"PRESENTLY" –A PERSONAL TRAINER APP TO SELF-TRAIN AND IMPROVE PRESENTATION SKILLS.

21_22-J 02

Project Proposal Report

Shehara AKGH

Sri Lanka Institute of Information Technology Sri Lanka

June 2022

"PRESENTLY" –A PERSONAL TRAINER APP TO SELF-TRAIN AND IMPROVE PRESENTATION SKILLS.

21_22-J 02

Project Proposal Report

Shehara AKGH

Sri Lanka Institute of Information Technology Sri Lanka

June 2022

Declaration of The Candidate & Supervisor

I declare that this is work of mine, and this proposal does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Name	Student ID	Signature
Shehara AKGH	IT18205152	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the supervisor   Date

Acknowledgement

I would like to show my gratitude Dr. Shyam Reyal, my research supervisor, Ms. Samanthi Eranga Siriwardene, my research co-supervisor, and Ms. Veerandi Kulasekara, my research external supervisor for providing me with their valuable advice and helped me in difficult periods giving me a very helpful guidelines for the project proposal throughout numerous consultations.

I would also like to express my gratitude to my team members Wanigasingha N.T, Wandana R.A.K, Wedage C.V , without their support and cooperation this research proposal could not have been accomplished. Finally, I would like to thank my parents for their love and blessings.

Abstract

Delivering a successful presentation, a presenter must conduct the presentation clearly, confidently and cogently. Speaking clearly without pronunciation issues and vocabulary errors is essential when presenting. The proposed system "Presently" is a mobile responsive web application that facilitates the speakers to evaluate their own presentation skills. The system will provide the user with a feedback and rating of his or her performance which will be valuable points when delivering the presentation. This paper presents a method to evaluate own pronunciation and vocabulary errors occurs during the presentation using automatic speech recognition (ASR) models and rule-based approach. User can upload his or her video, audio or presentation slides of the presentation to the proposed system and from that the system will extract the audio file and will analyze for the possible pronunciation and vocabulary errors. The ASR model will take the human speech in the form of an audio file and will convert into a sequence of words. ASR software detects the speech of the speaker and creates a wave file of the words that it hears. The wave file is cleaned to delete the background noises and normalize the volume. Then this filtered wave form is then broken down and analyzed in sequences. The ASR software analyzes these sequences to determine the whole words and then complete sentences. Using the rule-based system, a sentence does not have to be complete to be checked, instead the software can check the text and give immediate feedback. In rule-based checking a set of rules is matched against a text which has at least been POS tagged.

Keywords: – presentation, evaluation, pronunciation, vocabulary, ASR models, rule-based approach

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviations	Description
ASR	Automatic Speech Recognition
LM	Language Model
HMM	Hidden Markov Models
ANN	Artificial Neural Networks
IVI	iFlytek Voice Input
POS	Parts Of Speech
DNN	Deep Natural Network
CAPT	Computer Assisted Pronunciation Training
MFCC	Mel frequency cepstrum coefficient
WER	Word Error Rate
JSON	JavaScript Object Notation
NUCLE	National University of Singapore Corpus of Learner English

1 Introduction

1.1 Background & Literature Survey

Many people are afraid of giving presentations, however there are so many circumstances when we must stand up and give a presentation to an audience. It might happen at work, university, a social gathering, or as part of a class or volunteer activity. Furthermore, due to a lack of experience speaking in front of a big audience and English language competencies, novice presenters and speakers of English as a foreign or second language frequently feel burdened by public presentations. Therefore, presenters need to guide themselves as lower-intermediate or intermediate presenters already have problems with appropriate grammar and correct pronunciation [1]. A good presentation should be written and performed in error-free and understandable English, and the presenter should appear well-prepared and rehearsed from the audience's perspective [2]. Looking unprepared with poor English both audibly and in writing, specifically with incorrect and confused pronunciation, emphasizes the presentation's negative characteristics.

Self-assessment allows presenters to analyze gaps between speech delivery and effects, as well as self-reflection, which could help presenters enhance overall presenting performance [3]. It is critical for non-native English speakers to practice before giving a presentation. For oral presentations, practicing keeping time with a friend, roommate, or family member is important, but acknowledging one's own pronunciation faults, as well as vocabulary errors, is as critical. Because presenters are overwhelmed with a need to perform a successful speech, they frequently make pronunciation and vocabulary errors, resulting in an unsatisfactory speech delivery at the end of the day.

It is critical for anyone who want to give a successful presentation to the audience to identify the mistakes that will happen during the presentation beforehand so. A student, teacher, lecturer, professional worker, or other individual can speak. The presentation skills of an individual can be used to strengthen a speaker's competency as well as their

acceptance in the workplace, university, or school. As a result, assessing those qualities would be beneficial in an induvial employment as well.

The proposed system will evaluate an individual's presenting ability and provide feedback on areas that need to be improved in order to provide a more successful presentation in the end. The survey, which was done to see if there was a need for a tool to assess presentation skills, yielded some fantastic results. Participants in the study stated that having to self-evaluate presentation skills before to giving a presentation would be extremely beneficial to speakers. In addition, the survey's findings revealed that a system for detecting a speaker's pronunciation and vocabulary faults during a presentation would be beneficial to all speakers out there. There are some figures in this research that show a summary of the responses collected during the survey that was done to gather information regarding the need for a tool to evaluate presentation skills. A figure describes the details of a previously existing system for evaluating a speaker's presentation skills. The majority of those who responded to the survey stated that they had never used any tools to evaluate presentation skills previously.
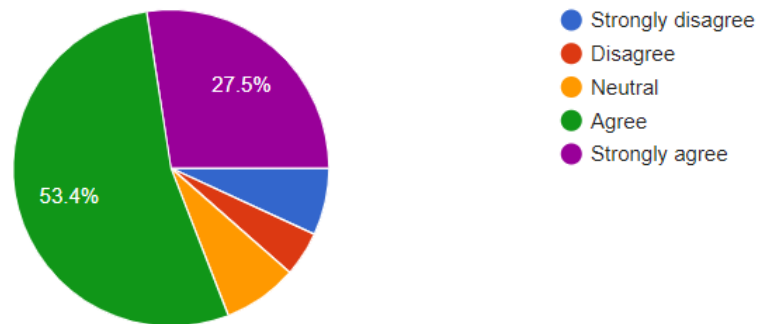

Figure 1 Summary of the responses received to check vocabulary before presentation

4) As people often make pronunciation mistakes, will suggesting those earlier be helpful during a presentation?
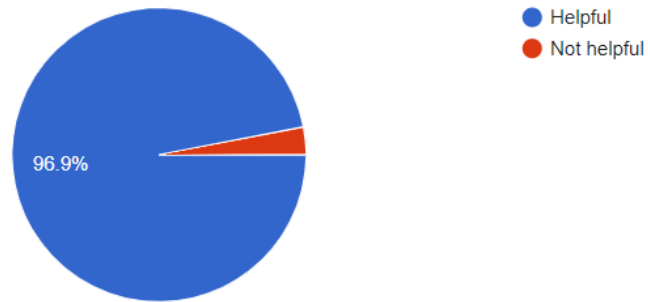
131 responses

- Helpful
- Not helpful

96.9%

Figure 2 Summary of the responses received to suggest pronunciation before presentation

1.2 Literature Survey

In this literature review, there are some research papers to identify existing research aspects of pronunciation and vocabulary errors by applying various methods to ensure the speaker is ready beforehand. During the search for research papers to write this literature survey, there were some recent related research papers of Automatic speech recognition (ASR) technology and its contribution to the identifying the errors occurs during the presentation and the way it can be uses to minimize those issues.

ASR technologies are used as a popular method to examine or practice English pronunciation. The Standard ASR systems are generally used to recognize words. The ASR system consists of a decoder which is known as the search algorithm, and three knowledge sources as the language model, the lexicon, and the acoustic models. The language model (LM) is the probabilities of words and sequences of words. Acoustic models describe the way of how the sounds of a language are pronounced. In most cases, hidden Markov models (HMMs) are used for this, but it is also possible to use artificial neural networks (ANNs). The lexicon is the connection between the language model and the acoustic models which contains information on how the words are pronounced, in terms of sequences of speech sounds. With the development of ASR technologies, it has been widely spread and proven to be very helpful in improving English pronunciation skills.

With the help of ASR technology, taking iFlytek Voice Input (IVI) as model speakers can realize clearly where they pronounce incorrectly and try to improve pronunciation by applying the meta-cognitive learning strategy, which is a new way of self-learning pattern [4]. Moreover, speech recognition indicates in case a word is not recognized or recognized as another word, then it is highly like to be mis-pronounced [5]. IVI based on ASR and speech visualization technology, can quickly recognize what the speakers say and transcribe it into the text. The IVI's accuracy rate reaches to more than 90% [4, 5]. In one research paper, there was a method that contained its research procedure in the

order of data collection in as first step, practicing its procedure as second step and data processing procedure as final step [4]. The Data collection and processing was done before the experiment called pre-test, which conducted to find the whether the speakers have problems in their pronunciation, and after experiment is called post-test conducted to testify whether there was any significant improvement of the pronunciation accuracy after the experiment. During the procedure of practicing the experiment step, speakers were required to learn by themselves. Read and self-check, reread and recheck repeatedly using selected material with IVI which can make their pronunciation visualized through transferring the speech into text. Moreover, they were required to record their speech of the presentation every week and took the pictures of text in the screen. Through this practice, they were required to learn by themselves. The goal of the data analysis was to check and prove the effect of improving English pronunciation accuracy through ASR technology using the result of the pre-test, the post-test, and the analysis of the comparison of the pre-test and the post-test. Furthermore, in this research, the IVI can visualize the speakers' pronunciation so that they can find out their own weaknesses in English Pronunciation. After self-correcting and practicing, most speakers can improve their English pronunciation accuracy. However, IVI has not widely applied for English pronunciation improvement yet because of some limitations [4]. One of the main points is this method can only give a visible feedback to tell the speakers which words they mispronounced, but this method fails in giving speakers the correct pronunciation, so speakers have to check the correct pronunciation.

By using a system built with python along with some use of R programming language where the system can be run as a Python script or even using Anacondas' Jupyter notebook, is another methodology that presents in the paper 2021, A. Deshpande et al.[5]. The interface for this framework built using a simple HTML page created using Flask, a Python web application platform. They use SpeechRecognition, a library that supports several speech recognition engines and APIs which enables users to try and test different speech recognizers together very conveniently [5]. Google Speech Recognition as it is a free engine to use that works best with their test data and provides an average

accuracy of around 90% with a Word Error Rate (WER) lower than 10% [5, 6]. The speech recognizer returns a transcript of the audio file. Words that are not present in the transcript or are present in a list of filler words get highlighted in red and this transcript is returned to the webpage. Using the values of vocal elements extracted using a Python library my-voice-analysis, they created a table that includes the values features needed to calculate the total number of pauses and filter words, total number of syllabuses in the speech etc. Words such as "uh", "um", "ah", "like", and "you know", are a few examples of filler words. Filler words must be avoided while delivering a speech since the use of filler words indicates pronunciations errors, a lack of confidence and can damage the credibility of the speaker [5]. Lastly, the table consisting of values of vocal elements helps in looking at possible improvements and keeping track of progress.

Based on Phoneme [7], which is the smallest unit of speech divided according to the natural attributes of speech, a system is designed for the auto-correction of English pronunciation errors. This is analyzed according to the pronunciation actions in syllables. An action constitutes a phoneme. Phonemes are divided into vowels and consonants. Firstly, the model of spoken English pronunciation signal is constructed, and then the voiceprint map of spoken English pronunciation is extracted. Through the Mel frequency cepstrum coefficient (MFCC) perception method [7], the voice recognition and map comparison of spoken English pronunciation is realized. This paper analyzes the input speech signals of spoken English pronunciation, digitizes them, and transforms them into digital signals. After speakers pronounce according to file system prompts, the system will combine with the standard pronunciation dictionary and pronunciation rules to form a phoneme level detection network. The current application of foe wrong pronunciation test has two kinds, one kind is based on acoustic linguistics methods, by looking for a new way of phonetic features for error detection. Another kind is to borrow from the speech recognition method, through calculating confidence scores for error detection. These two methods have been widely applied and popularized. In recent years, with the deepening of research, an algorithm of

pronunciation error detection to been optimized [7]. This system is mainly based on phonemes to detect spoken English pronunciation errors.

In 2016, S. P. Singh et al.[8] have used a rule-based approach for grammar checking. This is the approach where we can match the text with a set of rules and that has been at least Parts of Speech (POS) tagged. This method gives immediate feedback to the user, and it is easy to configure, as each rule can be turned on and off individually. In here by using the present, past and future tense are used to indicate the time, sometimes the continuation or completeness, of an action. Based on tense rules, certain rules are formed which can be understood by the system. Those are stored in JavaScript Object Notation (JASON) file. Using the noun type select the root in the rules (JSON file), then start matching the words in sentence against the nextAllowedValues. If matched, proceed to next word and continue the process until it reaches the leaf node in the rules tree. If it reaches to a leaf node, the sentence is correct. If the current word does not match with any of the rules in nextAllowedValues, the sentence is incorrect [8].

In the same paper [8] frequency based spell checker had been used for the tasks of Error detection and Suggesting prediction. There are various methods of error detection. This paper mainly focuses on using two methods Dictionary Lookup Method and N-Gram Techniques. Dictionary Lookup Method is the technique where data is arranged in rows and columns. If word is not found in dictionary then it is considered as incorrect. Advantage of Dictionary lookup technique is easy to implement but the problem is cost of searching in a large dictionary is often high, and also sometimes it is difficult to extract and store the complete vocabulary of a given language in it. But this technique is good for non- word type of error (words that don't exist in English Dictionary). N-gram analysis is a process to detect wrongly spelled words. In this approach rather than comparing each and every word in a dictionary, n-grams are used. To find an empty or deficient n-gram, the word is considered as an incorrect word, otherwise it is assumed to be correct. An n- gram is a collection of characters of length N. If N is equal to 1 then the term used is a unigram, if N is 2 then it is called Bigram, and if N is 3 then the term

is trigram and so on. Each word or string that is involved in the process of comparison is divided into pair of adjacent N-grams. The n-grams algorithm is also referred as neutral string matching or a language independent algorithm. However, in this method of grammar checking in the application it is only applied to the aspects of the grammar checking from the tenses in English language. As for the Active and Passive voice, Direct and Indirect speech the application doesn't provide the features.

A solution based relies on Deep Natural Network (DNN) mode whereas simpler pattern-based corrections are handled by rule-corrector, spelling correction, typos are handled by Spell-Checker [9]. The N-gram language model is used to filter any false corrections made by the DNN model. To generate grammatically incorrect sentences, in 2021, S. V. Gothe et al.[9] have defined certain operations namely, insert, delete and substitute that are applied to well-formed sentences. As for the Data Generation the system has used the data from various relevant online sources to procure 4M grammatically correct sentences and also obtained 1M correct sentences from a collaborative online database. For training, the system has merge different categories of artificially generated data with the publicly available datasets as National University of Singapore Corpus of Learner English (NUCLE) and Lang-8 Corpus of Learner. Final training data consists of 9M sentences including both erroneous and error-free sentence pairs and is split into the ratio of 98:2 for training and validation. They have tagged every sentence in the corpus using a POS tagger [8]. Using these tags, data can be generated. The responsibility of correcting different types of errors to different individual modules have been divided and the input sentence is passed to the Spell-Checker module, which corrects spelling mistakes and handles capitalization issues. This intermediate output is then passed to a DNN model to seek contextual corrections. To avoid any false corrections by DNN, this system has applied an N-gram based statistical filter [8, 9]. Finally, the sentence is passed to the Rule-Corrector module which further checks the sentence structure, applies corrections and generates a grammatically correct and profanity checked sentence..

There are many ways to train the pronunciation, including many features of prosody. A Computer Assisted Pronunciation Training (CAPT) system has been used to automatically assess the speaker's pronunciation and give a score showing how native-like the pronunciation is [10, 11]. By giving a precise evaluation, which includes scores for the features of pronunciation (the accent, the rhythm, etc.) to the speaker, a more effective learning can be expected. In this research paper, to evaluate the accent of a non-native utterance with known text information, first, a speech sample is synthesized using average voice of native speakers model. And then, the accent of the synthesized speech sample is substituted with the accent of non-native utterance have used. First, they have extracted the pitch contour. Then, a native speech sample is synthesized. Next, the accent of the native speech sample is modified to match the non-native utterance's accent. In 2021, Bogach et al [11] introduces a framework to detect the mispronunciations and maps each case to a phone-dependent decision tree where it can be interpreted and communicated to the non-native speakers in terms of the path from a leaf node to the root node.

In 2020, P. Gupta [12] have a built a system which takes a sentence as input, tokenizes the sentence, identifies misspelled words (if any), generates a list of suggestions and ranks them to return the top k corrections. For ranking the suggestions, they have used n-gram conditional probabilities [8]. As a preprocessing step, they have created a word frequency dictionaries that will aid in the generation of n-gram conditional probabilities. In 2020, Fouz-González et al.[13] has done a research to help foreign language learners improve their pronunciation using 52 Spanish learner participants enrolled in an English Studies degree. Pre- and post-tests were used to assess the participants' perception and production before and after training. Participants were randomly assigned to two groups as control and experimental. Training took place over a period of two weeks in which participants used the English File pronunciation app for around 20 minutes a day. Moreover, after the post-test, the group that had acted as control started to receive instruction and, after two weeks, took a second post-test, therefore acting as experimental too. However, the training showed substantial improvements in the

learners' perception and production of the target features, although the differences between groups were not statistically significant for every sound or in every task.

With the discussion on the performance of the ASR baseline systems, in 2018, Denis Jouvet et al.[14] presents the set of recent textual data from the internet for updating the speech recognition vocabularies and for training the language models, as well as the elaboration of development data sets necessary for the vocabulary selection process. This paper also compares the coverage of the training data collected from the internet, and of the GigaWord data, with finite-size vocabularies made of the most frequent words. Finally, this research paper presents and discusses the amount of out-of-vocabulary word occurrences, before and after the update of the speech recognition vocabularies. Here the speech recognition systems are based on the Kaldi speech recognition toolkit and Acoustic modeling relies on DNN, as such modeling provides the best performance.

1.3 Research Gap

According to the latest research papers obtained to complete the literature review for the individual proposal, it is possible to conclude that there have not been many research papers published in the last five years to improve a person's presentation skills. During a search for research papers, I discovered that there are no researches that have been conducted to evaluate a person's presentation skills. There is no appropriate approach to individually evaluate a speaker's talents whether they need to rehearse beforehand. If this is the case, the speaker will need to seek the assistance of a friend, roommate, or family member. This will be troublesome for the other party since it will take up time and energy. The other party may not always be accurate in providing feedback on the speaker's performance. The best solution for this would be a system that enables speakers to self-evaluate the presentation skills. However, after looking for a solution to this problem, the findings came up as zero.

As a conclusion, we proposed the "Presently" system, which will assess a speaker's presentation skills based on several factors such as pronunciation, vocabulary, emotions, facial expressions, intonation, prosody, and slide quality.

There were a few recent research papers available that could be used to verify pronunciation in classrooms [4], for university students [1, 3, 2], and so on. They were conducted with the sole purpose of examining the speech of a certain group of people. There aren't many research papers that analyze both the speakers' pronunciation and vocabulary at the same time. Some of the approaches utilized are not 100 % accurate in terms of the system's interpretation of particular words. Long and advanced words, for example, some words are not recognized by the IVI [4]. The word 'shuddering' will be recognized as 'shut the ring' when spoken. During the search for research papers, it was effective in targeting not just a specific group of people, but also a free system that would allow everyone to access the application from anywhere at any time. One of the research papers discovered had developed a mobile app to assess students' grammatical

errors[9]. However, the research found revealed, the children in schools are not permitted to use mobile phones during class time. As a result, students must limit the use of the system to times when they are not at school. The proposed system "Presently" will allow users to use the system at any time and from any location to accommodate their requirements.

	Pronunciation error detection	Vocabulary mistake detection	PRESENTLY
Research [1]	√		√
Research [2]	√		√
Research [3]	√		√
Research [4]	√		√
Research [5]	√		√
Research [6]		√	√
Research [7]	√		√
Research [8]	√	√	√
Research [9]		√	√
Research [10]	√		√
Research [11]	√		√
Research [12]	√		√
Research [13]	√		√
Research [14]		√	√

Table 1 Research comparison Table

1.4 Research Problem

Have you used any existing system to self evaluate your presentation performance before the actual presentation?

131 responses

● Yes
● No
● Maybe

94.7%
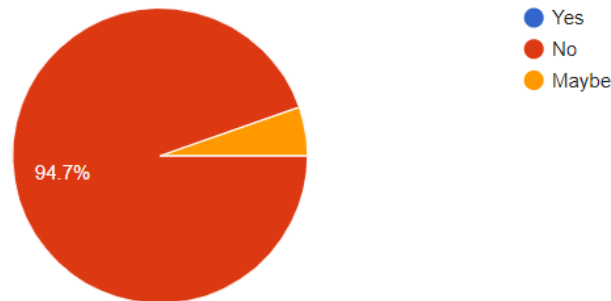
Figure 3 Summary of the responses for existing systems to evaluate presentation skills

When speaking, people who speak English as a second language face significant challenge. People are concerned about speaking English since they frequently make grammatical and vocabulary flaws. This is very common while performing a presentation. When people get nervous before a presentation, they are more prone to make these errors. People want to rehearse beforehand to avoid uneasiness and awkwardness during the presentation, as practice makes a man perfect. Furthermore, practicing requires seeking someone's assistance in providing feedback on one's performance. This is not very successful or 100 % accurate because the offered feedback may differ from one person to another due to their diverse evaluation methods. For the problem of an appropriate way of evaluating the presentation skills of an individual, we proposed "Presently" a mobile responsive web application as a solution.

Pre-evaluation of the presentation and memorizing the essential flow rather than all the words, would improve the presentation. As a result, practicing with the help of an app will improve the presentation's success.

2) Would you like to self-evaluate your presentation skills using a web application?

131 responses

98.5%

Yes
No

Figure 4 Summary of the responses for checking the interest to self-evaluate presentation skills using a web application

2 Objectives

2.1 Main Objective

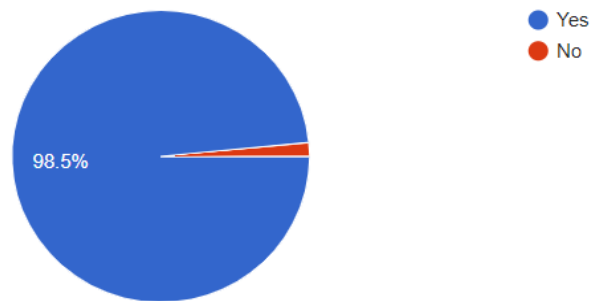The major goal of implementing "Presently," a mobile responsive web application, is to enable speakers to monitor their improvement in terms of presentation abilities based on the feedback they received. The others, preferably a friend, a roommate, a family member, etc., cannot flexibly schedule the time to seek feedback since practicing the presentation takes more time and effort. They may be unable to provide the precise feedback required to improve their presentation skills. In this research, we focus on analyzing the audio, video, and slides of the presentation uploaded by the speakers. Preparing for a presentation has been shown to be essential for improving emotional control, intonation and prosody, pronunciation and vocabulary, as well as the quality of the presentation slides. As a result, practicing has become one of the most critical parts of giving a good presentation.

As a result, the main objective of "Presently" is to implement a Mobile Responsive Web Application that will assist presenters in preparation beforehand for their presentations in order to create an effective speech to the audience.

2.2 Specific Objectives

Some specific objectives that must be met in order to achieve the main objectives in the implementation of the "Presently" are listed below.

Specific Objective 1: When delivering a presentation to provide a user with incorrect pronunciation and vocabulary mistakes.

- Identifying possible pronunciation issues that could arise throughout the presentation.

Speakers with excellent English pronunciation are much more likely to be understood even if they make other flaws, whereas speakers with poor pronunciation will not be

understood even if their grammar is flawless. We evaluate presenters by the way they speak during a presentation, thus speakers with poor pronunciation may be perceived as unskilled or lacking in expertise.

- Evaluation of grammatical mistakes in enhancing the audience's understanding of the presentation

Vocabulary is fundamental in the learning of a second or foreign language since learners cannot comprehend others or convey their own feelings without it. Vocabulary words are those that every learner may utilize in a variety of situations. A solid foundation of vocabulary knowledge is required for non-native speakers into becoming successful learners of English in any academic or professional context, in addition to grammar and pronunciation. Speakers cannot effectively communicate unless they have an adequate vocabulary. In addition, grammatical errors result in ungrammatical utterances, whereas improper vocabulary usage has an impact on interpersonal communications. The more often speakers are exposed to vocabulary, the more confident they are in their ability to grasp and interpret the meaning of unfamiliar terms in any situation. Learning vocabulary entails not just learning new words but also understanding their functions and applicability in various settings and situations.

As stated in specific objective 1, the user can upload their presentation recordings to the proposed system "Presently," from which they can evaluate their presenting capabilities. The system will extract and analyze the audio clip from the uploaded video. Audio analysis will assist the users in comprehending their pronunciation and vocabulary faults when presenting. Using the system's feedback on pronunciation and vocabulary faults, the user can gain a deeper grasp of the basics and reduce the number of mistakes that were made during the presentation.

3 Methodology
3.1 Project Overview

The system overview diagram, system overview, and resources and technologies listed below will aid in understanding the systematic, theoretical evaluation of the methodologies used in the "Presently" research field of study.

English is a language spoken by around 380-420 million people on the globe, and it is not easy to comprehend. The meaning of a statement varies depending on the circumstances and speaker's tone [1,2]. Poor communication skills can weaken an individual's self, limiting social connections, and negatively influence the speaker's credibility and capabilities. Speakers need pay more attention to the development of their competence and focus on a more effective and successful methodology in order to satisfy the expectations of modern society.

The proposed system "Presently" will evaluate a speaker's presentation capabilities in a variety of areas, involving emotion control, intonation and prosody, pronunciation and vocabulary, and presentation slide quality. In order to deliver a successful presentation to the audience, a speaker needs rehearse beforehand. This typically requires an excellent speaking delivery. The objective of giving a presentation is to captivate the audience's attention with a strong delivery, and that involvement contributes to establish sense of community.

The proposed system will aid anyone who needs to prepare a presentation by detecting emotions and body language using video, detecting emotions, tonality, and prosody using audio, verifying pronunciations and vocabulary, and assessing presentation slide quality using the proposed system. Overall, the system will offer the presenter a rating and feedback on their performance before they commence their official presentation.

3.2 System Overview Diagram
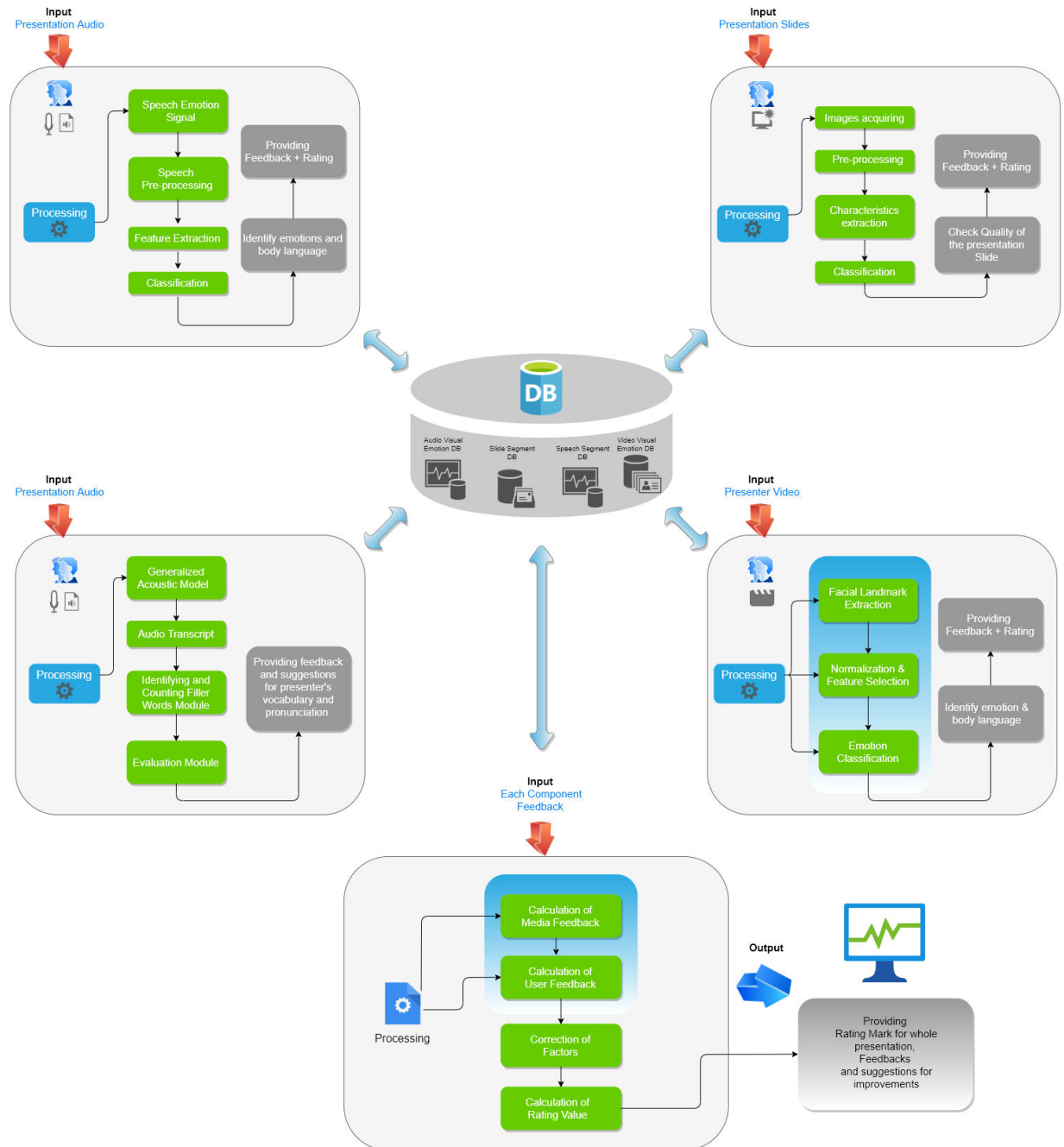
3.3 Individual Component - When delivering a presentation to provide a user with incorrect pronunciation and vocabulary mistakes.

The most crucial aspect of our everyday lives is communication. The wheels of communication propel our society forwards. As a result, accurate pronunciation is essential for effective communication, as pronunciation has a significant impact on understanding the meanings of words. Everyone should have a strong vocabulary and understand grammatical principles in order to communicate effectively. As English is Sri Lanka's second language, people often make more pronunciation and vocabulary errors when speaking. People have to always deal with a lot of challenges when giving a presentation because of these. As a result, rehearsing a presentation ahead of time will provide several opportunities to rectify these pronunciation and vocabulary faults.

When presenters upload an audio file, video file, or presentation slide, the system we proposed will suggest inaccurate pronunciations and vocabulary errors. Whenever a video clip is uploaded, the system we proposed extracts the audio clip and analyzes it independently for any inaccurate pronunciation and vocabulary errors. After the evaluation is conducted, the system will provide a rating and feedback on the presenter's performance, which will help them improve their capabilities when delivering the actual presentation or speech. Natural Language Processing techniques will be used to evaluate the audio and check for pronunciation issues (NLP).). Eventually, anybody who requires aid in rehearsing a speech or presentation will receive assistance from the system in order to deliver the speech satisfactorily.

Architectural Design

3.3.1 The flow of the system
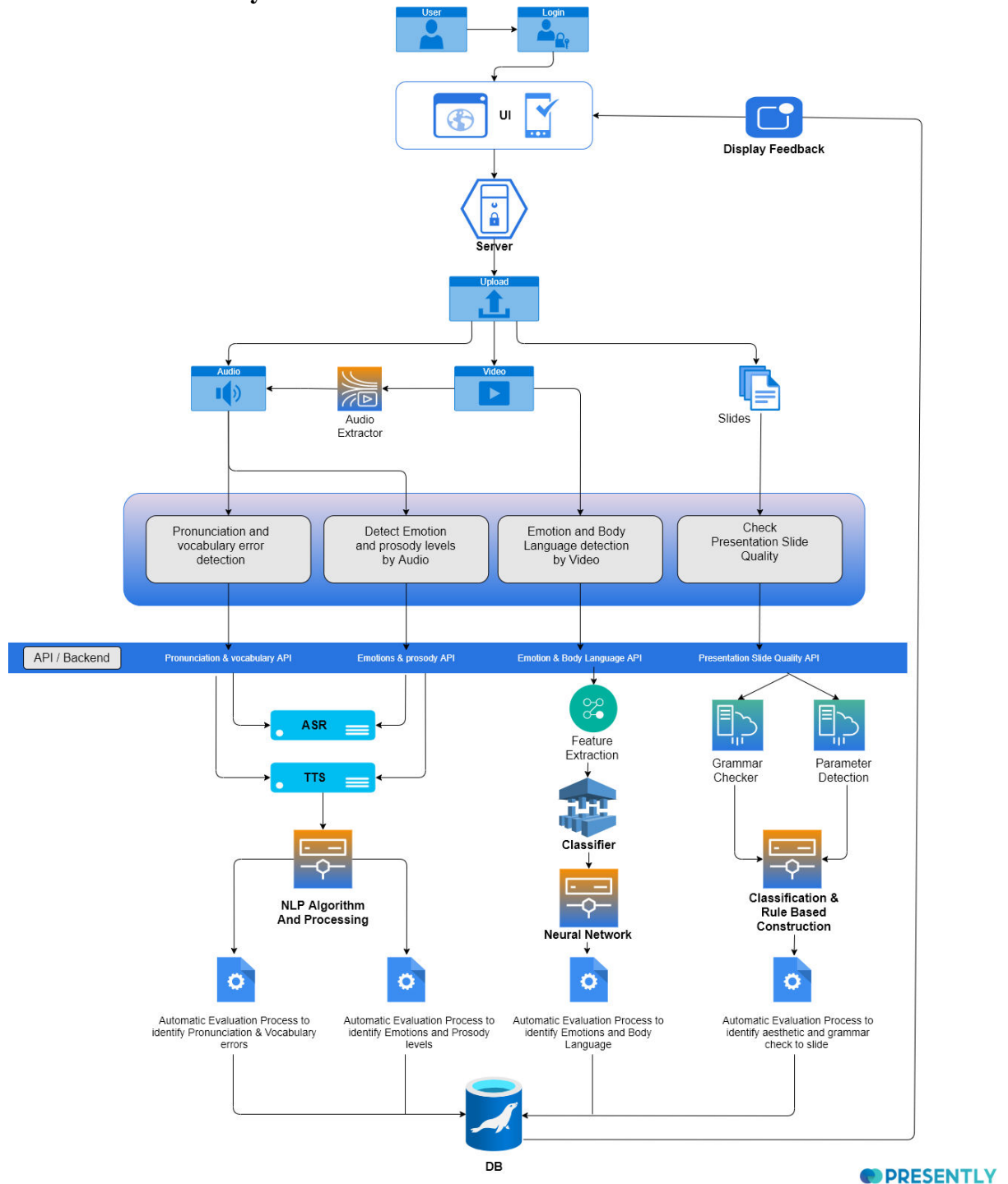
Architectural Diagram

3.3.2 The flow of the system – Individual Component
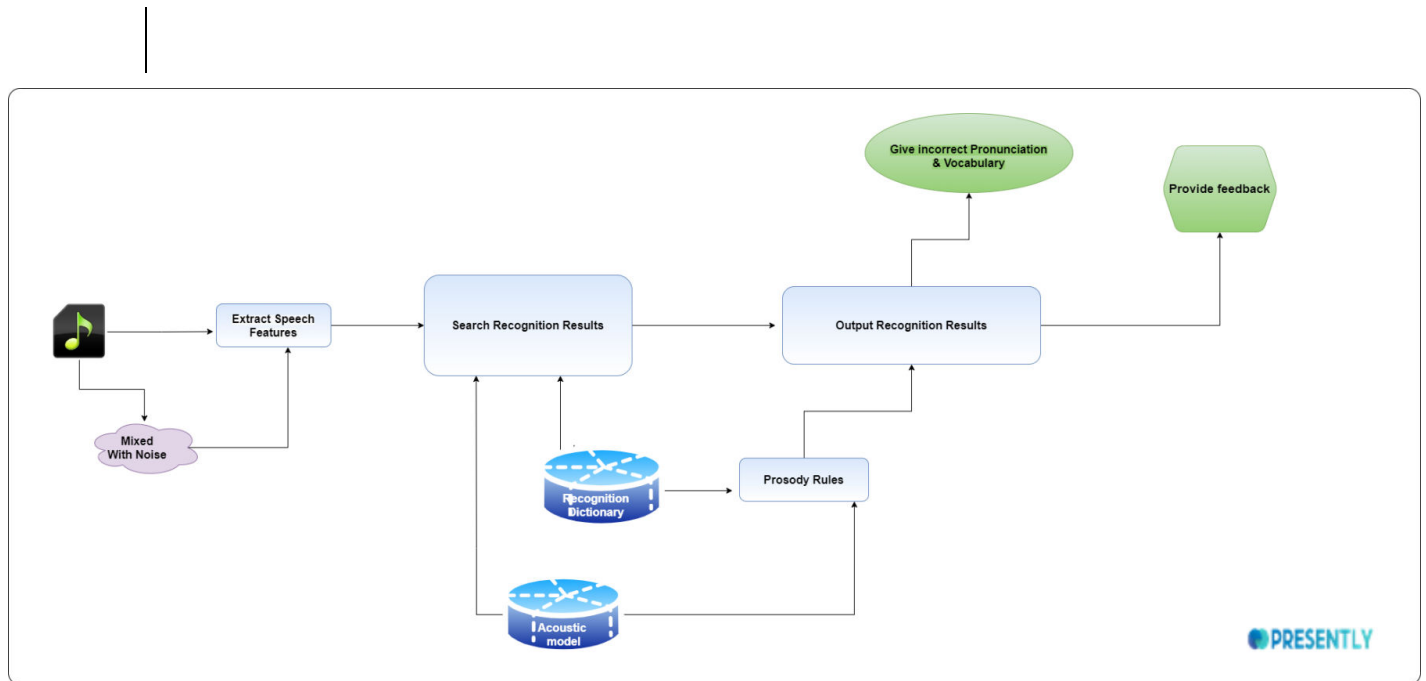
Figure 5 Overview of model architecture for pronunciation and vocabulary components
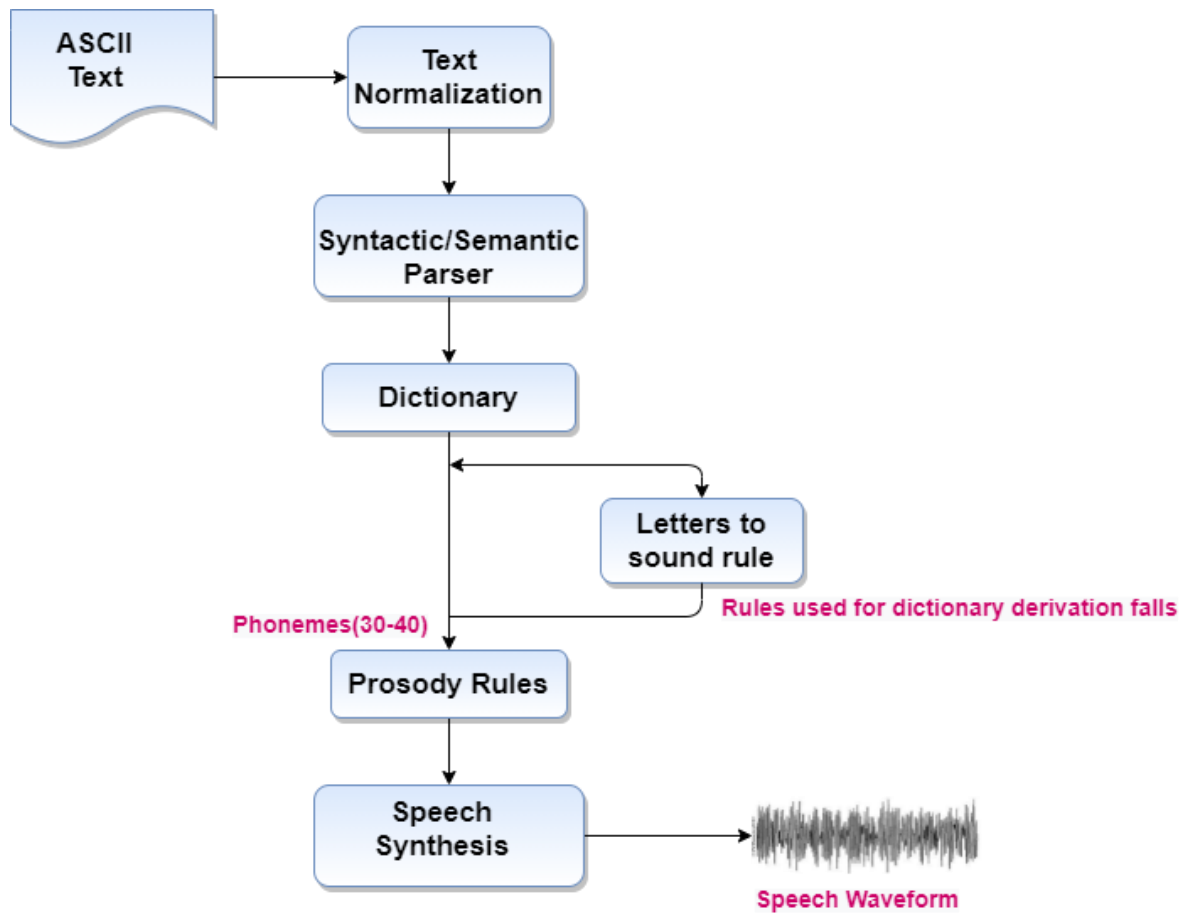
Figure 6 An Overview of Text Extraction for Vocabulary Checking

3.4 Wireframes of the proposed system
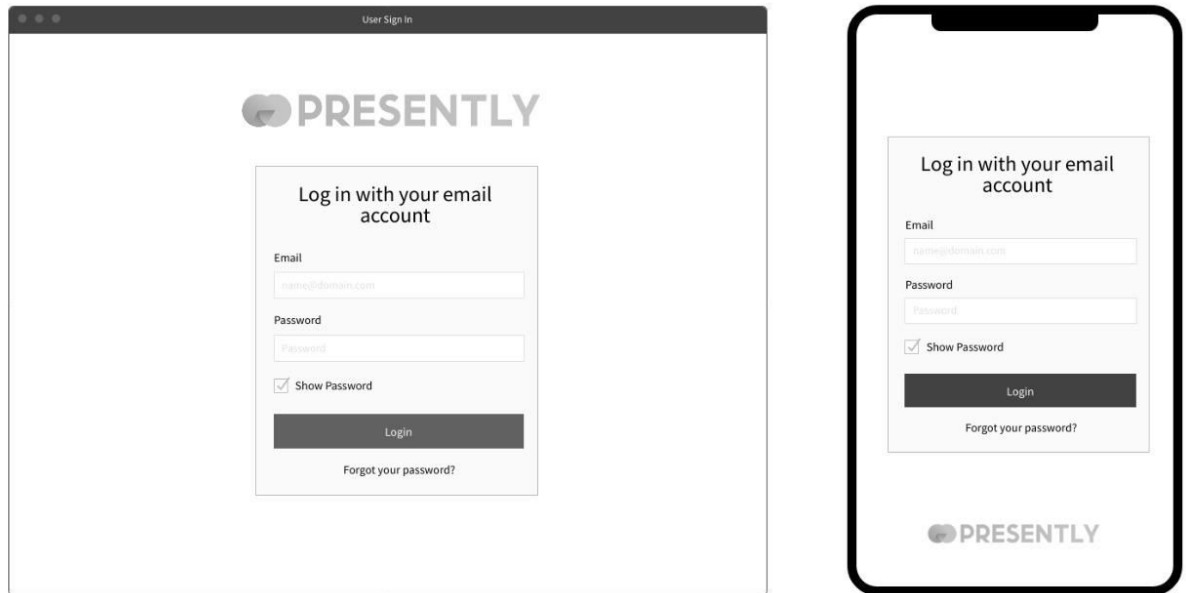
Figure 7 wireframe of the login page
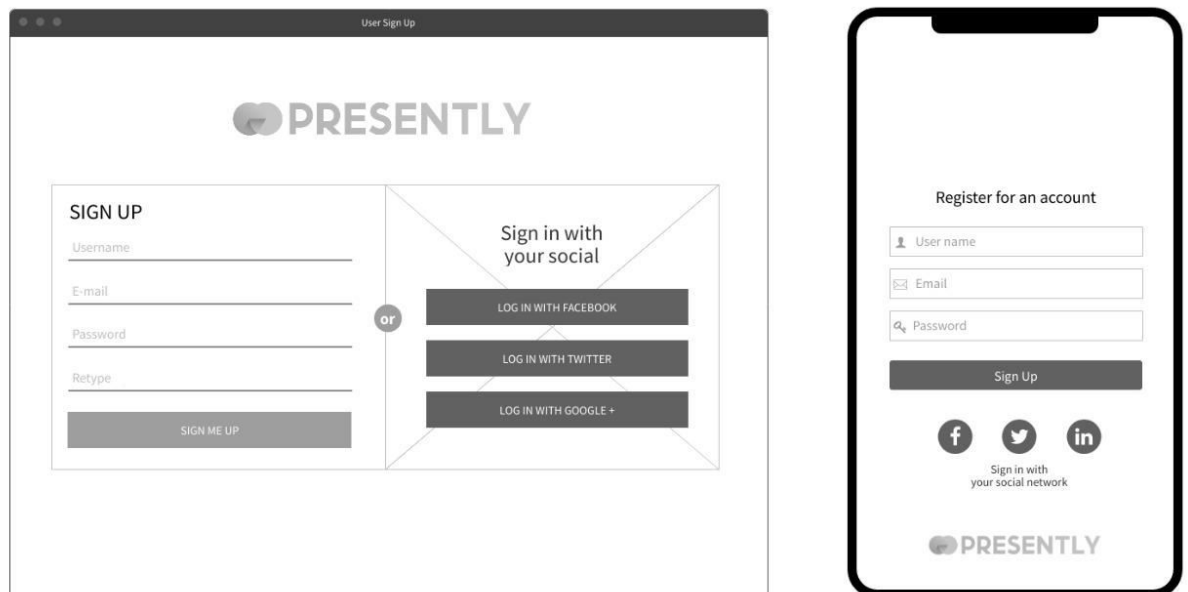
Figure 8 Wireframe of the Registration page

Figure 9 Wireframe of the Home page
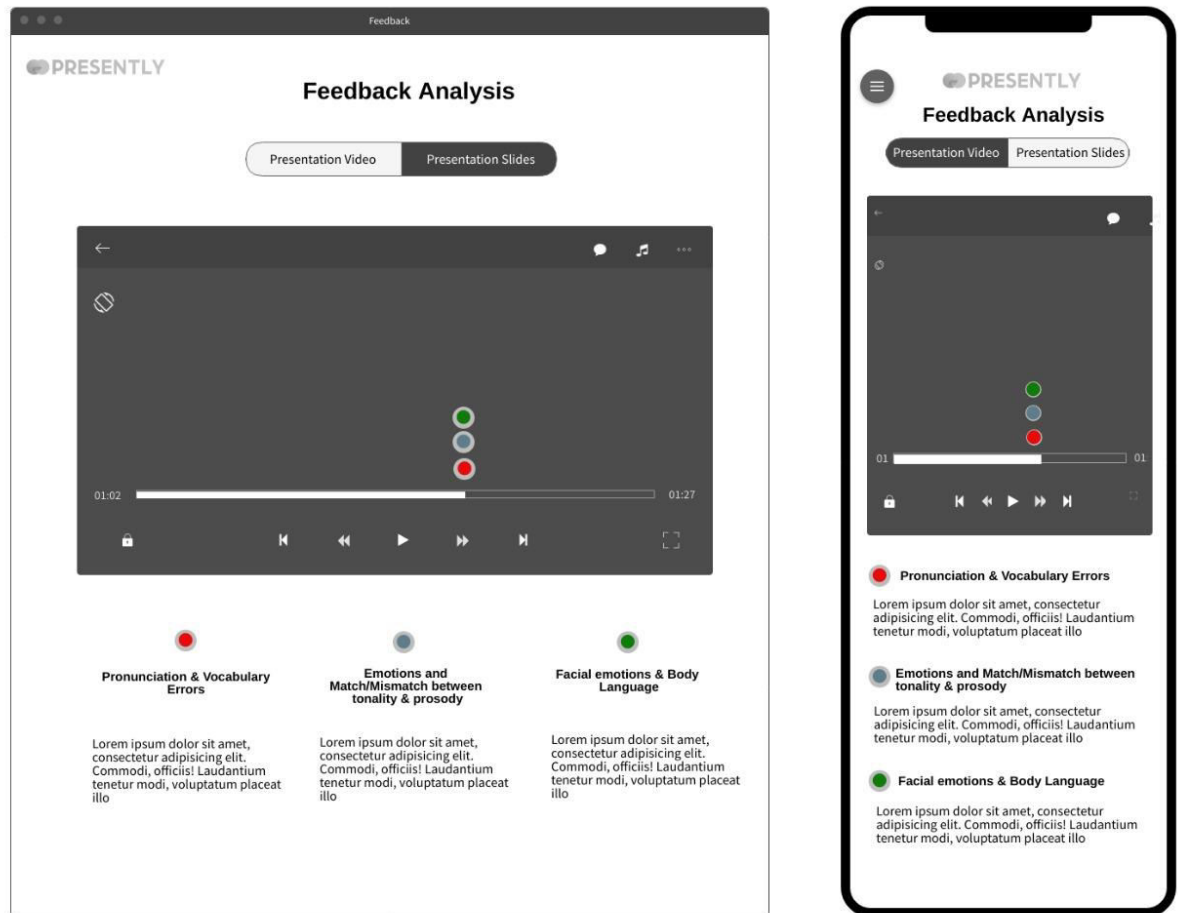
Figure 10 Feedback analysis of the Video and Audio files
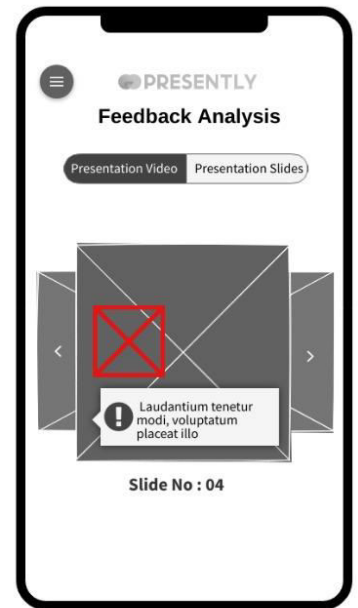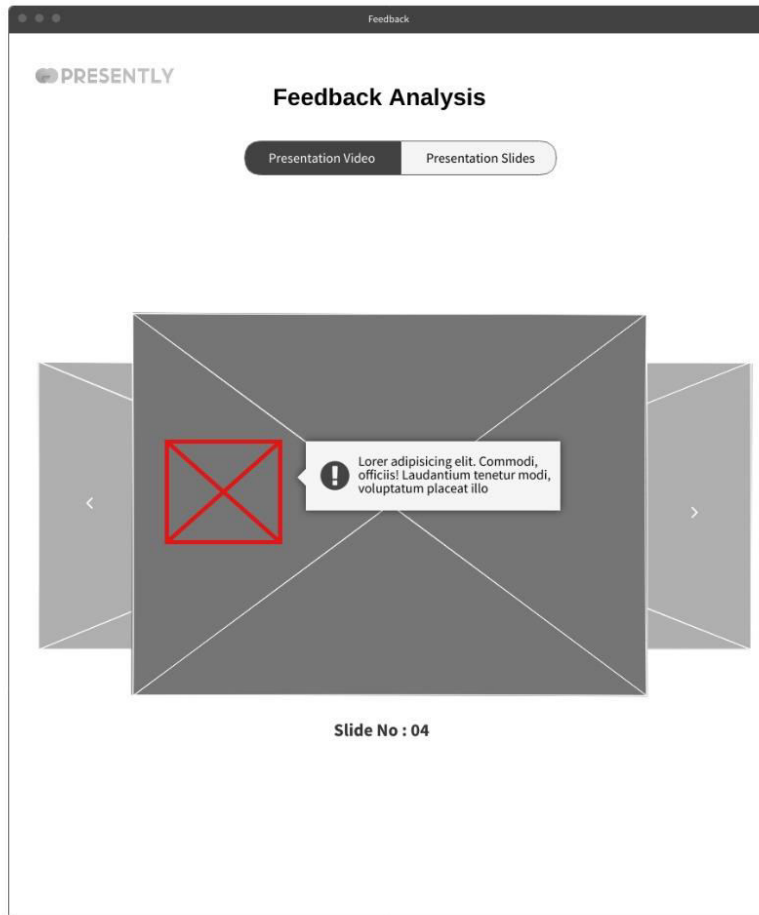
Figure 11 Feedback analysis of the presentation slides

3.5 Technologies and Implementations

1. The analyzation of the audio Input.

   This will be done using the Automatic Speech Recognition system and will turn the speech in to text.

2. Suggesting the vocabulary errors.

   This will be done using the rule based system.

Python will be used as the programming language because it is interpreted, high-level, general programming language.

Git hub will be the version control used for this research as it is helpful for the team members to track and manage changes to a software project's code.

Amazon S3 bucket which is a Storage service offered by Amazon Web Services and MySQL will be used as the database.

FastAPI which is a web framework for developing APIs with Python 3.6+ based on standard Python will be used.

3.6 Work Breakdown chart and Gnatt chart

3.6.1 Work Breakdown chart

3.6.2 Gnatt chart

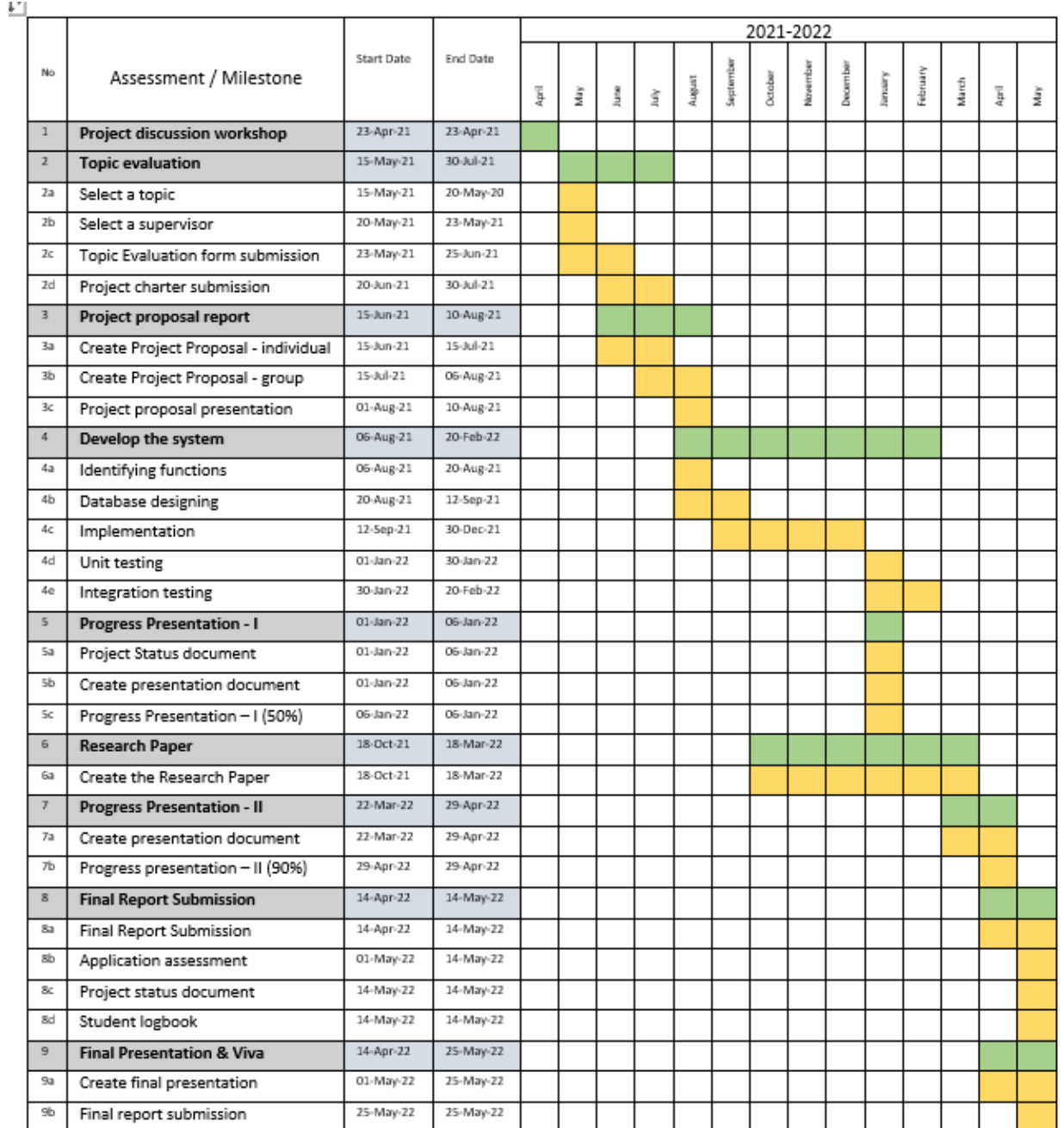| No	Assessment / Milestone	Start Date	End Date	April	May	June	July	August	September	October	November	December	January	February	March	April	May
1	**Project discussion workshop**	23-Apr-21	23-Apr-21	G													
2	**Topic evaluation**	15-May-21	30-Jul-21		G	G	G										
2a	Select a topic	15-May-21	20-May-20		Y												
2b	Select a supervisor	20-May-21	23-May-21		Y												
2c	Topic Evaluation form submission	23-May-21	25-Jun-21		Y	Y											
2d	Project charter submission	20-Jun-21	30-Jul-21			Y	Y										
3	**Project proposal report**	15-Jun-21	10-Aug-21			G	G	G									
3a	Create Project Proposal - individual	15-Jun-21	15-Jul-21			Y	Y										
3b	Create Project Proposal - group	15-Jul-21	06-Aug-21				Y	Y									
3c	Project proposal presentation	01-Aug-21	10-Aug-21					Y									
4	**Develop the system**	06-Aug-21	20-Feb-22					G	G	G	G	G	G	G			
4a	Identifying functions	06-Aug-21	20-Aug-21					Y									
4b	Database designing	20-Aug-21	12-Sep-21					Y	Y								
4c	Implementation	12-Sep-21	30-Dec-21						Y	Y	Y	Y					
4d	Unit testing	01-Jan-22	30-Jan-22										Y				
4e	Integration testing	30-Jan-22	20-Feb-22										Y	Y			
5	**Progress Presentation - I**	01-Jan-22	06-Jan-22										G				
5a	Project Status document	01-Jan-22	06-Jan-22										Y				
5b	Create presentation document	01-Jan-22	06-Jan-22										Y				
5c	Progress Presentation – I (50%)	06-Jan-22	06-Jan-22										Y				
6	**Research Paper**	18-Oct-21	18-Mar-22							G	G	G	G	G	G		
6a	Create the Research Paper	18-Oct-21	18-Mar-22							Y	Y	Y	Y	Y	Y		
7	**Progress Presentation - II**	22-Mar-22	29-Apr-22												G	G	
7a	Create presentation document	22-Mar-22	29-Apr-22												Y	Y	
7b	Progress presentation – II (90%)	29-Apr-22	29-Apr-22													Y	
8	**Final Report Submission**	14-Apr-22	14-May-22													G	G
8a	Final Report Submission	14-Apr-22	14-May-22													Y	Y
8b	Application assessment	01-May-22	14-May-22														Y
8c	Project status document	14-May-22	14-May-22														Y
8d	Student logbook	14-May-22	14-May-22														Y
9	**Final Presentation & Viva**	14-Apr-22	25-May-22													G	G
9a	Create final presentation	01-May-22	25-May-22														Y
9b	Final report submission	25-May-22	25-May-22														Y

3.7 Requirement Analysis

3.7.1 Functional Requirements

- Anyone who need assistance in practicing the presentation should be able to log in to the "Presently".
- A user should be able to upload their video or audio files to the system.
- System should facilitate the user to upload the presentation slides to the system only if the user desire.
- If the presentation slide is uploaded the system should analyze it.
- System should be able to provide the user with the relevant feedback.

3.7.2 Non-Functional Requirements

1. Performance

    - The system should facilitate the user to upload the recording of the presentation to the system in the form of audio or video.

    - The system should analyze the uploaded files quicker.

    - The websites load time should not be more than one second for the users.

2. Correctness

    - The system should be error free.

    - All the feedback provided by the system should be accurately to the highest level.

3. Compatibility

    - The system should be run in any device.

    - The system should be mobile responsive

4. Security

    - The system should allow the user to log into the system with correct credentials.

    - The user should not have the access to the details of the other users the use the system

5. Usability

- The system should be easy to navigate by the users.
- The features of the system should be function well.

4 Budget and Budget Justification

Component	Amount (Rs.)
Internet	3000.00
Stationery	2000.00
Documentation and printing cost	5000.00
Server cost	4000.00
Educational survey cost (online payments)	1000.00
Electricity	1000.00
Transport	2500.00
Total	17500.00

Table 2 Budget and budget justification

5 References

[1] S. Bosley and B. Anuyahong, "A development of EFL presentation skills by using the communicative approach for university students," 2018 5th International Conference on Business and Industrial Research (ICBIR), 2018, pp. 522-525, doi: 10.1109/ICBIR.2018.8391253.

[2] J. Shi, "Developing research presentation skills for international conferences," Proceedings of the 2013 International Conference on Advanced Mechatronic Systems, 2013, pp. 406-411, doi: 10.1109/ICAMechS.2013.6681817.

[3] Z. Ma, Y. Wang and X. Guan, "Developing oral presentation skills through formative self- and peer-assessment in the context of China," 2019 Eighth International Conference on Educational Innovation through Technology (EITT), 2019, pp. 199-205, doi: 10.1109/EITT.2019.00046.

[4] M. Li, M. Han, Z. Chen, Y. Mo, X. Chen and X. Liu, "Improving English Pronunciation Via Automatic Speech Recognition Technology," 2017 International Symposium on Educational Technology (ISET), 2017, pp. 224-228, doi: 10.1109/ISET.2017.58.

[5] A. Deshpande, R. Pandharkar and S. Deolekar, "Speech Coach: A framework to evaluate and improve speech delivery," 2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP), 2020, pp. 1-5, doi: 10.1109/ICCCSP49186.2020.9315267.

[6] P. Vanajakshi and M. Mathivanan, "A detailed survey on large vocabulary continuous speech recognition techniques," 2017 International Conference on Computer Communication and Informatics (ICCCI), 2017, pp. 1-7, doi: 10.1109/ICCCI.2017.8117755.

[7] Y. Xin, "The Design of Automatic Correction System for Spoken English Pronunciation Errors Based on Phonemes," 2019 6th International Conference on Dependable Systems and Their Applications (DSA), 2020, pp. 247-254, doi: 10.1109/DSA.2019.00038.

[8] S. P. Singh, A. Kumar, L. Singh, M. Bhargava, K. Goyal and B. Sharma, "Frequency based spell checking and rule based grammar checking," 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016, pp. 4435-4439, doi: 10.1109/ICEEOT.2016.7755557.

[9] S. V. Gothe, S. Dogra, M. Chandra, C. Sanchi and B. R. K. Raja, "An efficient system for grammatical error correction on mobile devices," 2021 IEEE 15th International Conference on Semantic Computing (ICSC), 2021, pp. 147-154, doi: 10.1109/ICSC50631.2021.00034.

[10] H. Prafianto, T. Nose and A. Ito, "A precise evaluation method of prosodic quality of non-native speakers using average voice and prosody substitution," 2016 International Conference on Audio, Language and Image Processing (ICALIP), 2016, pp. 208-212, doi: 10.1109/ICALIP.2016.7846620.

[11] Bogach, Natalia & Boitsova, Elena & Chernonog, Sergey & Lamtev, Anton & Lesnichaya, Maria & Lezhenin, Iurii & Novopashenny, Andrey & Svechnikov, Roman & Tsikach, Daria & Vasiliev, Konstantin & Pyshkin, Evgeny & John, Blake. (2021). Speech Processing for Language Learning: A Practical Approach to Computer-Assisted Pronunciation Teaching. Electronics. 10. 235. 10.3390/electronics10030235.

[12] P. Gupta, "A Context-Sensitive Real-Time Spell Checker with Language Adaptability," 2020 IEEE 14th International Conference on Semantic Computing (ICSC), 2020, pp. 116-122, doi: 10.1109/ICSC.2020.00023.

[13] Fouz-González, Jonas. (2020). Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. Language, Learning and Technology. 24.

[14] Denis Jouvet, David Langlois, Mohamed Menacer, Dominique Fohr, Odile Mella, et al.. Adaptation of speech recognition vocabularies for improved transcription of YouTube videos. Journal of International Science and General Applications, ISGA, 2018, 1 (1), pp.1-9. ffhal-01873801f

[15] Cavus, N. (2016). Development of an Intelligent Mobile Application for Teaching English Pronunciation. In Procedia Computer Science (Vol. 102, pp. 365–369). Elsevier B.V. https://doi.org/10.1016/j.procs.2016.09.413

[16] Modeling, P.S., and St, N.C. (2019). IMPROVING END-TO-END SPEECH RECOGNITION WITH Hainan Xu , Shuoyang Ding , Shinji Watanabe Center for Language and Speech Processing ,. Ieee 7110–7114.

[17] Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2021). Effects of ASR-based websites on EFL learners' vocabulary, speaking anxiety, and language enjoyment. System, 99. https://doi.org/10.1016/j.system.2021.102496

[18] Mukherjee, P., Santra, S., Bhowmick, S., Paul, A., Chatterjee, P., & Deyasi, A. (2018). Development of GUI for text-to-speech recognition using natural language processing. In 2018 2nd International Conference on Electronics, Materials Engineering and Nano-Technology, IEMENTech 2018. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/IEMENTECH.2018.8465238