ENGLISH SCHOOL MATE

S.N.S.Divyanjali

(IT17400596)

Bachelor of Science Special (honors) in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2020

ENGLISH SCHOOL MATE

Project id: 2020-069

(Design documentation submitted in partial fulfillment of the requirement for the

Degree of Bachelor Science Special (honors)

In Information Technology)

Bachelor of Science Special (honors) in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2020

1. Declaration

Declaration of Candidates

"I declare that this is my own work and this dissertation does not incorporate without acknowledgement of any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also, I hereby grant to Sri Lanka Institute of Information Technology, the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic, or other media. I retain the right to use this content in whole or part in future works (such as articles or books)."

Signature: Date:

Declaration of the Supervisor

"The above candidate has carried out research for the bachelor's degree dissertation under my supervision."

Signature of the supervisor: Date:

2. Abstract

With the expansion of modern technology, the lifestyle of people is mostly managed and relied on their computers and mobile phones to an increasing extent. On that premise, learning spoken English using their mobile phone and the computer can be introduced to education platforms as a teaching and learning method and it allows reducing the lack of teaching methods and students can learn with entertainment. Traditionally most students learn spoken English with their teacher and their friends in the school or their tuition classes. Less count of students learn from their family background. This research targeted those school learners of above mentioned and proposed the online educational system to them because of some sort of problems they faced. Mainly they are faced with a lack of teachers and the physical resources and environmental challenges so in rural areas that can be extremely difficult to attract great teachers. This proposed module is like a virtual English teacher. There are lessons in an entertaining way and will give marks after exams. The main thing is students can practice spoken English as with their teacher or friend because using speech recognition. Several versatile techniques are used throughout the exercise development process to improve the effectiveness of the product including conversations, dialogs, model speeches, poems and pronunciations. Speech recognition technology is highly used.

3. Acknowledgement

I would like to present my heartfelt gratitude for the immense support, guidance, and motivations provided by my supervisor Mr. Dhammika de Silva throughout the research and for his valuable feedback in helping make this research more effective and a success. The enthusiasm he showed towards research motivated us to try different approaches to discuss with industry experts, which otherwise would not have been possible.

Moreover our special gratitude goes to all the teachers, students and parents who helped to collect information on the English subject and their data. Special thank you goes to all the industry experts and the graduates who helped in determining the best technologies when designing the solution.

I am glad to mention something about all the volunteers who participated in answering the questionnaires during the testing phase of our solution. Their feedback gave us hope to do well in this worthy research.

Furthermore, we would like to thank our parents for their patience, time, and providing resources to acquire the needful and for other expenses. A special thanks to the fellow research members who worked alongside to make this research more effective and successful.

Table of Contents

List of Tables

List of Figures

1 Introduction

Governments of the country since the 1950s are teaching of English as a second language in schools. Within the current context of globalization, technological advancement and also the development of a contemporary and diversified labor market, the employment of English language skills is recognized as essential. The need for the English language may also be seen particularly in emerging entrepreneurial and innovative businesses. Despite the need for English language skills, particularly within the private sector, the dearth of English proficiency remains an issue for graduates entering the labor market and for those already working and looking out for development of their careers. Furthermore, for college kids considering instruction options overseas, the knowledge of land language is important.

English language education must be attention at the first level, junior secondary level and senior secondary level of all schools. All told schools, students must incline the chance, no matter socio-economic and regional differences, to accumulate a level of English proficiency that aids them in instruction and career advancement [7].

Access to English education across the country is partly determined by the supply of teachers who are well trained in teaching English to students. The allocation of teachers to different schools determines access to English learning skills for college kids across the country. Disadvantaged schools or schools in extremely rural areas of the country require good English teachers. As well as many of the problems they have to fulfill their education. Some of those are lack of teaching methods, physical distance of students to school, difficulty finding teachers interested in relocating, poor connection and poverty [1]. This research proposes a solution to help mainly rural area students to learn English language entertainingly.

So the main focus is to present an integrated solution to improve all the main aspects of English Language while preserving the interest of students to learn. This also introduces an Instructor Dashboard to help mentors visualize results easily and to predict student proficiency levels. This document is based on the Spoken English module which is one of the components of the above mentioned integrated solution.

1.1 Background literature

Anyone wants a language to communicate with others so speaking is one of the major parts in any language. Speaking English allows you to truly broaden your world, from job opportunities to the flexibility to relate to people from every country. Knowing the language makes it way more interesting every trip. Education is incredibly important to enhance yourself but learning English also improves the standard of life.

- **Yuichi Ono, Takumi Ishii, Akio Ohnishi "Construction of a Voice-based Asynchronous Communication System Utilizing Speech Recognition and Its Potential for EFL Learners' Speaking Ability: A Pilot Study"[2]**

  The present paper deals with the construction of an asynchronous voice-based computer-mediated communication (CMC) system for less confident English as a Foreign Language learners. The results from this pilot evaluation of the system are discussed in terms of its usability and effectiveness at reducing foreign language anxiety. The proposed system incorporates a browser-driven Automatic Speech Recognition (ASR) into a blog to provide real-time feedback on their pronunciation before posting. With the results from the questionnaire survey conducted in this pilot study, we demonstrate that this system reduces foreign language anxiety in speaking and increases motivation for less motivated learners. [2]

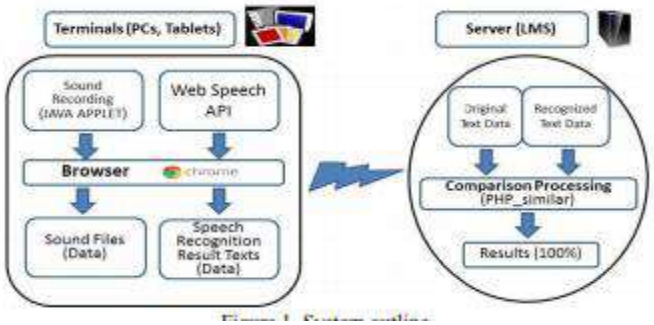

Figure 1.1.1 : System outline

1. **Use of Web Speech API -** As a speech recognition tool, the system Speech API, which was formulated by the W Consortium. Google Chrome was selected for the pilot study because it offers partial technology. For sound recording, Java apply Multiple, simultaneous processing of sound voice recognition was made possible b JavaScript. [2]

- **Valéria Farinazzo Martins, Eduardo Lombardi, Luana Felix da Silva, Marcelo de Paiva Guimarães "Using the recognition and speech synthesis to assist the practice of English pronunciation "[3]**

  Recognition and voice synthesis systems have been used in diverse situations, such as by phones, GPS´s, games and consumer services. This is because speaking is inherent in human beings which makes it a user-friendly computer interface. This paper aims to present a mobile application (Talk2Practice) based on recognition and voice technology to support English pronunciation teaching. In order to evaluate the usability of this application, tests were performed with 19 users.[3]

Figure 1.1.2. : Interfaces of the Talk2Practice system

- **Rebecca Hincks,"SPEECH RECOGNITION FOR LANGUAGE TEACHING AND EVALUATING: A STUDY OF EXISTING COMMERCIAL PRODUCTS" [4]** [1]

Educators and researchers in the acquisition of L2 phonology have called for empirical assessment of the progress students make after using new methods for learning. This study investigated whether unlimited access to a speech-recognition-based language learning program would improve the general goodness of pronunciation of a group of middle-aged immigrant professionals studying English in Sweden. Eleven students were given a copy of the program Talk to Me by Auralog as a supplement to a 200-hour course in Technical English, and were encouraged to practice on their home computers. Their development in spoken English was compared with a control group of fifteen students who did not receive software. Talk to Me uses speech recognition to provide conversational practice, visual feedback on prosody and scoring of pronunciation. A significant limitation of commercial systems currently available is their inability to diagnose specific articulatory problems. However, in this course students also met at regular intervals with a pronunciation tutor who could steer them in the right direction for finding the most important sections to practice for their particular problems. Students reported high satisfaction with the software and used it for an average of 12.5 hours. Students were pre- and post-tested with the automatic PhonePass SET-10 test from Ordinate Corp. Results indicate that practice with the program was beneficial to those students who began the course with a strong foreign

accent but that students who began the course with intermediate pronunciation did not show the same improvement[4].

- Speech recognition for teaching and evaluating -

With proper adaptation, speech technology allows beginning language students to practice spoken language outside the classroom. Dialogue-based software using fixed-response ASR lets learners have a simulated conversation with a computer. Practicing with such programs should help students improve fluency and confidence. Furthermore, the software can provide individual feedback on pronunciation, which is something that is often lacking in the language classroom. Algorithms calculate by how much a given pronunciation has deviated from a model, and give a score on phonetic accuracy[4].

- Software used for this project -

One company that has successfully marketed language learning programs that use ASR is the French company Auralog. Auralog thoroughly integrated the technology into its Talk to Me language series, which is the product we chose for our study. The core of the software consists of six dialogues, where the user utters one of three responses. Each dialogue consists of thirty question and-answer screens with accompanying photographic illustrations and occasionally music and video clips as well. The act of choosing a response initiates a degree of spontaneity into the 'conversation' and hopefully allows more natural language than would be enabled by just reading one specific response. While the dialogues in practice communication skills, more specific pronunciation training is carried out at sentence, word, or phoneme level. At phoneme level, users are shown 3D animations of a sagittal section, showing how the sound is articulated. At word and sentence levels, each response from the dialogues is practiced individually. Users compare the waveform and pitch curve of their own production with that of the model speaker. A score for the production is given, on a scale from one to seven. If the program has found particular difficulties recognizing a specific word in the phrase, that word is highlighted in the text screen, to help the student notice what sounds he needs to work on. The user's responses are recorded and she can listen to any of them at any time. The program is thus a development of a record/playback model, with the added input of feedback in the form of a score from the system, extraction of the most serious deviation from the models, and the visual comparison of speech signals [4].

- Framework -

This study involved two groups of students, the control group taking a course in the fall of 2000, and the experimental group taking the same course the

following term, spring 2001. The experimental group received Talk to Me English (1) as supplemental courseware. The course was a 200-hour, ten-week course in Technical English for Immigrants offered at the Royal Institute of Technology (KTH) in Stockholm. One requirement of the course was five hours of individualized, tutor- and computer assisted pronunciation tutoring. In the fall term of 2000, students followed the normal course plan and were tested for the purposes of future comparison. Their five hours of pronunciation tutoring were assisted by a software program that helped the students learn IPA notation (Skytalk) so that they could use dictionaries more effectively. They did not receive their own pronunciation program. In the spring term of 2001, students were offered the opportunity to trade one hour's tutoring for a copy of Talk to Me for use on their home computers. They still received four hours of tutoring, using Talk to Me instead of the IPA program. The course content was generally the same for the two groups, but the teachers were different[4].

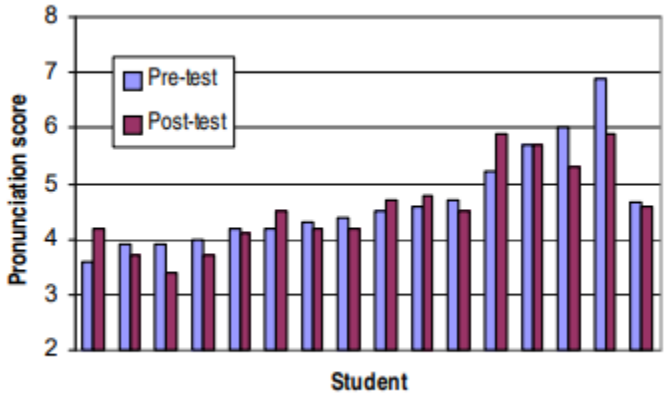

Figure 1.1.3:. Control group: Pre- and post-test scores in Pronunciation from PhonePass test.

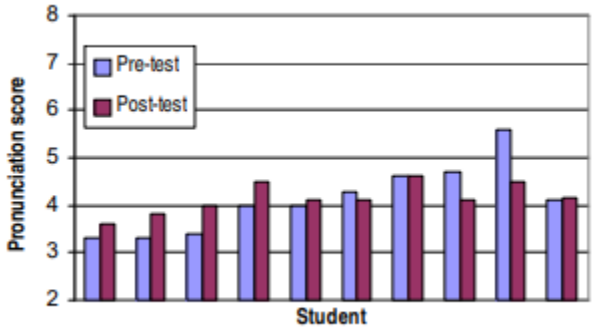

Figure 1.1.4 : Experimental group: pre- and post-test scores in pronunciation from PhonePass test.

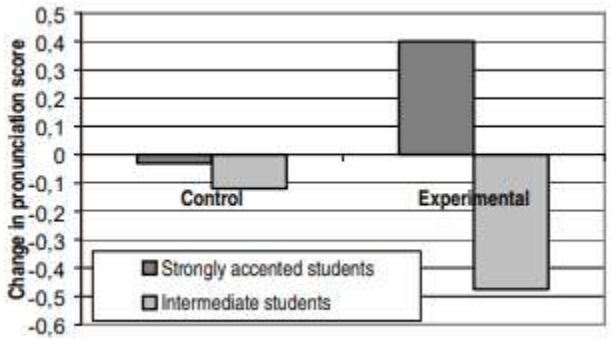

Figure 1.1.5: Change in pronunciation score for students according to beginning level. Weak students improved significantly in the experimental group. While all other groups showed no improvement.

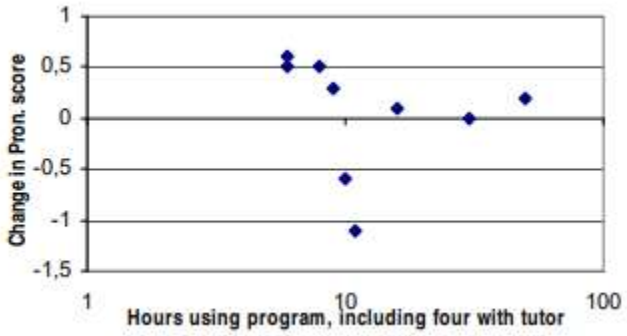

Figure 1.1.6: Relationship between times spent using program and change in pronunciation score on PhonePass test

- **Charl van Heerden, Etienne Barnard, Marelie Davel "Basic speech recognition for spoken dialogues"[5]**

  Spoken dialogue systems (SDSs) have great potential for information access in the developing world. However, the realization of that potential requires the solution of several challenging problems, including the development of sufficiently accurate speech recognizers for a diverse multitude of languages. We investigate the feasibility of developing small vocabulary speaker-independent ASR systems designed for use in a telephone-based information system, using ten resource scarce languages spoken in South Africa as a case study. We contrast a cross-language transfer approach (using a well-trained system from a different language) with the development of new language-specific corpora and systems, and

evaluate the effectiveness of both approaches. We find that limited speech corpora (3 to 8 hours of data from around 200 speakers) are sufficient for the development of reasonably accurate recognizers: Error rates are in the range 2% to 12% for a ten word task, where vocabulary words are excluded from training to simulate vocabulary-independent performance. This approach is substantially more accurate than cross-language transfer, and sufficient for the development of basic spoken dialogue systems [5].

Important challenges to address in this regard are the following:

• The design of spoken interfaces that are usable and friendly in diverse cultures, by users with limited or no computer literacy [5].
• The development of speaker-independent automatic speech recognition (ASR) systems that function reliably in the local languages of the developing world [5].
• The development of text-to-speech (TTS) systems that are easily understood in these same languages [5].

Language	code	# total minutes	# speech minutes	# distinct phonemes
isiZulu	zul	525	407	46
isiXhosa	xho	470	370	52
Afrikaans	afr	213	182	37
Sepedi	nso	394	301	45
Setswana	tsn	379	295	34
Sesotho	sot	387	313	44
SA English	eng	304	255	44
Xitsonga	tso	378	316	54
siSwati	ssw	603	479	39
Tshivenda	ven	354	286	38
isiNdebele	nbl	564	465	46

Table 1.1.1: The official languages of South Africa, their ISO 639- 3:2007 language codes, and the amount of speech contained in the Lwazi corpus

Language	% Corr	% Acc	Ave # phones	Phone ppl
Afrikaans	71.76	63.14	16.55	14.45
SA English	62.51	54.26	14.61	15.80
isiNdebele	74.21	65.41	28.66	10.26
isiXhosa	69.25	57.24	17.79	10.67
isiZulu	71.18	60.95	23.42	11.20
Tshivenda	76.37	66.78	19.53	9.99
Sepedi	66.44	55.19	16.45	11.54
Sesotho	68.17	54.79	18.57	10.40
Setswana	69.00	56.19	20.85	11.15
siSwati	74.19	64.46	30.66	10.38
Xitsonga	70.32	59.41	14.35	10.34
NTIMIT	64.07	55.73		

Table 1.1.2: Phone-recognition correctness ("Corr") and accuracy ("Acc") achieved for each of the languages in the Lwazi corpus. "Ave # phones" refers to the average number of occurrences of each phone for each speaker, and the final column lists the phono tactic perplexity of each language in our corpus. NTIMIT results from are provided for comparative purposes.

Language	Lwazi models	Lwazi eng model	Ntimit	WSJ
isiZulu	90.53	80.00	37.57	69.19
isiXhosa	95.29	77.78	34.34	61.28
Afrikaans	96.11	90.35	60.36	79.15
Sepedi	89.49	83.72	54.91	43.41
Setswana	87.66	76.95	39.02	52.09
Sesotho	97.14	79.48	30.65	50.65
SA English	91.94	91.94	82.86	81.95
Xitsonga	97.90	77.58	54.99	60.60
siSwati	96.62	77.01	46.46	61.09
Tshivenda	97.74	66.37	57.56	52.14

Table 1.1.3: Small vocabulary word recognition accuracies for 10 languages. Each system is required to distinguish between ten different semantic categories, with each category represented by one to three different lexical items.

- **Rebecca Hincks "Using speech recognition to evaluate skills in spoken English"[6]**

This paper analyzes some of the results of the use of PhonePass, a telephone-based test of spoken English that uses automatic speech recognition. It finds that the test provides sensitive measures of speech rate and phonetic accuracy.

- The PhonePass Test

The PhonePass test uses automatic speech recognition to assess facility in spoken English. It is designed as a simple way for organizations to test the English skills of potential employees or students. To administer the test, an organization purchases test papers from Ordinate Corp. Each test paper is unique, though the items are recombined to make other tests. To take the test, the examinee calls a phone number in California and is connected with a computer. The examinee enters his test paper number on the telephone keypad and then follows instructions. The test results are soon available on the company website. The test consists of five parts. This paper concerns results derived from Part A, Reading, where the examinee is instructed to read a set of sentences. The recognition engine can assess how the examinee's pronunciation of each word in the sentence compares with acceptable pronunciations, and measure the rate at which

the examinee reads. The speech processing used by Ordinate Corporation was trained with the speech of native speakers and adapted for use by non-native speakers. It uses forced alignment to locate the relevant parts of the speech signal, an HMM-based speech recognizer, a pronunciation dictionary, and an expected-response network constructed from responses collected in over 4000 administrations of the test (Bernstein 1999) [6].

- Reading Fluency

  In Part A, Reading, the examinee is asked to read aloud eight of the twelve sentences on the test paper. Data gathered from this part of the test determine the scores for two subscores: Reading Fluency and Pronunciation (Townshend 1998) [6].

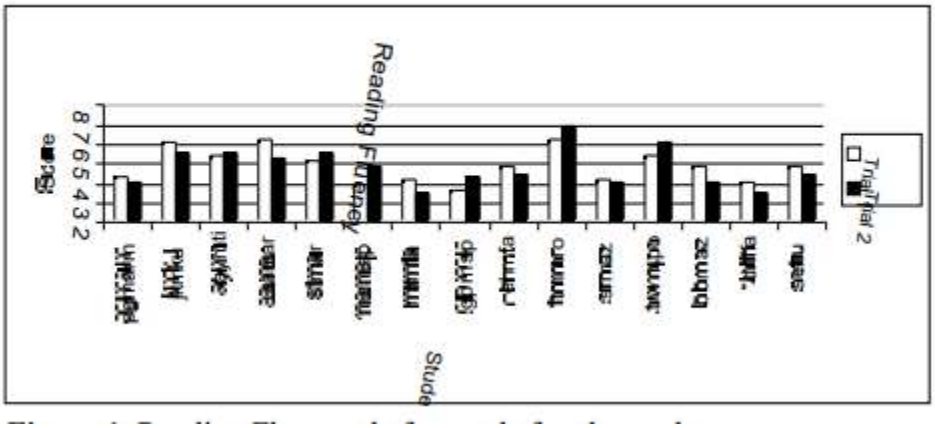

Figure 1.1.7: Reading Fluency, before and after, by student.

The lowest score shown in Figure 1 was Mrmfa's second trial. The highest result was Tmmro's second trial. Coincidentally, these two tests contained three sentences in common. Examinee Tmmro, with a score of 6.9, read these sentences twice as quickly as Mrmfa, who had a score of 3.5. Table 1 shows the sentences and the speed at which they were read [6].

Sentence	Tmmro2 (6.9)	Mrmfa2 (3.5)
"It's really expensive, but his friends eat there a lot."	4.06 seconds	7.92 seconds
"He gives them a pretty big discount."	2.05 seconds	3.89 seconds
"And they, in turn, always leave him a generous tip."	3.66 seconds	7.42 seconds
Total	9.77 seconds	19.23 seconds

Table 1.1.4: Reading speed for selected utterances, best and worst scoring students

- Pronunciation
  Part A of the PhonePass test also provides data for the Pronunciation subscore. Students' before and after results on the pronunciation subscore are shown in Figure 2.

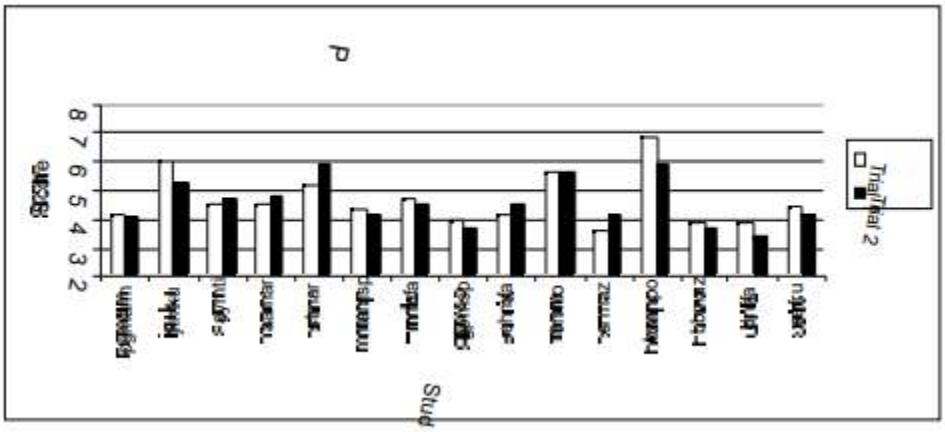

Figure 1.1.8 : Pronunciation subscore, before and after, by student.

Bwmpo's Pronunciation score can have been pulled down by the fact that the second set of sentences contained more words that he had particular problems with. The positive effect of the more natural prosody present in the second trial is not reflected in this score, though it can be seen in Figure 1 that his Reading Fluency score increased by more than half a point [6].

Error	First trial (score 6.9)	Second trial (score 5.9)
devoicing final /z/, /d/, /v/	9 out of 11 possible	13 out of 15 possible
/h/>[x]	2 out of 5 possible	7 out of 10 possible
Total (for these phonemes)	11 out of 16 possible	20 out of 25 possible

Table 1.1.5: Bwmpo errors

1.2 Research gap

Categories	**ENGLISH SCHOOL MATE**	**Mondly app**	**ELSA speak**

	(Spoken English module)		
Easy to learn	✓	✓	✓
Interesting lessons	✓	✓	✓
Practice conversations	✓	✓	✓
Practice poems	✓	✗	✗
Practice speeches	✓	✗	✗
Scoring	✓	✗	✓
Error checking	✓	✗	✓

Table 1.2.1 : Research gap

1.3 Research Problem

Nowadays knowing English language increases your chances of getting an honest job during a multinational company within your home country or for locating work abroad. it is also the language of international communication, the media and therefore the internet, so learning English is very important for socializing and entertainment additionally as work. So for that reason English learning is so important. English may be a language. Language plays a very important role in communication. The quality methods of communication are speaking or writing by a sender and listening or reading the receiver. Most communication is oral, with one party speaking and listening. So, the spoken English category is more important within the English.

Therefore in Sri Lanka, English language has been one of the main subjects in the school syllabus. For that reason, everyone can start learning English from their school time. Sri Lanka's current education system has been divided as figure 1.3.1. According to that chart I found the best time period which stage is the most suitable to start and improve the Basic spoken English properly. The stage of junior secondary education is more suitable to it because it's between

primary education and senior secondary GCE O-Level. So students can improve their English education before the O-Level examination.

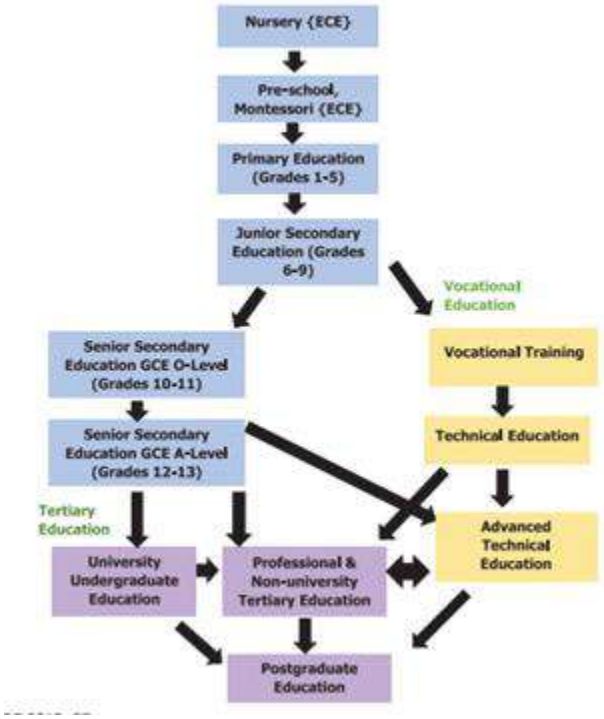

Figure 1.3.1 : Education stages of the student

So grade 6 is the starting point in the junior secondary education. That's the proper grade to teach English in proper way because they have some knowledge about the English language from the primary level. Good start is so helpful to the higher results so if teachers can properly teach the lessons and basic concept proper it's helpful to learn anything in the future path but in case today have some problems in this basic education in English language. It has so many reasons. Mainly in this case we can divide students in

two groups that rural areas student and urban areas. In the urban area student normally in good level than the rural areas.

Literacy Level	Urban			Rural		
	Sinhala	Tamil	English	Sinhala	Tamil	English
non-literacy	4%	55%	23%	5%	85%	47%
quite literacy	8%	2%	10%	13%	7%	24%
fairly literacy	6%	3%	7%	19%	0%	11%
very good literacy	75%	38%	53%	46%	2%	5%
ability to read & write	3%	1%	4%	16%	5%	13%
reading only	4%	1%	3%	1%	1%	0%
Total	100%	100%	100%	100%	100%	100%

Table 1.3.1: language literacy between urban and rural areas

According to the table 1.3.1 rural area student literacy level lower than the urban area students. Another one is the lack of teaching source. Because most of the time best teacher sources are lower in the rural areas and some problems have occurred as physical distance of students to school, difficulty finding teachers interested in relocating, poverty and physical teaching methods as well as the poor internet connection and lack of technically knowledge.

So another way the both rural and urban areas student mostly like to learn any subject with modern methods than traditional methods. They not interested with learning with books and notes. But they interested with online platforms. In example most of the time parent use the YouTube and some educational sites. So they know the different of the interested of their children.

1.4 Research Objectives

Speech recognition could be a software invention that enables the user to interact with their mobile devices through speech. It's simply an application that permits a machine to single out words or phrases in an exceedingly language, thereafter it converts them to a machine-readable format. Now a days most of the time use the mobile phones to easy their lives. So in this research product targeted capture those users and their children and in this module mainly targeted the student knowledge with entertainment. So rural areas have some poor internet connection and the problem of less of devices. So this product mainly targets the projects of donating computer laboratories. Because the application can use the computers. Then they can reduce the lack of teachers.

1.4.1 Main Objective

The main objective of this research is to develop a hybrid solution to improve and evaluate the Spoken English, Written English, English Listening and English Reading abilities of the grade 06 students. Spoken lesson mainly focuses to student's pronunciation to improve, the module target to cover every spoken lessons of their English pupil book, improving and evaluating the Spoken English knowledge of the students

As a whole, the product was developed by taking the above mentioned four research areas as the main objectives. In addition to the main objective, several specific objectives were designed to increase the productivity and efficiency of the product while maintaining commercial quality.

1.4.2 Specific Objectives

The main focus of this document is on the Spoken English module which covers one specific and special aspect of English language. The other components are built assuming that the Spoken English module works correctly. Therefore to ensure that the made assumption is correct and to increase the quality of this component several specific objectives were designed based on the Spoken English module.

- Collection of lessons relevant to the grade 06 English syllabus.
- Categorization of lessons based on the identified aspects of Spoken English language.
- Selection of relevant types of questions and Construction of exercises based on the identified criteria.
- Maintaining the functional independence of the component to make sure that the final product can be switched accordingly.
- Implementation of a user-friendly environment to allow the users to operate the system with minimum knowledge to gain maximum Performance.

1.5 Audience

This research mainly focuses on grade 6 students and teachers. As well as the rural area students are special than the urban area students. Same way the teachers also that focuses those students.

The rural area student has some problems than the rural areas student so that the reason they focused mainly. But urban area student also can use those system and improve their knowledge.

2 Methodology

The solution proposed under the research topic is broken down into 4 major components in order to fulfill all the objectives required to cover the prevailing research problem.

1. Spoken English Module
2. Written English Module
3. Prediction and Visualization mentor Dashboard
4. Brain Development and Vocabulary Improvement Game

Within the Spoken English component the users are allowed to attempt the build in exercises provided under different categories as levels. Several versatile techniques are used throughout the exercise development process to improve the effectiveness of the product including conversations, dialogs, model speeches, poems and pronunciations. Further it is important to note that in some instances the questions are organized into lessons where a lesson contains questions of all the types.

The Spoken English module follows a unique method to load the questions into the front end which is known as fetched HTTP calls. The Hypertext Transfer Protocol (HTTP) is designed to enable communications between clients and servers. HTTP works as a request-response protocol between a client and server. Here the questions are stored on a remote database and fetched through HTTP calls. The Fetch API is basically a modern replacement for XHR; it was introduced in browsers recently to make asynchronous HTTP requests easier to do in JavaScript, both for developers and other APIs that build on top of Fetch.

The next most important feature implemented is the progress status and the score related to the component. Users will be able to see their progress, rank, scores and other statistics for spoken quizzes. The score calculation process is totally based on the identification of the answers based on the given voice output to the product. This is totally implemented using Google speech recognizer. Google has a great Speech Recognition API. This API converts spoken text into written text, briefly Speech to Text. You can simply speak in a microphone and Google API will translate this into written text. The API has excellent results for English language. The complete mechanism of the Spoken English module can be divided into 6 main phases. This can also be referred to as the speech recognition and answer validation process in technical terms.

- User click speak button
- Google speech recognizer module is configured to identified English and launched
- User speak and Google speech recognizer capture the audio
- Google speech recognizer process audio and output the identified text
- Identified text is displayed to the user
- Identified text is compared to the original text and marks are calculated

Speech-to-Text has three main methods to perform speech recognition. These are listed below:

- **Synchronous Recognition** (REST and gRPC) sends audio data to the Speech-to-Text API, performs recognition on that data, and returns results after all audio has been processed. Synchronous recognition requests are limited to audio data of 1 minute or less in duration.
- **Asynchronous Recognition** (REST and gRPC) sends audio data to the Speech-to-Text API and initiates a *Long Running Operation*. Using this operation, you can periodically poll for recognition results. Use asynchronous requests for audio data of any duration up to 480 minutes.
- **Streaming Recognition** (gRPC only) performs recognition on audio data provided within a gRPC bi-directional stream. Streaming requests are designed for real-time recognition purposes, such as capturing live audio from a microphone. Streaming recognition provides interim results while audio is being captured, allowing results to appear, for example, while a user is still speaking.

The diagram below illustrates the speech recognition and answer validation process for the Spoken English module.

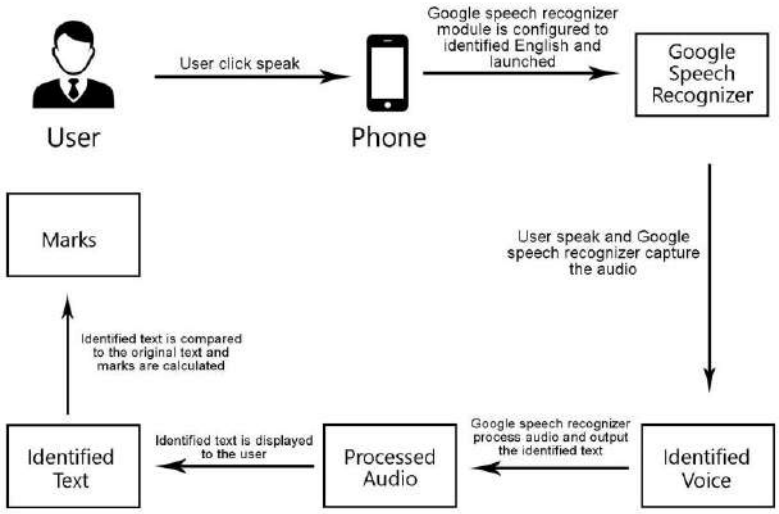

Figure 2.1 : spoken module architecture

2.1 Commercialization aspects of the product

This application is the educational application so that mainly focus the grade 6 students and their teachers. As well as can target their parent. Now a days online education is in the high place that reason of the corona situation. In the trend every path change the online base. There are paths of shopping, education, banking etc. changed the online systems after the corona season. Another one is student most probably not interested to the traditional methods to learn their subject. According to those reasons that educational system value goes high.

Introduced system has two parts that teaching module and the android application for the student. Its help to teacher that the add lessons their children and view the student evaluation. It's best reason and best point to market the product. And the other thing rural areas has the challenge of the lack of teachers so this product can be introduce them to resolve their problem. They can use the app teaching the English from one teaching module to many students.

2.2 Testing & Implementation

The main implementation of the Smart Student system has 2 major aspects as the Authentication segment and the Backend and Database. The Authentication segment is implemented using Firebase Authentication as a service. This is composed of a Sign-in screen and a Sign-Up functionality. Further, a Navigation Drawer update option and a user profile picture functionality are added to make the Authentication process complete. The backend and the database are implemented as a REST API using Java Spring-Boot and MongoDB as the database. Moreover, CRUD endpoints for spoken questions are also implemented. All the analyses of the question results will be performed here.

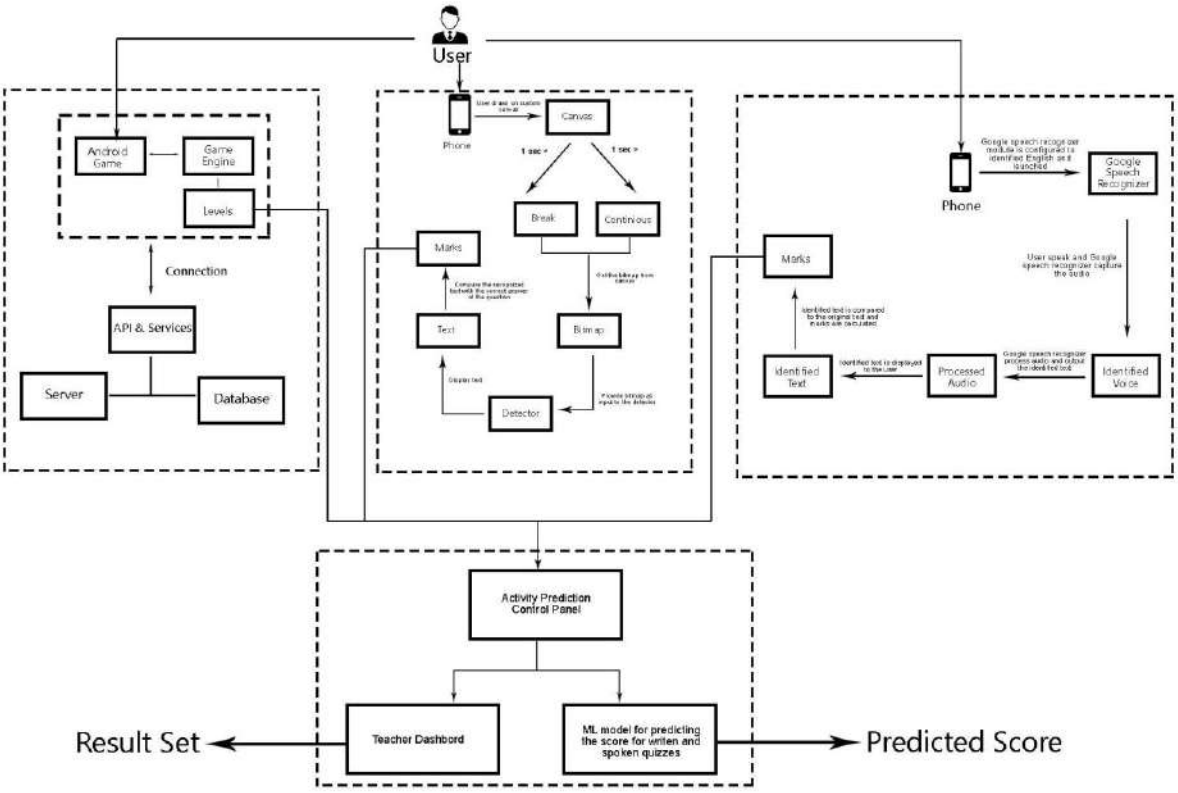

Figure 2.2.1: High Level Architecture

The android application architecture differs from the system of the Smart Student solution. The Activity sector handles UI and the rendering logic. The View Model handles business logic and change states of live data based on results from operations. The Live data notifies observers (activities/fragments) when the state changes. The Repository connect the backend and android application. It also Fetch and submit data from and to the backend. The diagram below shows the architectural design of the Android Application.

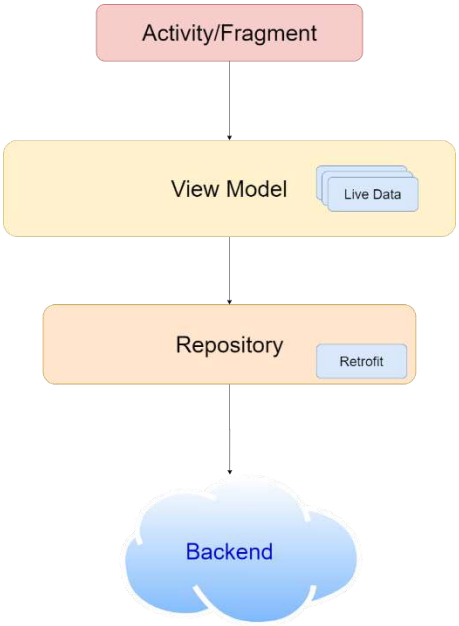

Figure 2.2.2: Android Application Architecture

Activity: Handles UI and rendering logic

View Model: Handles business logic and change states of live data based on results from operations

- Live data: notifies observers (activities/fragments) when state changes

Repository: Connect backend and android application

- Fetch and submit data from and to backend

Spoken module implemented using Google speech recognizer and there are three types of questions. Those are the

- Speech
- Poem
- Dialogue

Audio and video files are fetched from the Google Drive. Scores and question numbers will be displayed per question. Questions are organized into lessons and a lesson will contain questions of all the types [may be changed into one lesson containing questions from only one type]

3 Evaluating the user experience of the interfaces

Figure 3.1: English school mate home page

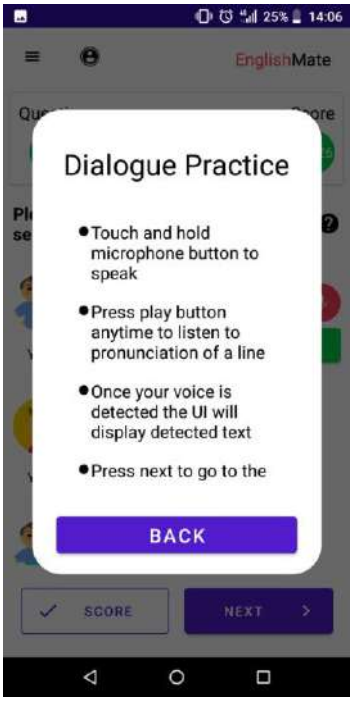

Figure 3.2: English school mate help page

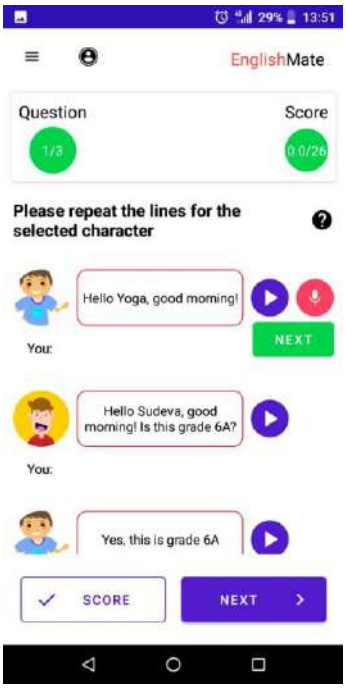

Figure 3.3: English school mate dialogue activity page

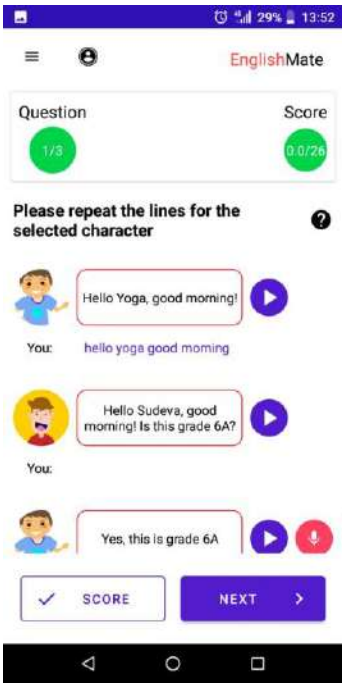

Figure 3.4: English school mate working dialogue activity

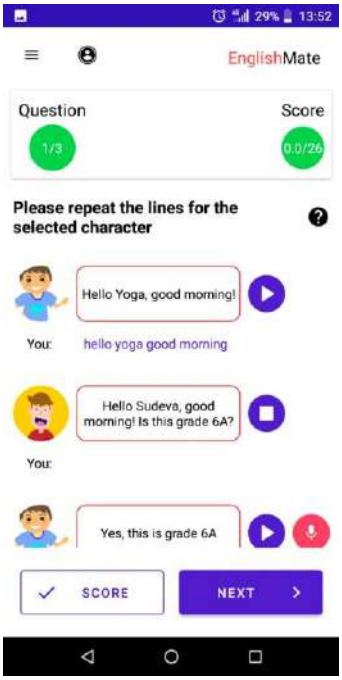

Figure 3.5: English school mate dialogue activity, playing audio of the other part

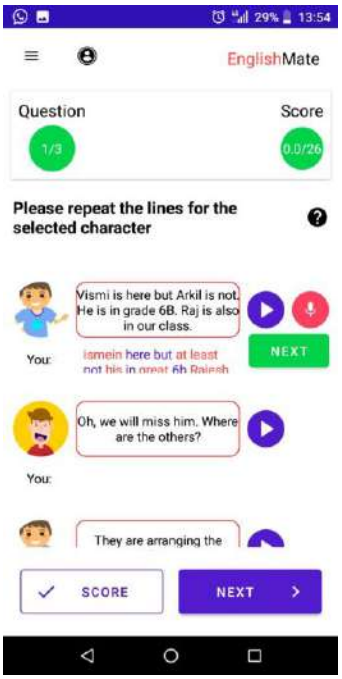

Figure 3.6: English school mate dialogue activity, display error

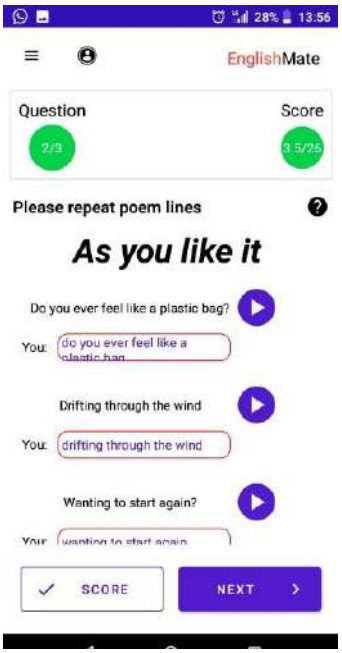

Figure 3.7: English school mate poem activity, reading session

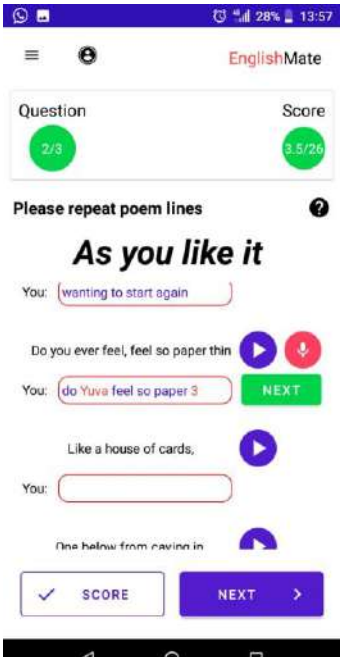

Figure 3.8: English school mate poem activity, display error

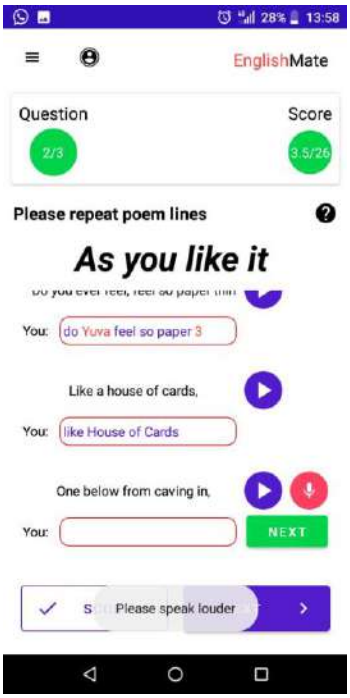

Figure 3.9: English school mate poem activity, view validate message 1

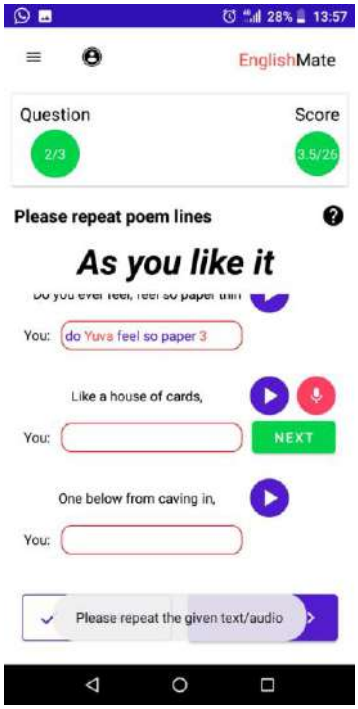

Figure 3.10: English school mate poem activity, view validate message 2

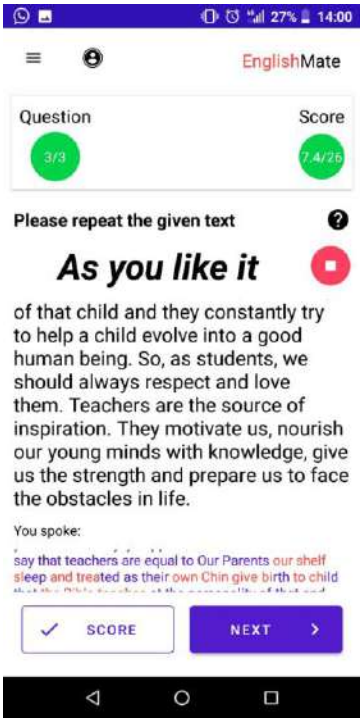

Figure 3.11: English school mate speech activity, playing record button and record

Figure 3.12: English school mate speech activity, view errors

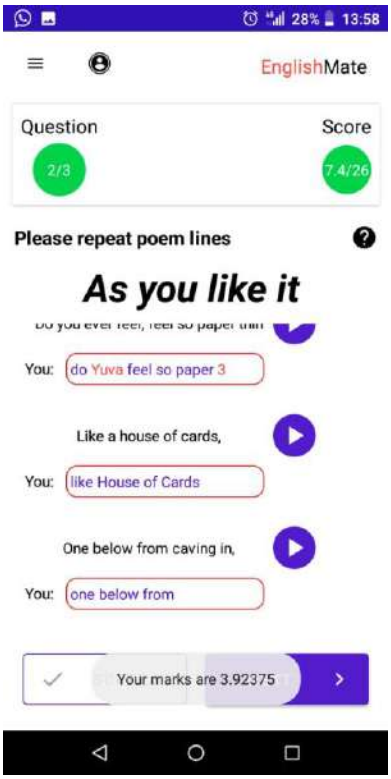

Figure 3.12: English school mate, view activity score from whole lesson

Figure 3.12: English school mate, view lessons score

4 Results & Discussion

4.1 Results

Test ID	01
Test Case Scenario	Navigate to the speech module page after log in to the system
Test input data	Sapnadivyanjali221@gmail.com Sapna1996
Test Steps	1. Successfully log in to the system 2. Press the 'Speech Quiz' button 3. Navigate to the lesson dashboard of the speech module
Expected Outcomes	Successfully user can navigate to the speech module
Actual Outcomes	As expected
Test results	Pass

Table 4.1.1: Test case 01

Test ID	02
Test Case Scenario	Verify user can read and record the selected character's dialogue sentences while pressing the record button.
Test input data	
Test Steps	1. Long press the record button until stop the reading. 2. Read aloud 3. Stop the button press after stop the reading

Expected Outcomes	Successfully user can read and record the dialogue
Actual Outcomes	As expected
Test results	Pass

Table 4.1.2: Test case 02

Test ID	03
Test Case Scenario	Verify user can automatically listen the other character voice after read the own character part.
Test input data	
Test Steps	1. Read and record the own character part 2. Press next button
Expected Outcomes	Successfully user can automatically listen the other character voice after reading own part.
Actual Outcomes	As expected
Test results	Pass

Table 4.1.3: Test case 03

Test ID	04
Test Case Scenario	Verify user can get the score of the activity from the whole lesson

Test input data	
Test Steps	1. Finished the activity 2. Click the score button 3. Display the score of the activity
Expected Outcomes	Successfully user can get the score for the activity from the whole lesson.
Actual Outcomes	As expected
Test results	Pass

Table 4.1.4: Test case 04

Test ID	05
Test Case Scenario	Verify display the error pronouncing word with highlighted to the user.
Test input data	
Test Steps	1. Read and record the wrong word in the sentence part
Expected Outcomes	Successfully user can view the error pronouncing word with highlighted
Actual Outcomes	As expected
Test results	Pass

Table 4.1.5: Test case 05

Test ID	06
Test Case Scenario	Verify user can navigate to the next activity after finished the dialogue activity
Test input data	
Test Steps	1. Finished the dialogue activity 2. Click the next button 3. Navigate the next activity from the dialogue activity
Expected Outcomes	Successfully navigate to the nest activity from the dialogue activity
Actual Outcomes	As expected
Test results	Pass

Table 4.1.6: Test case 06

Test ID	07
Test Case Scenario	Verify user can read and record the line by line in the poem activity
Test input data	Read and record poem lines
Test Steps	1. Long press the record button 2. Read aloud the line 3. Stop the read and release the record button
Expected Outcomes	Successfully user can practice the poem with read and record in line by line

Actual Outcomes	As expected
Test results	Pass

Table 4.1.7: Test case 07

Test ID	08
Test Case Scenario	Verify user can listen the poem line by line
Test input data	
Test Steps	1. Click the play button in front of the poem line
Expected Outcomes	Successfully user can listen the poem line by line
Actual Outcomes	As expected
Test results	Pass

Table 4.1.8: Test case 08

Test ID	09
Test Case Scenario	Verify user can view the wrong pronouncing word in the poem line
Test input data	
Test Steps	1. Read and record with the error pronunciation

Expected Outcomes	Successfully user can view the wrong pronunciation word with highlighting
Actual Outcomes	As expected
Test results	Pass

Table 4.1.9: Test case 09

Test ID	10
Test Case Scenario	Verify user can view the score in the poem after finished the activity
Test input data	
Test Steps	1. Finished the poem activity 2. Click the score button
Expected Outcomes	Successfully user can view the score of the poem activity
Actual Outcomes	As expected
Test results	Pass

Table 4.1.10: Test case 10

Test ID	11
Test Case Scenario	Verify user can navigate the next activity from the poem

Test input data	
Test Steps	1. Finished the poem lesson 2. Click the next button
Expected Outcomes	Successfully user can navigate the to the next activity from the poem
Actual Outcomes	As expected
Test results	Pass

Table 4.1.11: Test case 11

Test ID	12
Test Case Scenario	Verify user can read and record the speech to practice the speech
Test input data	
Test Steps	1. Short press the record button 2. Speak aloud 3. Again click the record button to stop recording
Expected Outcomes	Successfully user can practice the speech that recording
Actual Outcomes	As expected
Test results	Pass

Table 4.1.12: Test case 12

Test ID	13
Test Case Scenario	Verify user can view the wrong pronouncing word with highlighting.
Test input data	Read with the wrong pronouncing
Test Steps	1. Click the record button 2. Speak aloud with wrong words 3.
Expected Outcomes	Successfully user can view wrong pronouncing word with highlighting
Actual Outcomes	As expected
Test results	Pass

Table 4.1.13: Test case 13

Test ID	14
Test Case Scenario	Verify the user can view the notification that user sound is very low
Test input data	Read with low sound
Test Steps	1. Click the record button 2. Read with low sound
Expected Outcomes	Successfully user can view the message that user sound is low.
Actual Outcomes	As expected result with the notifications of 'sorry I couldn't understand you' and 'speak louder'

Test results	Pass

Table 4.1.14: Test case 14

Test ID	15
Test Case Scenario	Verify the system drive user's path smoothly from the notifications.
Test input data	
Test Steps	1. Record with any wrong path
Expected Outcomes	Successfully system view the notifications that drive to user in the correct path
Actual Outcomes	As expected result with the notification of 'please repeat the given text/audio'
Test results	Pass

Table 4.1.15: Test case 15

Test ID	16
Test Case Scenario	Verify the user can view the whole score of the lesson after complete the dialogue, poem and speech activities.
Test input data	
Test Steps	1. Finished three activities 2. Click the score button

	3. Display score
Expected Outcomes	Successfully user can view the whole score of the lesson
Actual Outcomes	As expected
Test results	Pass

Table 4.1.16: Test case 16

4.2 Research Findings

- Use the backend part to connect the application and the web portal to improve the scalability of the application

- So lesson can be change day by day so then can change only the backend not the application. It's easy to data get to the teaching web portal.

- In the speech model using the speech recognition library but this mainly used to short forms. In this research handle the temporary array from tracking the record press button. Every 30 second library stopped and again started. But it not feel the user because that continuation handle from above mentioned temporary array.

- Use the 3T architecture, its helps to the reduce errors in the future update versions.

4.3 Discussion

The focus of this section is to explain how the implemented solution and the results generated are fulfilling the objectives discussed above. As discussed under the sections above, the system was developed as a mobile application to improve the spoken knowledge of the grade 6 students based on speech recognition. The results generated proved that the accuracy of the solution is 90% which can be considered as the best outcome expected by the research team.

5 Summary

Component	Task
	Designing the user interfaces
	Implemented using Google speech recognizer
	Get the data from the inputs and convert to text
	Compared the added audio and the input
	Error tracking the input
	Calculate marks
	Display the score with interesting view

Table 5.1: Summery of the module

6 References

[3] R. Hincks, "SPEECH RECOGNITION FOR LANGUAGE TEACHING AND EVALUATING: A STUDY OF EXISTING COMMERCIAL PRODUCTS".

[7] "https://www.newsfirst.lk/2019/09/21/english-education-for-students/".

[4] C. J. v. Heerden, "Basic speech recognition for spoken dialogues," Brighton, United Kingdom, 2009.

[1] "English Language Teaching in Rural Areas: A Scenario and Problems and Prospects in Context of Bangladesh," 2016.

[5] Y. Ono, "Construction of a Voice-based Asynchronous Communication System Utilizing Speech," 15th International Conference on Advanced Learning Technologies, 2015.

[2] V. F. Martins, "Using the recognition and speech synthesis to assist".

[6] R. Hincks, "Using speech recognition to evaluate skills in spoken English".