials Science and Engineering. 1199. 012005. 10.1088/1757-899X/1199/1/012005.