AI-enabled Intelligent Assistant to Improve Reading and Comprehension Skills in English Language

RP 24-25J-027

Project Proposal Report

A.P. Ranaweera - IT21182396

B.Sc. (Hons) in Information Technology Specializing in Software Engineering

Department of Computer Science and Software Engineering

Sri Lanka Institute of Information Technology
Sri Lanka

July 2024

AI-enabled Intelligent Assistant to Improve Reading and Comprehension Skills in English Language

(Phoneme-Level Speech Error Detection Module)

RP-24-25J-027

Project Proposal Report

Supervisor: Prof. Dasuni Nawinna

B.Sc. (Hons) Degree in Information Technology Specialized in Software Engineering

Department of Computer Science and Software Engineering

Sri Lanka Institute of Information Technology
Sri Lanka

July 2024

DECLARATION

We declare that this is our own work, and this proposal does not incorporate without acknowledgement 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 acknowledgement is made in the text.

Name	Student Number	Signature
A.P. Ranaweera	IT21182396	

The above candidate is carrying out research for the undergraduate Dissertation under supervision of the undersigned.

Signature of the supervisor
(Prof. Dasuni Nawinna)

Date
26 / 8 / 2024

ACKNOWLEDGEMENT

I extend my deepest gratitude to my supervisor, Prof. Dasuni Nawainna, and my co-supervisor, Mr. Jeewaka Perera, from the Faculty of Computing, for their invaluable guidance and unwavering support throughout my research journey. Their insightful feedback and constructive critiques were instrumental in refining my work and enhancing its overall quality.

I am also grateful to the CDAP team for their continuous assistance and encouragement during my studies. Their commitment to both teaching and research has been a significant source of inspiration for me.

My sincere thanks also go to our panel members for their thoughtful comments and suggestions, which have been vital in shaping the direction of my research.

Lastly, I wish to express my heartfelt appreciation to my fellow team members. Their support and collaborative spirit, along with our shared discussions and idea exchanges, have greatly enriched my understanding of the subject matter.

ABSTRACT

The proposed project aims to develop an innovative computer-aided pronunciation training (CAPT) system specifically designed for English language learners, with a focus on improving pronunciation accuracy through accurate phoneme-level analysis. This system will leverage the power of powerful voice recognition technologies, such as CMU Sphinx, to methodically translate spoken words into a sequence of phonemes, allowing the discovery and correction of pronunciation mistakes in real time. Users will be given terms that vary in complexity, allowing for a progressive learning process. The algorithm will assess their pronunciation, detect faulty phonemes, and offer quick feedback to optimize their learning experience.

To further help learners, the system will use Large Language Models (LLMs) to produce practice words with comparable phonetic patterns, offering targeted exercises to encourage proper pronunciation. Both methods of phoneme recognition and LLM-driven content production address a fundamental gap in current educational technologies, which typically fail to provide comprehensive, phoneme-specific feedback and lack adaptable training routes tailored to individual user needs. Through this project, the efficiency of integrating phoneme identification with AI-powered content production will be studied, aiming to develop a tailored and dynamic learning environment that greatly boosts English pronunciation abilities for non-native speakers.

Keywords: Computer-Aided Pronunciation Training (CAPT), Phoneme Analysis, Speech Recognition, Pronunciation Error Detection, CMU Sphinx, Large Language Models (LLMs), Pronunciation Feedback, English Language Learning

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

Abbreviation	**Full Term**
IPA	International Phonetic Alphabet
CAPT	Computer-Aided Pronunciation Training
ASR	Automatic Speech Recognition
LLM	Large Language Model
UI	User Interface
UX	User Experience
DB	Database
API	Application Programming Interface
UAT	User Acceptance Testing
AWS	Amazon Web Services,
AI	Artificial Intelligence
CSS	Cascading Style Sheets

1. INTRODUCTION

1.1 BACKGROUND & LITERATURE SURVEY

English has become an essential global language not only for business and technology, but also for education, science and international communication. As a result, proficiency in English, especially proficiency in spoken communication, is increasingly seen as a key skill for personal and professional development. However, achieving correct pronunciation remains a significant obstacle for many non-native English speakers.

Mispronunciations can lead to misunderstandings, impacting both the speaker's confidence and the listener's comprehension. This challenge is particularly evident in regions where English is taught as a second or foreign language, where learners may have limited exposure to native pronunciation patterns and may be influenced by the phonetic structure of their mother tongue.

The importance of correct pronunciation in mastering English cannot be underestimated. Pronunciation involves the correct production of phonemes, the smallest units of sound that distinguish one word from another [1]. For example, the difference in pronunciation between "ship" and "sheep" hinges on the accurate articulation of the vowel phonemes /ɪ/ and /iː/ [2]. Mispronouncing phonemes can alter the meaning of a word entirely, leading to communication breakdowns. Therefore, there is a growing recognition of the need for educational tools that can provide learners with targeted pronunciation training, focusing on the correct articulation of phonemes.

Traditionally, language teachers have made use of the phonetic alphabet, and activities, such as transcription practice, diagnostic passages, detailed description of the articulatory systems, recognition/discrimination tasks, developmental approximation drills, focused production tasks (e.g., minimal pair drills, contextualized sentence practice, reading of short passages or dialogues, reading aloud/recitation), tongue twisters, and games (e.g., Pronunciation Bingo). Other trendy methods are listening and imitating, visual aids, practice of vowel shifts, and stress shifts related by affixation, and recordings of learner’s production. All these techniques are based on teachers having their students learn each sound and then apply them in real speech. Some students benefit from these techniques; however, others do not learn the pronunciation of the other language easily from them. For this reason, new techniques are being developed to supplement the learning of English pronunciation [3].

Computer-Aided Pronunciation Training (CAPT) systems have emerged as a promising solution to these challenges. By leveraging advances in speech recognition technology, CAPT systems can analyze a learner's speech in real-time and provide immediate feedback on their pronunciation. However, many existing CAPT systems are limited in their ability to provide detailed, phoneme-level feedback. Instead, they often focus on broader aspects of pronunciation, such as stress, rhythm, and intonation, without addressing the specific phoneme-level errors that are critical to achieving native-like pronunciation.

The proposed research seeks to address these limitations by developing a CAPT system that integrates advanced speech recognition technologies with phoneme-level analysis. The system will recognize spoken words into a sequence of phonemes and identify pronunciation errors at the phoneme level. To achieve this, the system will use the grapheme-to-phoneme model and a suite of speech recognition tools, among other approaches.

In addition to phoneme-level analysis, the proposed CAPT system will incorporate large language models (LLMs) to generate personalized training words for learners, according to the mispronounced phoneme set. In the context of the proposed CAPT system, LLMs will be used to create practice words that are similar in phonetic structure to the words that the learner has difficulty pronouncing. This approach ensures that learners receive targeted practice on the specific phonemes they struggle with, thereby improving their pronunciation over time.

The Role of Computer-Aided Pronunciation Training (CAPT)

Computer-Aided Pronunciation Training (CAPT) plays a significant role in helping language learners improve their pronunciation through interactive, technology-driven methods. CAPT systems provide individualized, instant feedback using advanced technologies like automated speech recognition (ASR), allowing learners to focus on phonemes, intonation, and stress patterns effectively. CAPT has proven particularly beneficial for non-native speakers, as it offers a cost-effective and scalable way to enhance pronunciation skills compared to traditional methods [4].

Research shows that CAPT improves learners' pronunciation skills by helping them practice repeatedly and receive real-time corrections. Moreover, these systems are found to be highly motivating, especially when integrated into personalized learning environments [5]. CAPT tools also emphasize the importance of intelligibility over perfection, aiding learners in achieving effective communication [6].

The proposed research seeks to develop a CAPT system that addresses these limitations by integrating phoneme-level analysis and real-time feedback. This system will utilize advanced speech recognition technologies to convert spoken words into their constituent phonemes, identify pronunciation errors, and provide immediate feedback on the specific sounds that need improvement. By focusing on the phoneme level, learners can target the precise areas where they struggle, rather than receiving generalized feedback.

Current Approaches to Pronunciation Error Detection and Feedback Mechanisms

Despite the availability of several digital tools aimed at improving pronunciation, few focus specifically on phoneme-level feedback. Popular language-learning platforms such as Duolingo, Rosetta Stone, BoldVoice and Linguacoach emphasize overall language skills such as vocabulary, grammar, and comprehension, while offering limited support for detailed pronunciation training [7] [8] [9] [10]. These platforms often evaluate spoken input at the word or sentence level, indicating whether the pronunciation was correct or incorrect, without identifying the specific phonemes that were mispronounced.

The journal of Pronunciation Error Detection and Correction [13] demonstrates a practical approach to real-time pronunciation error detection and correction. Combining speech recognition, the CMU Pronouncing Dictionary, and text-to-speech capabilities, the script offers a valuable tool for language learners and individuals aiming to enhance their spoken communication skills. While the focus of this implementation centers on individual word pronunciation, it sets the foundation for more advanced and comprehensive error detection and correction systems. This script addresses a critical aspect of language proficiency by providing immediate feedback and suggestions for improving pronunciation.

This technology operates by converting spoken language into text, leveraging speech recognition algorithms to transcribe audio input into words or phrases. It then utilizes the CMU Pronouncing Dictionary or similar linguistic resources to compare the pronunciation of these transcribed words with their expected phonetic representations. Additionally, text-to-speech capabilities are employed to provide immediate feedback to users. When a mispronunciation is detected, the system can suggest the correct pronunciation either through auditory feedback or by displaying the correct phonetic transcription [13].

The research paper [11] paper presents a method that automatically detects pronunciation error in learners' speech and generates corrective feedback. The methods target at a very common use case in CAPT. Learners try to imitate a sentence after they listen to the gold standard and wait for the system to tell them if they pronounce good enough. After training with annotated data, their system can detect phoneme errors like deletion, insertion, substitution and distortion with high accuracy, and provides feedback that could significantly help learners to correct their errors. The model, which they trained with only voice data from 10 learners, already has good performance. In industrial usage, if learners allow their voice data to be collected, a more capable model can be expected.

This article [12] proposes an improved random forest model and applies it to pronunciation error detection and correction in English teaching to use artificial intelligence technology to assist learners in detecting and correcting errors in spoken English pronunciation. The detection framework in this article primarily employs MFCC for feature extraction, while PCA is employed for feature data dimensionality reduction. When learners pronounce, the improved RF algorithm classifies and detects pronunciation errors caused by nonstandard position, action, and pronunciation duration of pronunciation-related organs. The experimental design demonstrates that a combined classification framework based on MFCC, PCA, and RF can clarify the learner’s pronunciation problem, making it possible to provide feedback and correction opinions for various error types. During the experiment, we discovered that the method of multifeatured fusion may improve feature extraction performance, because the feature extraction effect of WPC is also very good. Further research in this study will focus on the fusion of multi-featured methods.

The paper [13] considering pronunciation errors as divided into accent and lexical errors and a methodology for detecting each is presented and evaluated. The paper is investigated in the context of three corpora, two on which humans were asked to annotate pronunciation errors, and one where they were asked to transcribe actual pronunciation. Results are consistent with accent and lexical errors being defined as distinct categories of error that can be detected separately. The system was successfully able to detect word-level accent and lexical errors on the latter corpus but not the former two. It was, however, able to diagnose lexical and general and specific accent error tendency with satisfactory performance across all three datasets. Analysis suggested that the annotators of the first two corpora were systematically under-annotating accent errors and that therefore the phonetic transcription technique is a superior method of annotation for error detection tasks [14].

Studies have consistently shown that learners benefit from targeted, specific feedback when mastering pronunciation. The absence of detailed phoneme-level guidance in most current tools creates a gap in language-learning resources that needs to be addressed. By focusing on phoneme recognition and error correction, a more effective system can be developed to help learners tackle their specific pronunciation challenges.

Phoneme Recognition and Error Identification in English Pronunciation

Phoneme recognition is a critical aspect of pronunciation training. Existing CAPT systems that use Automatic Speech Recognition (ASR) technology typically compare a learner’s spoken input to a predefined word model, delivering feedback at the word level. While this is useful for general language learning, it does not offer the granularity needed to correct specific sound errors. This limitation is particularly significant for English learners, where even small phoneme-level mistakes can change the meaning of a word [11].

The proposed system will provide advanced phoneme recognition algorithms to break down words into individual phonemes and identify errors at this level. The system will detect the error and provide corrective feedback. This precise, phoneme-level feedback allows learners to focus on correcting individual sounds rather than grappling with entire words or sentences.

Leveraging Large Language Models (LLMs) for Dynamic Word Generation

Large Language Models (LLMs) have revolutionized the way we generate dynamic and contextually appropriate words and sentences. By leveraging vast datasets and neural architectures, LLMs can generate words that fit specific contexts based on semantic understanding. These models, such as Gemini, are capable of adapting word generation to suit user requirements, allowing them to respond to nuanced prompts and deliver highly personalized content [15]. LLMs not only improve in terms of generative capabilities but also in understanding the user's input, making word generation more relevant and context aware. By adapting practice exercises based on learner performance, the system can offer a personalized learning experience that is more effective than traditional, static word lists. In the context of the proposed CAPT system, LLMs will generate practice words that are similar in phonetic structure to the words learners have difficulty pronouncing.

According to existing systems tend to focus on word- or sentence-level evaluations, offering limited support for correcting specific sound errors. The proposed CAPT system seeks to address this gap by providing real-time, phoneme-level analysis and personalized practice exercises generated by LLMs. By focusing on the smallest units of sound and offering targeted feedback, the system will enable learners to improve their English pronunciation more effectively and with greater confidence.

1.2 RESEARCH GAP

The development of English pronunciation tools has seen considerable advancements over the past few years, with popular platforms such as Duolingo, Rosetta Stone, and Babbel providing learners with general feedback on their spoken English. However, a significant gap still exists in terms of providing phoneme-level analysis and feedback. Current tools primarily focus on word or sentence-level feedback, leaving users unaware of the specific phoneme or sound where their pronunciation falters. This lack of detailed feedback limits the effectiveness of these tools for non-native speakers, who require precise corrections to refine their pronunciation skills.

Furthermore, existing platforms often offer generic feedback and do not adapt to individual learner needs. For instance, while tools may provide users with an indication of mispronounced words, they fail to offer targeted exercises that focus on the specific sounds that need improvement. The absence of adaptive learning mechanisms, such as generating similar words with the same phonetic structure for additional practice, leaves learners without sufficient resources to effectively address their pronunciation challenges.

In addition, most of the widely used pronunciation platforms rely on fixed content and do not leverage advanced technologies like Large Language Models (LLMs) to dynamically generate new practice material based on user performance. As a result, learners may not be exposed to diverse words that challenge their weak phonemes, leading to slower progress in mastering accurate pronunciation.

The proposed system seeks to address these research gaps by introducing a phoneme-level feedback mechanism, real-time error detection, and adaptive learning. The integration of LLMs allows for the generation of similar words that target specific phoneme errors, creating a personalized and dynamic learning path. This system provides a more efficient way for users to identify, understand, and correct their pronunciation errors, something that is largely missing from the current landscape of pronunciation tools.

Table 1-1 Comparison of Similar Products

Research paper / Tools	Phoneme level error Detection	Phoneme level feedback	Similar words generation for practice	Focus on phoneme level practice	Use of large language model (LLMs)
Duolingo	✕	✕	✕	✕	✕
Rosetta Stone	✕	✕	✕	✕	✕
BoldVoice	✕	✕	✕	✕	✕
Linguacoach	✕	✕	✕	✓	✕
Research paper A [11]	✕	✕	✕	✕	✕
Research paper B [12]	✕	✕	✕	✕	✕
Research paper C [13]	✕	✕	✕	✕	✕
Research paper D [14]	✕	✕	✕	✕	✕
Purposed system	✓	✓	✓	✓	✓

The comparison table 1.1 clearly illustrates the differences between existing systems and the proposed platform. They focused on providing feedback at the word or sentence level. While these platforms help users understand whether a word pronounced correctly, they do not identify the specific sound (phoneme) that causes the mistake. In contrast, the proposed system offers phoneme-level feedback, enabling users to pinpoint and correct errors with greater precision.

Another key gap addressed by the proposed system is real-time feedback. Unlike current platforms that only provide feedback after the user completes speaking a word or sentence, the proposed solution gives immediate corrections, allowing users to adjust their pronunciation on the spot.

Moreover, while existing tools lack adaptive learning features, the proposed system dynamically generates similar words with the same phonetic patterns using LLMs. This allows users to practice problematic sounds through targeted exercises that directly address their weaknesses.

In summary, by using these technologies, the proposed research will provide a more effective tool for improving English pronunciation skills.

1.3 RESEARCH PROBLEM

The challenge of mastering English pronunciation remains a significant barrier for non-native speakers, particularly in regions where English is not the primary language. Despite advances in language learning technologies, most current tools primarily focus on general language acquisition, offering limited emphasis on phonetic precision and individualized feedback. This lack of tailored pronunciation guidance often leaves learners struggling with specific sounds, resulting in persistent errors that hinder effective communication and reduce learner confidence. The problem is further compounded by the scarcity of accessible, user-friendly systems capable of delivering detailed phoneme-level analysis and adaptive feedback, especially those that can evolve with a learner's progress and needs.

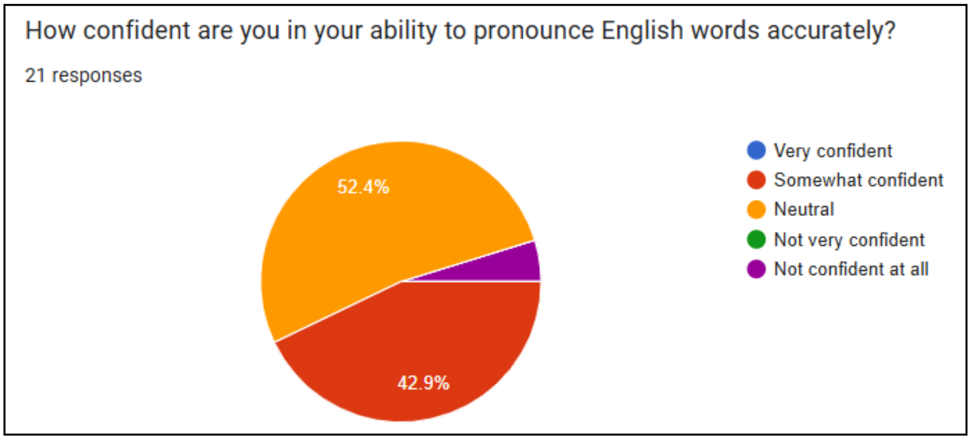
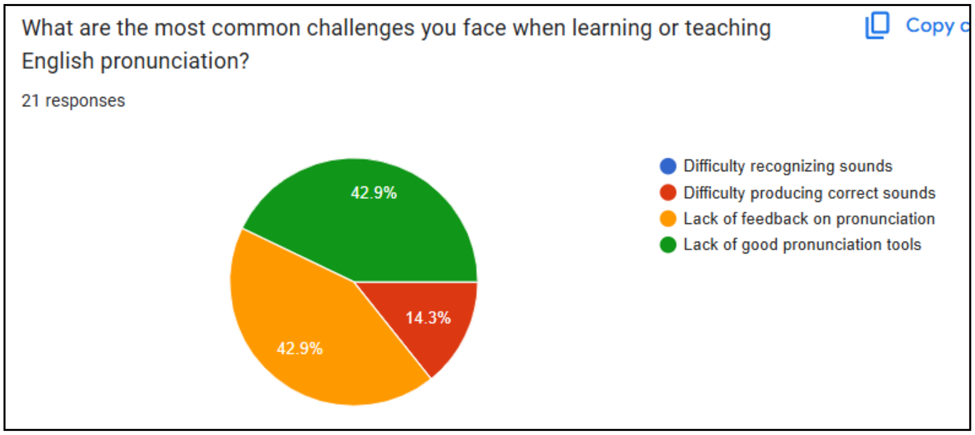


Figure 1-1 User Confidence in English word pronunciation.

The responses (Figure 1.1) indicate that most participants do not feel fully confident in their ability to pronounce English words accurately. A significant portion expressed only moderate confidence, while others remained neutral or unsure about their pronunciation skills. This suggests that learners are aware of their limitations and may benefit from tools that offer more focused support. It highlights the need for pronunciation training solutions that build user confidence through clear, phoneme-specific feedback and guided practice, rather than relying solely on generic or binary evaluations.

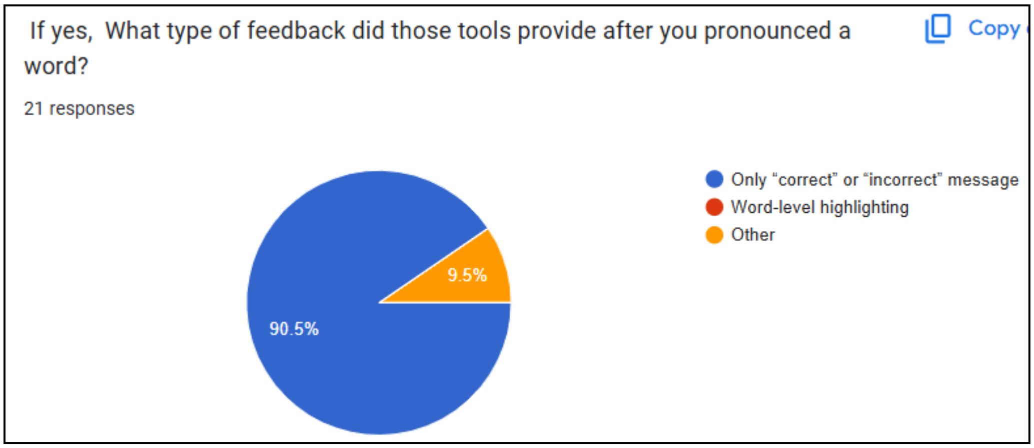


What are the most common challenges you face when learning or teaching English pronunciation?

21 responses

Figure 1-2 Challenges facing English pronunciation

The responses (Figure 1.2) highlight two major challenges learners face in mastering English pronunciation: the lack of effective pronunciation tools and the lack of useful feedback on their pronunciation efforts. These issues are just as common as the difficulty in producing correct sounds, showing that learners are not only struggling with speech itself but also with the quality of support available to them. This emphasizes the need for advanced learning systems that provide meaningful, targeted feedback and offer practical tools designed to address individual pronunciation challenges.



If yes, What type of feedback did those tools provide after you pronounced a word?

21 responses

Figure 1-3 Pronunciation Feedback Types

The chart (Figure 1.3) shows that most learners using pronunciation tools receive very basic feedback, typically limited to a simple "correct" or "incorrect" message. This kind of evaluation lacks depth and fails to inform the user about what specifically went wrong. Only a few reported receiving more detailed responses such as word-level highlighting. This clearly indicates a gap in the quality of feedback provided by existing tools and highlights the need for systems that deliver more informative, phoneme-specific, and actionable feedback to help learners truly improve.

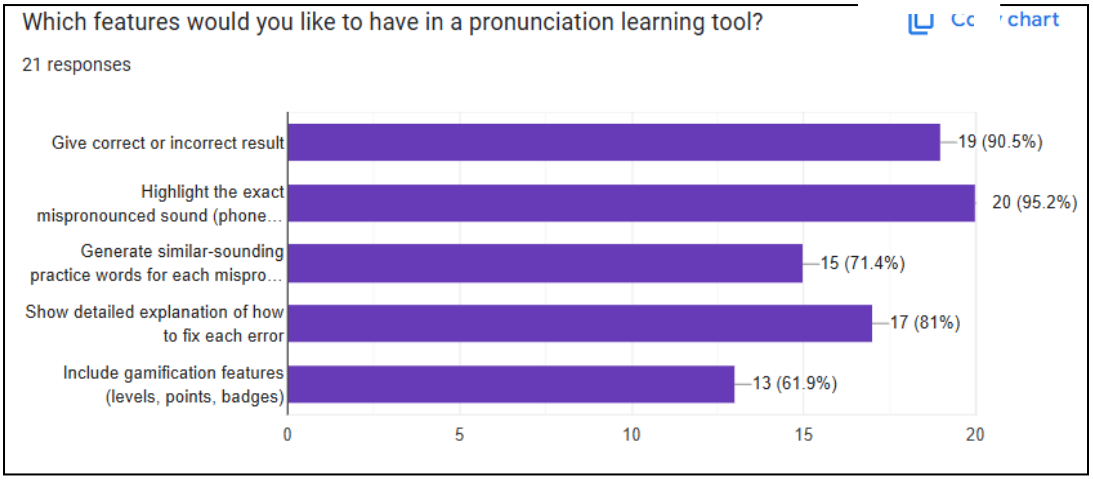


Figure 1-4 User Selected Features for Pronunciation Tool

The figure 1.4 responses clearly show that learners are not just looking for basic feedback they expect a more advanced, supportive, and personalized pronunciation learning experience. Most participants expressed interest in tools that can highlight the exact mispronounced sound within a word, explain how to fix that error, and provide similar-sounding words for further practice. Additionally, many also favor having motivational elements like gamification to keep the learning engaging. These preferences confirm a strong demand for a pronunciation system that offers phoneme-level analysis, real-time correction, and intelligent guidance, far beyond the capabilities of most existing tools.

2. OBJECTIVES

2.1 MAIN OBJECTIVE

To design and implement an intelligent speech error detection and feedback system that operates at the phoneme level. The goal is to enhance English pronunciation learning by analyzing the learner’s spoken input, accurately identifying mispronounced phonemes, and delivering personalized corrective feedback. This system aims to provide real-time, targeted guidance that helps learners understand their specific pronunciation errors and improve through focused practice. By leveraging phoneme-level analysis, the tool will offer a more precise and effective learning experience, especially for non-native speakers seeking to improve their spoken English clarity and confidence.

2.2 SPECIFIC OBJECTIVES (SUB OBJECTIVES)

1. **Phoneme Recognition and Analysis:**

   - Develop and integrate a phoneme recognition system using technologies such as CMU Sphinx and advanced speech recognition tools to accurately convert spoken words into phoneme sequences.
   - Implement algorithms to analyze the phoneme sequence of the learner's pronunciation and compare it against the correct phoneme sequence of the target word.
   - Identify and highlight specific phoneme-level pronunciation errors, providing detailed feedback on incorrect sounds.

2. **Dynamic Practice Generation:**

   - Create a system that generates practice words and sentences based on the learner’s pronunciation errors. This involves using Large Language Models (LLMs) to produce phonemically similar words for targeted practice.

- Design an adaptive learning path that adjusts the difficulty of practice words based on the learner's progress and error patterns, ensuring continuous and effective skill development.

3. **Personalized Feedback Mechanism:**

- Develop a feedback mechanism that provides real-time, personalized insights into the learner's pronunciation, including explanations of common phoneme errors and suggestions for improvement.
- Implement features that allow users to review their pronunciation history and track their progress over time, facilitating ongoing improvement and learning.

4. **User Interface and Experience Design:**

- Design an intuitive and user-friendly interface for the CAPT system, ensuring that it is accessible and engaging for users of varying proficiency levels.
- Incorporate features that enhance user experience, such as interactive exercises, and audio samples, to support effective learning and practice.

3. METHODOLOGY

3.1 SYSTEM ARCHITECTURE

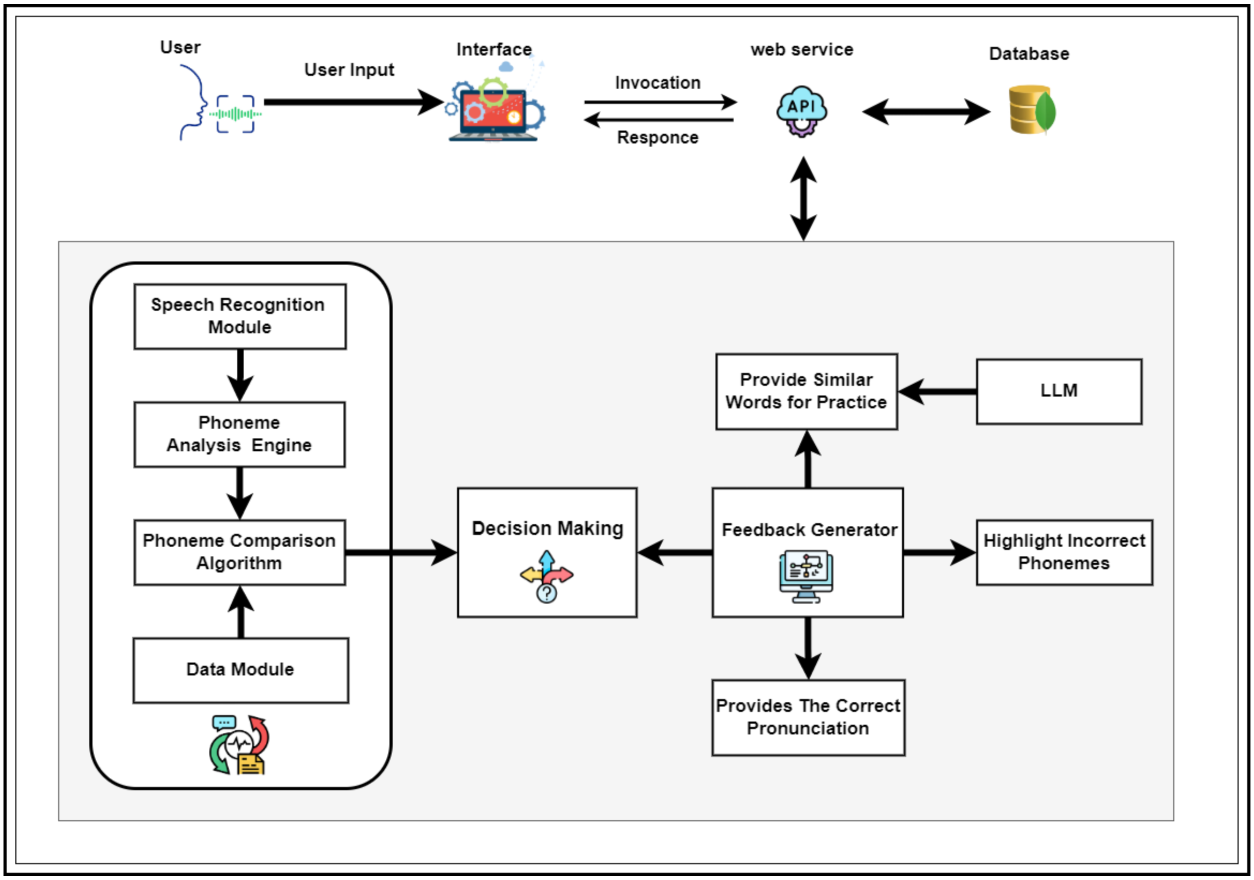

User
User Input
Interface
Invocation
Responce
web service
API
Database
Speech Recognition Module
Phoneme Analysis Engine
Phoneme Comparison Algorithm
Data Module
Decision Making
Feedback Generator
Provide Similar Words for Practice
LLM
Highlight Incorrect Phonemes
Provides The Correct Pronunciation

Figure 3-1 System Architecture

The methodology for this research component is designed to develop a system that helps English language learners improve their pronunciation by providing real-time phoneme-level feedback and personalized exercises. The system will leverage speech recognition technology, phoneme analysis algorithms, and large language models (LLMs) to generate corrective feedback and similar pronunciation exercises. The detailed workflow is explained based on the attached system diagram.

1. User Input and Interface Design

The user interacts with the system through an interface where they are prompted with words to pronounce. The system records the user's spoken input using a microphone. The interface is designed to be intuitive and user-friendly, allowing the user to easily navigate through the pronunciation exercises. The words displayed for pronunciation are carefully chosen based on the user's current proficiency level.

2. Speech Recognition and Phoneme Extraction

The user’s spoken word is first processed by the Speech Recognition Module, which converts the audio input into a sequence of phonemes. This module utilizes speech-phoneme technology with a focus on phonetic-level analysis. The system identifies and extracts the phonemes from the user's speech.

3. Phoneme Analysis and Comparison

Once the phonemes are extracted, they are passed to the Phoneme Analysis Engine. Here, the system compares the user's phonemes with the correct phoneme sequence for the target word. A Phoneme Comparison Algorithm is used for this purpose, which evaluates each phoneme in the spoken word and determines whether it was pronounced correctly or incorrectly.

The system uses a Data Module that stores phoneme sequences for each word, allowing for accurate phoneme-to-phoneme comparisons.

4. Decision-Making Process

The Decision-Making component of the system determines the next course of action based on the user's pronunciation accuracy. If the pronunciation is correct, the user moves to the next word or exercise. If an error is detected, the system identifies the incorrect phonemes and proceeds to generate feedback.

5. Feedback Generation and Error Highlighting

The Feedback Generator provides real-time corrective feedback to the user. If the user’s pronunciation is incorrect, the system highlights the specific phonemes that were mispronounced. This feedback is crucial as it allows users to focus on the exact areas where they make mistakes, thus improving their learning process.

In addition to highlighting errors, the system also provides the correct pronunciation of the word, guiding the user towards improvement.

6. LLM Integration for Similar Word Suggestions

The system uses Large Language Models (LLMs) to dynamically generate similar words that have the same phonetic pattern as the mispronounced phonemes. This provides the user with additional practice opportunities, allowing them to work on improving the specific phonemes they struggled with. Similar words are chosen based on their relevance to the incorrect phonemes, ensuring targeted practice.

7. Real-Time Pronunciation Feedback Loop

Once the user practices with the similar words provided, the system continues to evaluate their progress in real-time. If the user successfully corrects their pronunciation, the system progresses to more challenging exercises, adapting to the learner's skill level.

8. Data Storage and System API

All data, including the user's input, feedback, and progress, is stored in a Database using a cloud-based system. The web service API facilitates communication between the user interface and the backend, ensuring smooth interaction and data flow. The API handles requests for pronunciation evaluation and provides responses in real-time to the user interface.

9. Continuous Improvement and Adaptive Learning Path

The system monitors the user's progress over time, adapting the learning path to ensure continuous improvement. As the user masters certain phonemes, the system introduces more complex pronunciation tasks, helping the user build their pronunciation skills incrementally.

Table 3-1 Used technologies and techniques

Technologies	ReactJS, Python, Firebase database, LLM model, CMU Sphinx Dictionary
Techniques	Phoneme comparison Algorithm, Voice recognition API, Audio -phoneme model
Architecture	Microservices Architect
CI/CD	GitHub, Docker, Kubernetes

3.2 SOFTWARE SOLUTION

3.2.1 Software Development Life Cycle (SDLC)

The development of the proposed system follows the traditional Software Development Life Cycle (SDLC) process, ensuring a structured and phased approach to deliver the system effectively. The key stages include Requirements Gathering, Feasibility Study, Design, Testing, and addressing Constraints.

Requirements Gathering

Data Gathering

To ensure the effectiveness and relevance of the pronunciation training system, data collection was a foundational step. The following methods were employed,

- **Analysis of Language Learner Challenges**: Review of academic literature and existing studies on English pronunciation difficulties faced by non-native speakers.
- **Dataset Preparation**: Compilation of a dataset containing English words categorized by difficulty level (easy, medium, hard) and corresponding phoneme sequences (from sources such as CMUdict).
- **Phoneme Error Patterns**: Collection of common mispronunciations and phoneme-level errors specific to learner demographics.

Conducting a Survey

A structured survey was administered among English learners and language instructors to gain insights on:

- Current challenges in pronunciation improvement.
- Preferred methods for receiving corrective feedback.
- Familiarity and comfort with technology-based learning tools.
- Features learners expect in a pronunciation training platform.

The survey responses helped in defining user needs, validating system features (such as phoneme-level feedback and adaptive word selection), and shaping the user interface design.

Feasibility Study (Planning Phase)

Economic Feasibility

- Low-cost Infrastructure: The system leverages cloud services with pay-as-you-go models, open-source phoneme processing tools, and modern web development frameworks.
- Licensing: Open-source tools and educational licenses keep the budget manageable.
- Analyzing potential revenue streams (subscription fees, advertising)

Scheduled Feasibility

- Phase-Based Development Plan.
- Creating a project timeline with milestones and deadlines

Technical Feasibility

- Evaluating the availability of necessary technologies and tools
- Considering scalability and future-proofing the architecture

Design

Use Case Diagram

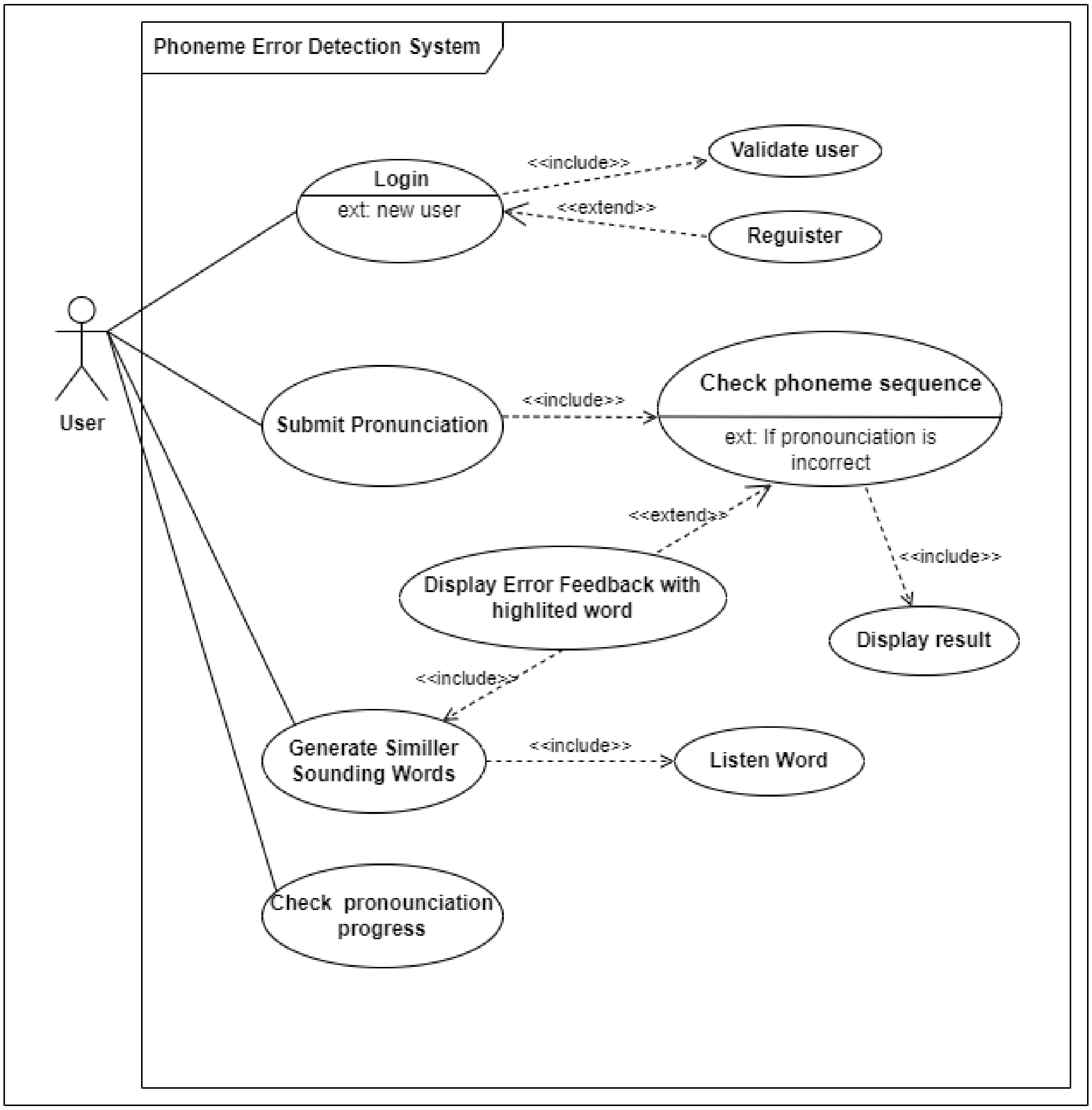


Figure 3-2 Use Case Diagram

Implementation

Audio Recording

The frontend interface allows users to record their pronunciation using a built-in microphone. JavaScript is used to capture the audio and send it to the backend through an API endpoint for analysis.

Audio to Phoneme Conversion

Once the audio is received by the backend, it is processed using a speech recognition system, which recognizes the spoken word. Then, the audio-to-Phoneme model is used to convert the word into its expected phoneme sequence. The spoken word is also aligned into phonemes using phoneme extraction tools.

Phoneme Comparison Algorithm

The phoneme sequence extracted from the user's speech is compared with the correct phoneme sequence using a algorithm. This helps detect the exact phonemes that were pronounced incorrectly, even if the speaking speed or length varies slightly.

LLM Integration

If a pronunciation mistake is found, a Large Language Model (LLM) such as Google Gemini is used to generate a list of similar practice words containing the mispronounced phonemes. These words are shown to the user for targeted practice, helping them correct specific sounds they struggle with.

This pipeline enables the system to give accurate, real-time, and personalized pronunciation feedback to users.

Testing

Unit Testing: Each module, such as audio recording, phoneme extraction, comparison algorithm, and LLM integration were tested individually. Sample audio files were used to validate that.

Integration Testing: End-to-end testing was performed to ensure smooth data flow between frontend and backend.

User Testing: The system was tested by a group of English learners with varying proficiency levels.

Performance Testing: Tests were conducted to ensure real-time response under normal usage. The system consistently delivered results within 1–2 seconds per request, meeting performance expectations.

4. PROJECT REQUIREMENTS

4.1 FUNCTIONAL REQUIREMENTS

- The system should allow users to input speech for pronunciation analysis.
- Phoneme analysis and comparison should be conducted at the phoneme level.
- The system should provide feedback highlighting incorrect phonemes in real time.
- It should suggest similar words for practice based on error patterns using LLMs.
- The system must allow users to practice the correct pronunciation after feedback.
- The user should be able to view a visual breakdown of correct and incorrect phonemes.
- Users should have the ability to log in and save their progress for future sessions.
- The system Track and store user progress, including the words practiced, phoneme errors, and improvements over time.

4.2 NON-FUNCTIONAL REQUIREMENTS

- **Performance**: The system should process and return pronunciation feedback within 1-2 seconds.
- **Scalability**: The system must support multiple concurrent users without degrading performance.
- **Reliability**: The system should achieve 99% uptime to ensure availability for users.
- **Security**: User data and progress should be encrypted and stored securely.
- **Usability**: The interface should be intuitive and user-friendly, particularly for non-technical users.
- **Maintainability**: The system should be designed in a modular fashion for easy maintenance and updates.
- **Compatibility**: The system should be accessible via major browsers and mobile devices.

- **Extensibility**: It should allow for the easy addition of new languages or dialects in the future.

4.3 SYSTEM REQUIREMENTS

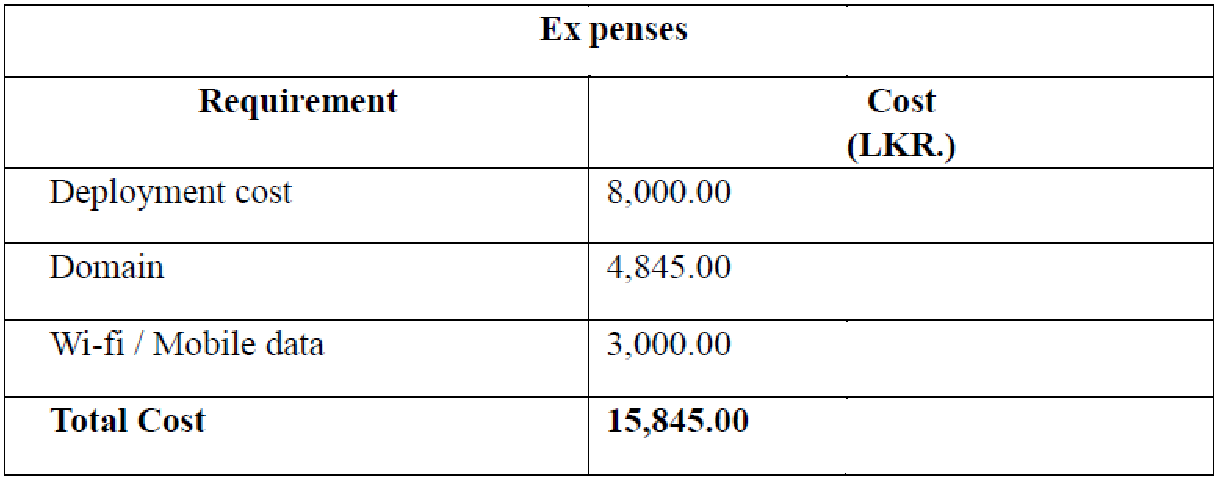
- **Server**: Cloud-based (AWS) for hosting the API.
- **Database**: NoSQL database (firebase) to store user data, phoneme data, and progress.
- **Frontend**: React.js and Tailwind CSS for the user interface.
- **Backend**: python for handling API requests, Python for advanced phoneme analysis and ML components.
- **APIs**: Integration with speech recognition and Gemini AI for pronunciation analysis and feedback generation.

4.4 4. USER REQUIREMENTS

- Users should be able to easily navigate the system interface.
- The system must provide clear, real-time feedback on pronunciation errors.
- Users should be able to view a history of their progress and improvements.
- The system should allow users to replay their pronunciation and compare it with the correct version.
- Users should have the option to practice suggested words for error correction.
- User profiles should allow customization (skill level).

5. BUDGET AND BUDGET JUSTIFICATION

Table 5-1 Expenses for the proposed system



Ex penses	
Requirement	**Cost (LKR.)**
Deployment cost	8,000.00
Domain	4,845.00
Wi-fi / Mobile data	3,000.00
Total Cost	**15,845.00**

Justifications:

1. **Deployment Cost** (8,000 LKR):

   This includes the cost of deploying the system on a cloud platform (such as AWS, Google Cloud, or Firebase). The expense covers server hosting, security, and database storage needed for the proper functioning of the application.

2. **Domain** (4,845 LKR):

   This cost is for purchasing a domain name for the application, which is essential to make the system accessible via the web. A custom domain enhances the professional appearance of the platform.

3. **Wi-Fi / Mobile Data** (3,000 LKR):

   This covers the necessary internet connection required for the development, testing, and maintenance phases of the project. Reliable internet access is crucial for communication with APIs, cloud services, and real-time processing needs.

The total estimated cost for the project amounts to **15,845 LKR**.

6. COMMERCIALIZATION

6.1 TARGET AUDIENCE AND MARKET

1. **Target Audience:**

- **English Language Learners (ELL):** The primary users of the platform are individuals who want to improve their English pronunciation skills, particularly non-native speakers and students.
- **Educational Institutions:** Schools, universities, and language training centers can incorporate the platform into their curriculum to help students improve their pronunciation through guided practice and feedback.
- **Professionals Seeking English Proficiency:** Professionals looking to enhance their communication skills for work, especially those in global industries where English is the primary business language.
- **Students Preparing for Exams:** Learners preparing for exams such as IELTS, TOEFL, and others, where accurate pronunciation is critical.

2. **Market Overview:**

- **Global Reach:** With English being the most widely spoken second language in the world, the potential market is vast. The platform can serve users worldwide, particularly in regions where English is a crucial skill (e.g., South Asia, East Asia, Latin America).
- **Local Focus:** In Sri Lanka and neighboring countries, where English education is prioritized but high-quality pronunciation tools are lacking, the platform can gain significant traction among students and professionals.

Business Strategy

1. **Freemium Model:**

- Offer a free version of the platform that provides basic pronunciation practice, feedback, and a limited set of exercises.

- Introduce premium plans with advanced features such as detailed phoneme-level feedback, personalized training exercises, and a larger library of practice words.
- Premium users could also access features like practice for specific professional contexts (e.g., medical, legal, or business English) and real-time assessments.

2. **Subscription Plans:**

- **Monthly/Yearly Subscriptions:** Provide different tiers of subscription plans (Basic, Professional, Educational Institution) to cater to individual learners, professionals, and institutions.

- Offer discounts for educational institutions or businesses that adopt the system for their training programs.

3. **B2B Partnerships:**

- Collaborate with schools, language institutes, and corporate training providers to integrate the system into their curriculum.
- Offer enterprise solutions with bulk subscriptions, tailored lesson plans, and dedicated support for institutions.

4. **Localization:**

- Localize the system for different markets by supporting multiple languages for instructions and feedback while focusing on English pronunciation improvement.

Marketing Strategies

1. **Digital Marketing:**

- **Social Media Campaigns:** Leverage platforms like Facebook, Instagram, and LinkedIn to reach language learners, professionals, and educational institutions. Tailor ads to highlight key features such as phoneme-level feedback and personalized learning paths.
- **Content Marketing:** Create a blog and video tutorials that provide tips on English pronunciation, the importance of phoneme-level correction, and how the platform helps in mastering it.

- **Influencer Collaborations:** Partner with language learning influencers or educators who can review and promote the platform to their followers.

2. **Educational Partnerships:**

   - Collaborate with universities, language institutes, and schools to offer the platform as part of their language programs.
   - Provide free trials or discounted rates for bulk sign-ups to educational institutions.

3. **SEO Optimization:**

   - Ensure the website is optimized for search engines with keywords like "English pronunciation practice," "phoneme correction tool," and "AI-based pronunciation feedback."
   - Create informative content (blogs, guides) that targets long-tail keywords to increase organic traffic.

4. **Referral Program:**

   - Implement a referral program where users can earn discounts or free premium access by inviting their peers or colleagues to the platform.
   - Target educators and institutions by offering additional perks for referrals made through their network.

5. **Localized Marketing:**

   - Focus on region-specific marketing efforts, especially in countries where English learning is a priority. This could include online ads in local languages, partnerships with regional educational platforms, or even collaborations with local influencers.

7. WORK BREAKDOWN STRUCTURE

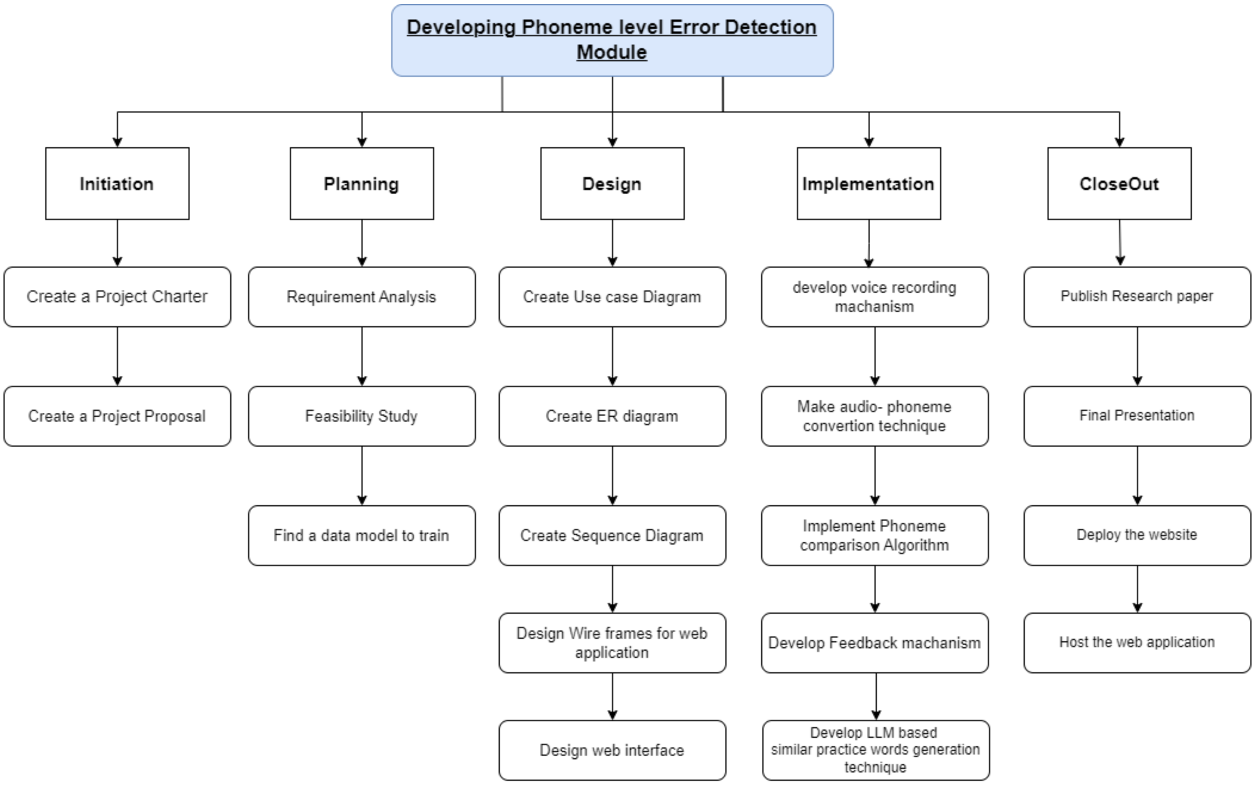
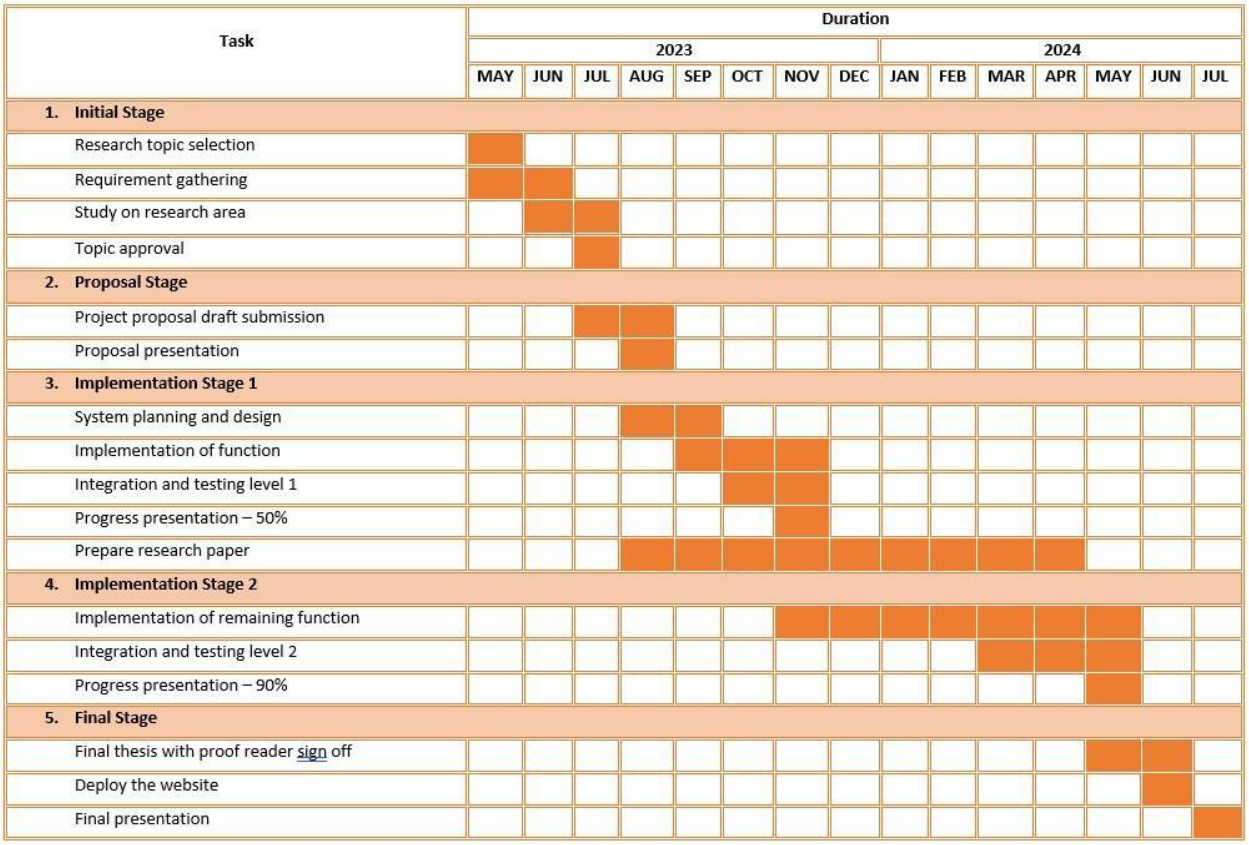

Developing Phoneme level Error Detection Module
Initiation
Planning
Design
Implementation
CloseOut
Create a Project Charter
Create a Project Proposal
Requirement Analysis
Feasibility Study
Find a data model to train
Create Use case Diagram
Create ER diagram
Create Sequence Diagram
Design Wire frames for web application
Design web interface
develop voice recording machanism
Make audio- phoneme convertion technique
Implement Phoneme comparison Algorithm
Develop Feedback machanism
Develop LLM based similar practice words generation technique
Publish Research paper
Final Presentation
Deploy the website
Host the web application

Figure 7-1 Work Breakdown Structure

8. GNATT CHART



Task	Duration														
	2023								2024						
	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL
1. Initial Stage															
Research topic selection	🟧														
Requirement gathering	🟧	🟧													
Study on research area		🟧	🟧												
Topic approval			🟧												
2. Proposal Stage															
Project proposal draft submission			🟧	🟧											
Proposal presentation				🟧											
3. Implementation Stage 1															
System planning and design				🟧	🟧										
Implementation of function					🟧	🟧	🟧								
Integration and testing level 1						🟧	🟧								
Progress presentation – 50%							🟧								
Prepare research paper				🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧			
4. Implementation Stage 2															
Implementation of remaining function							🟧	🟧	🟧	🟧	🟧	🟧	🟧		
Integration and testing level 2											🟧	🟧	🟧		
Progress presentation – 90%													🟧		
5. Final Stage															
Final thesis with proof reader sign off													🟧	🟧	
Deploy the website														🟧	
Final presentation															🟧

Figure 8-1 Gnatt Chart

9. REFERENCES

[1] A. Noiray, K. Iskarous, and D. H. Whalen, "Variability in English vowels is comparable in articulation and acoustics," *Lab. Phonol. J. Assoc. Lab. Phonol.*, vol. 5, no. 2, Jan. 2014, doi: 10.1515/lp-2014-0010.

[2] Peter, "AE 466 – Ship or Sheep? | English Pronunciation of /iː/ vs /ɪ/," *Aussie English*, Aug. 25, 2022. [Online]. Available: https://aussieenglish.com.au/ae-466-ship-or-sheep-english-pronunciation-of-i-vs-%C9%AA/

[3] M. Hismanoglu and S. Hismanoglu, "Language teachers' preferences of pronunciation teaching techniques: traditional or modern?," *Procedia - Soc. Behav. Sci.*, vol. 2, no. 2, pp. 983–989, Jan. 2010, doi: 10.1016/j.sbspro.2010.03.138.

[4] P. M. Rogerson-Revell, "Computer-Assisted Pronunciation Training (CAPT): Current Issues and Future Directions," *RELC J.*, vol. 52, no. 1, pp. 189–205, Jan. 2021, doi: 10.1177/0033688220977406.

[5] D. Korzekwa, J. Lorenzo-Trueba, T. Drugman, and B. Kostek, "Computer-assisted pronunciation training—Speech synthesis is almost all you need," *Speech Commun.*, vol. 142, pp. 22–33, Jun. 2022, doi: 10.1016/j.specom.2022.06.003.

[6] S. Coulange, "Computer-aided pronunciation training in 2022: When pedagogy struggles to catch up," *HAL (Le Centre Pour La Communication Scientifique Directe)*, Jan. 2023, doi: 10.5281/zenodo.8137754.

[7] S. Nita, E. R. N. Sari, K. Sussolaikah, and S. M. F. Risky, "The Implementation of Duolingo Application to Enhance English Learning for Millennials," *J. Int. Lingua Technol.*, vol. 2, no. 1, pp. 1–9, Jun. 2023, doi: 10.55849/jiltech.v2i1.215.

[8] R. N. Hermana, "Rosetta Stone Application on Students' Pronunciation," *J. English Teach. Linguist. Stud. (JET Li)*, vol. 5, no. 2, pp. 92–102, Oct. 2023, doi: 10.55215/jetli.v5i2.8779.

[9] E. Miller, "BoldVoice Review: Despite its usefulness, it has a BIG flaw," *Medium*, Mar. 5, 2024. [Online]. Available: https://medium.com/@emmamillerw1990/boildvoice-review-despite-its-usefulness-it-has-a-big-flaw-33c944b1150c

[10] "LinguaCoach an English Phonetics Approach The best place to master english pronunciation." [Online]. Available: https://linguacoach.appspot.com

[11] R. Ai, "Automatic Pronunciation Error Detection and Feedback Generation for CALL Applications," in *Lecture Notes in Computer Science*, 2015, pp. 175–186, doi: 10.1007/978-3-319-20609-7_17.

[12] A. Neri, C. Cucchiarini, and H. Strik, "Selecting segmental errors in non-native Dutch for optimal pronunciation training," *Int. Rev. Appl. Linguist. Lang. Teach.*, vol. 44, no. 4, pp. 357–404, 2006.

[13] K. Kyriakopoulos, K. M. Knill, and M. J. F. Gales, "Automatic Detection of Accent and Lexical Pronunciation Errors in Spontaneous Non-Native English Speech," in *Proc. INTERSPEECH*, Oct. 2020, doi: 10.21437/interspeech.2020-2881.

[14] J. Levis, "L2 pronunciation research and teaching," *J. Second Lang. Pronunc.*, vol. 7, no. 2, pp. 141–153, 2021, doi: 10.1075/jslp.21037.lev.

[15] Q. Ai *et al.*, “Information Retrieval meets Large Language Models: A strategic report from Chinese IR community,” *AI Open*, vol. 4, pp. 80–90, Jan. 2023, doi: 10.1016/j.aiopen.2023.08.001.