WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Automatic Pronounciation Mistake Detector

K. Sanath A[1], Vignesh Chandra[2], Mrs. V. Veena[3]

[1,2]Student, [3]Guide, **Asst. Professor, IT**
Department of Information Technology, Mahatma Gandhi Institute of Technology (A)
DOI : https://doi.org/10.55248/gengpi.6.0425.14117

ABSTRACT

The Automatic Pronunciation Mistake Detection project is an innovative and efficient system designed to help students improve their English pronunciation skills. By utilizing speech recognition technologies, such as PyAudio and pyttsx3, the project focuses on reducing error rates and enhancing the accuracy of error detection. The system provides real-time feedback to users, enabling them to identify and correct pronunciation errors effectively.

The system works by allowing users to select a word they want to practice, record their pronunciation, and compare the recorded audio against the correct pronunciation. If the recorded pronunciation does not match, the system notifies the user with a pop-up indicating a mistake. Users can view their mispronounced words and listen to the correct pronunciation by clicking on an audio file. This interactive approach makes it easier for users to learn and refine their pronunciation skills.

The project is built using HTML, CSS, and JavaScript for the front-end interface, while Python powers the backend. Django is used as the framework, with MySQL serving as the database for storing user information and word details. Admin users have privileges to manage the system by adding, updating, deleting, and viewing words, as well as monitoring user details and their pronunciation mistakes. This robust and user-friendly design ensures an effective learning experience for students.

1. INTRODUCTION

The Automatic Pronunciation Mistake Detection System is designed to address the challenges non-native speakers face in mastering pronunciation. By leveraging advanced speech recognition and machine learning, the system provides real-time feedback to help learners identify and correct errors. It caters to users of all proficiency levels, from beginners to advanced learners, and supports multiple languages and regional accents. This innovative system bridges the gap left by traditional classroom-based learning by offering personalized, accessible, and scalable solutions. With features like progress tracking and adaptive learning levels, it ensures an engaging learning experience. Its flexibility makes it suitable for use in educational institutions, corporate training, and mobile apps. Ultimately, the system empowers learners to enhance their communication skills and build confidence in their spoken language abilities.

1.1 Background

Pronunciation is a critical aspect of language learning, as it directly impacts effective communication and comprehension. For non-native speakers, mastering accurate pronunciation is a significant challenge due to variations in phonetics, accents, and linguistic structures. The advent of technology, particularly in speech recognition and artificial intelligence (AI), has opened up new possibilities for addressing this challenge. Research in this domain has laid the groundwork for developing automated systems capable of detecting and correcting pronunciation errors in real time, thereby improving learners' language skills and confidence. Several studies have significantly influenced the conceptualization and design of this project, focusing on frameworks, machine learning algorithms, and advanced feedback mechanisms.

The work of Bandar Ali Al-Rami and Yousef Houssni Zrekat (2023) presents a comprehensive framework for detecting and correcting pronunciation errors specifically tailored for non-native Arabic speakers learning English. Their study emphasizes the importance of developing systems that account for the linguistic and cultural nuances of target users. The proposed framework integrates error detection with real-time corrective feedback, allowing learners to understand and improve their pronunciation mistakes effectively. This approach underscores the value of personalization in language learning tools, ensuring that the system addresses the unique challenges faced by specific learner groups. The insights from this study serve as a cornerstone for this project, particularly in designing user-centric features and adaptive learning pathways.

Yuhua Dai (2022) conducted a study leveraging an improved random forest model to enhance the accuracy and efficiency of pronunciation error detection. The research demonstrated the potential of machine learning algorithms in processing speech data, identifying errors, and offering feedback

at a granular level. One of the key contributions of this study is its emphasis on the scalability and adaptability of machine learning-based systems, enabling them to cater to learners with varying proficiency levels. The findings have directly influenced this project's technical foundation, particularly in selecting algorithms that balance computational efficiency with detection precision. Moreover, the random forest model's success highlights the importance of combining robust algorithms with real-world educational applications.

Lei Chen et al. (2022) explored the use of deep learning models to detect reduced-form pronunciations in English, focusing on subtle pronunciation errors often overlooked by traditional systems. Their study introduced innovative methodologies for analyzing speech at the phoneme level, enabling precise identification of errors that impact fluency and intelligibility. The integration of deep learning models demonstrated improved accuracy in detecting these nuances, paving the way for real-time corrective feedback. This research inspires this project's adoption of deep learning techniques to enhance the system's ability to provide targeted and actionable feedback. By enabling learners to address specific pronunciation challenges, the system aligns with modern pedagogical goals of fostering autonomy and engagement in language learning.

Drawing from these foundational studies, the Automatic Pronunciation Mistake Detection System aims to combine advanced frameworks, machine learning algorithms, and deep learning models to create a comprehensive solution. The system incorporates real-time feedback mechanisms inspired by Al-Rami and Zrekat's framework, ensuring that learners receive immediate guidance tailored to their unique needs. Building on Dai's research, the project adopts scalable machine learning models, ensuring adaptability across diverse user groups and proficiency levels. Finally, Chen et al.'s insights into deep learning drive the project's focus on phoneme-level analysis, allowing for precise detection and correction of subtle pronunciation errors.

These foundational studies collectively highlight the growing demand for automated language learning tools that are accurate, efficient, and learner-focused. By integrating the key findings and methodologies from these studies, this project seeks to bridge existing gaps in pronunciation training tools. It aims to address challenges such as real-time feedback, support for diverse accents, and user-friendly interfaces, providing a robust solution for learners worldwide. Furthermore, the project's design aligns with educational goals, enhancing accessibility, and empowering learners to develop their pronunciation skills independently and effectively.

1.2 Objectives

The primary objective of this project is to enhance pronunciation accuracy among language learners by detecting pronunciation errors at the phoneme level and providing precise, actionable feedback. By focusing on the nuances of speech, the system ensures that users can make targeted improvements to their spoken language skills, ultimately aiding in better communication.

Another significant goal is to offer real-time feedback, enabling users to correct mistakes as they speak. This immediate guidance creates an interactive and engaging learning process, which is essential for effective skill development. The system aims to empower learners by providing a dynamic and supportive environment for practicing pronunciation.

Supporting multi-language learning is a key objective, with the system initially focusing on English. Over time, the project aspires to expand its scope to include multiple languages, catering to diverse linguistic backgrounds and ensuring inclusivity. This multilingual capability makes the system versatile and globally relevant.

A critical consideration is the design of a user-friendly interface that ensures accessibility for individuals of all ages and technical abilities. The system will feature an intuitive and interactive design, enabling seamless navigation and encouraging users to engage actively in their learning journey. Furthermore, the project aims to accommodate learners at all proficiency levels, from beginners to advanced, through adaptive learning features that provide personalized and progressive experiences.

To promote flexibility, the system will be accessible anytime and anywhere, allowing users to practice pronunciation at their convenience. This 24/7 availability ensures that learners can incorporate language improvement into their daily routines without time constraints, leveraging the power of technology to support continuous learning.

Incorporating progress-tracking tools is another important objective. By monitoring user performance and providing data-driven insights, the system helps learners identify areas for improvement and measure their progress over time. This feature motivates users to set achievable goals and maintain a steady learning curve.

Additionally, the system seeks to address the challenges posed by regional accents and dialects. By incorporating machine learning models that can recognize and adapt to such variations, the system ensures accurate detection and feedback for users from different linguistic regions. This inclusivity enhances the system's effectiveness across diverse user bases.

Scalability and flexibility are also prioritized in the project design, ensuring that the system can be integrated into educational platforms, corporate training programs, and individual use cases. This adaptability makes the system suitable for a wide range of applications and user needs.

The system is also designed to foster independent learning by providing users with the tools to practice and improve their pronunciation skills without relying heavily on human instructors. This autonomy empowers learners to take charge of their language development and build confidence in their spoken communication.

Finally, this project aims to contribute to the broader advancement of speech recognition and machine learning technologies in the educational domain. By setting a benchmark for pronunciation learning systems, the project aspires to influence future innovations and redefine the standards of language education technology.

1.3 Scope

The scope of this technology spans multiple domains, focusing on enhancing real-time communication and accessibility. Key areas include lip-reading and speech recognition for accurate transcription, natural language processing (NLP) for contextual understanding and machine translation, and video processing for synchronizing audio-visual content. Real-time captioning and translation further enable seamless interaction across languages, supported by advanced models and efficient pipelines. These innovations have applications in diverse fields, such as assistive technologies, multilingual communication, media production, and education, ensuring accessibility and inclusivity in various scenarios.

Educational Use

The Automatic Pronunciation Mistake Detection System holds immense potential for revolutionizing language education. It can be seamlessly integrated into the curriculum of schools, colleges, and universities to assist students in mastering pronunciation. By providing real-time feedback and tailored suggestions, the system enables learners to enhance their oral language skills effectively. Language learning platforms and specialized institutes can also leverage this technology to supplement their teaching methods. With its ability to adapt to individual learning speeds and cater to various proficiency levels, the system promises to bridge the gap between traditional language instruction and modern technological solutions.

Professional Development

In the professional realm, effective communication is a vital skill. The system can play a crucial role in corporate training programs, helping employees refine their pronunciation and communication abilities. This is particularly beneficial in multinational companies where clear articulation is essential for collaboration and client interactions. Additionally, professionals such as public speakers, voice artists, and broadcasters can use the system to perfect their pronunciation and deliver impactful speeches or performances. By fostering clear and confident communication, the tool contributes to personal growth and career advancement.

Personal Language Learning

For individual learners, the system offers a highly personalized and flexible learning experience. It caters to a diverse audience, including children, adults, and senior learners, who wish to improve their pronunciation at their own pace. The tool is designed to address varying proficiency levels, from beginners struggling with basic sounds to advanced users aiming for accent refinement. With features like progress tracking and adaptive learning, it empowers users to achieve their language goals independently, making it a valuable resource for personal development.

Multilingual Applications

One of the standout features of the system is its potential for multilingual support. While the initial focus is on English, the design allows for expansion into other languages, catering to the needs of a global audience. Furthermore, the system can adapt to regional accents and dialects, ensuring accurate feedback for users from different linguistic backgrounds. This inclusivity makes it a versatile tool for learners worldwide, breaking barriers in language education and making it accessible to a diverse user base.

Accessibility and Convenience

The system is designed to be accessible 24/7, enabling users to practice their pronunciation anytime and anywhere. Whether at home, during a commute, or in a classroom setting, learners can use their preferred devices, such as smartphones, tablets, or computers, to engage with the tool. Its user-friendly interface ensures that even those with limited technical skills can navigate the platform effortlessly. By offering convenience and ease of use, the system encourages consistent practice, which is key to mastering pronunciation.

Integration with Technology

Technological integration is another significant scope of this system. It can be incorporated into existing educational platforms, enhancing their offerings and enabling seamless learning experiences. The system's compatibility with mobile applications, smart assistants, and IoT devices opens up new avenues for interactive and immersive learning. Such integrations make the tool more accessible and engaging, appealing to tech-savvy users and broadening its adoption.

Speech Therapy and Rehabilitation

Beyond education and personal learning, the system can play a transformative role in speech therapy and rehabilitation. Speech therapists can use the tool to help patients with speech disorders practice and improve their pronunciation. The real-time feedback and adaptive features make it an ideal aid for individuals recovering from speech impairments caused by medical conditions such as strokes or surgeries.

Research and Development

The system also holds potential for advancing research in speech recognition and language learning. Researchers can use it to study speech patterns, identify common pronunciation challenges, and develop more effective language learning strategies. The integration of cutting-edge technologies such as machine learning and natural language processing ensures that the system remains at the forefront of innovation. This makes it a valuable asset for academic and technological research communities.

Global Outreach

With its scalability and multilingual capabilities, the system can cater to users across the globe. It bridges geographical barriers in language education, making pronunciation improvement tools accessible to learners in remote or underserved areas. By facilitating better communication skills, the system equips non-native speakers to participate more confidently in a globalized world, fostering cultural exchange and mutual understanding.

Scalability and Versatility

The scalability and versatility of the system are key to its wide-ranging applications. It can be tailored to meet the needs of various user groups, from individual learners to educational institutions and corporate organizations. The system's ability to handle large user bases and adapt to specific requirements makes it a flexible and robust solution. Whether used for personal development, classroom learning, or professional training, the tool ensures a consistent and impactful learning experience for all users.

LITERATURE SURVEY

[1] Introduction to Automatic Pronunciation Mistake Detector

The Automatic Pronunciation Mistake Detector is an advanced system designed to help language learners, especially non-native English speakers, improve their pronunciation skills through real-time feedback and actionable guidance. By leveraging speech recognition technologies like PyAudio and pyttsx3, the system analyzes user recordings, detects pronunciation errors, and provides corrections by comparing the user's pronunciation with the correct one. Users can select words, record their speech, and listen to the accurate pronunciation, fostering an interactive and effective learning process. Built using technologies such as HTML, CSS, JavaScript, Python, Django, and MySQL, the system is designed to be accessible and user-friendly. It also includes administrative features for managing user progress and word databases. By bridging the gap between traditional language learning methods and modern technology, this tool empowers learners to enhance their communication skills confidently and efficiently.

[2] Key Technologies in Automatic Pronunciation Mistake Detector

The key technologies used in the Automatic Pronunciation Mistake Detector project include a combination of speech recognition tools, web development frameworks, and database management systems. For speech recognition, PyAudio is utilized to capture user audio input, while pyttsx3 provides text-to-speech functionality, allowing users to hear the correct pronunciation of words. The frontend of the application is built using HTML, CSS, and JavaScript, ensuring a visually appealing and interactive user interface. On the backend, Python serves as the primary programming language, with Django as the web framework to manage application logic and workflows. MySQL is employed for secure storage and management of user data, word lists, and pronunciation records. Additionally, APIs and libraries are integrated as needed for advanced speech-to-text conversion and error detection. Together, these technologies create a robust, secure, and user-friendly platform for improving pronunciation skills effectively.

[3] Applications and Future Prosects

The Automatic Pronunciation Mistake Detector has a wide range of applications across various domains. It is particularly beneficial in language learning, helping students and non-native speakers improve their English pronunciation. Educational institutions can integrate this tool into their language courses to provide personalized learning experiences. Language training centers and online learning platforms can use it as a supplemental resource for accent training and communication skill enhancement. Additionally, it can be employed in corporate training programs to help employees working in international environments improve their spoken English.

In terms of future prospects, the system has immense potential for expansion. It can be enhanced to support multiple languages, catering to diverse learners worldwide. Integration of advanced AI algorithms can enable more granular feedback on phoneme-level pronunciation errors, making it even more effective. The tool could also evolve into a mobile application, allowing users to practice on the go. With advancements in machine learning and natural language processing, the system could incorporate real-time conversational practice, adaptive learning techniques, and gamified elements to increase user engagement. These developments can position the system as a versatile and indispensable tool for language learning in the digital age..

Table 2.1 Literature Table

Sl. No	Title	Author	Journal & Name year	Methodology Adapted	Key Findings	Gaps

1	End-to-End Automatic Pronunciation Error Detection Based on Hybrid CTC/Attention Architecture.	X. Zhang, L. Li.	2023/Research gates.	CTC/Attention, deep learning.	Integration of attention mechanisms improves phoneme detection accuracy.	Performance dependent on ASR accuracy, struggles with noisy data. Sensitive to the accuracy of the.
2	Attention-Based Multi-Encoder Automatic Pronunciation Assessment	J. Kim et al.	2023/Research Gates	Attention mechanism, multi-encoder models	Multi-encoder models help focus on specific speech features. Attention mechanisms allow for better modeling of pronunciation variations.	Limited multilingual support, generalization issues Generalization issues in low-resource environments or uncommon accents.
3	Transfer Learning for Multilingual Pronunciation Assessment	A. Gupta, S. Roy	2022/IEEE	Transfer learning, CNN, RNN	Transfer learning improves multilingual assessment. Improves generalization across multiple languages, especially for high-resource languages.	Requires large training datasets, low-resource languages underrepresented. Relies on large datasets, making it challenging to apply in low-resource languages.
4	Mispronunciation Detection with End-to-End ASR	N. Singh, D. Kuma	2022/IEEE	End-to-end ASR, neural networks	End-to-end ASR streamlines error detection. They support multilingual capabilities in error detection.	Accuracy heavily depends on ASR quality. Performance is heavily dependent on ASR system quality.
5	Pronunciation Error Detection Using Segmental Deep Neural Networks	C. Li, Y. Wang	2021/IEEE	Segmental DNNs, phoneme extraction	Segmental DNNs outperform traditional methods in precision.	Language dependency limits generalization. The approach shows limited generalization to non-native speakers from diverse
6	Improved Articulatory Models for Pronunciation Error Detection	H. Chen, S. Zhao	2022/IEEE	Articulatory DNN models	Modeling speech articulation enhances feedback accuracy. Articulatory models offer superior feedback on articulation accuracy.	High computational cost, real-time application not tested. Computationally intensive, leading to challenges in real-time applications.

7	Speech-Based Pronunciation Evaluation Using Deep Learning	M. Lee, J. Cho	2021/IEEE	DNN, MFCC, feature extraction	Deep learning models like DNNs improve the classification of pronunciation mistakes.	Struggles with detecting errors in non-standard accents or dialects. Low accuracy for context-sensitive pronunciation mistakes.
8	Confidence Measures in Automatic Pronunciation Detection	L. Wei, T. Sun	2021/IEEE	Log-likelihood, posterior probability	Refined measures improve feedback on non-native pronunciation.	Sensitivity to noisy data impacts robustness
9	Segmental Pronunciation Error Detection Using Acoustic-Phonetic Models	R. Patel, S. Bose	2020/IEEE	Acoustic-phonetic models, phoneme analysis	Phoneme segmentation improves error localization.	Lack of real-time processing capability.
10	Neural Network-Based Mispronunciation Detection for Non-Native Learners	K. Zhang, Y. Luo	2020/IEEE	CNN, LSTM, neural networks	Combining CNN and LSTM provides robust phoneme sequence modeling.	Pronunciation mistakes in low-resource languages remain unaddressed

3. DESIGN

3.1 Uml Diagrams

Use case Diagram:

The diagram illustrates the use case model of an **Automatic Pronunciation Mistake Detection System**, detailing the interactions between two primary actors: the student and the administrator. Students can use the system to enhance their pronunciation skills. They start by selecting a language and recording their speech. The system then analyses the pronunciation and generates feedback based on detected mistakes. Students can view the feedback to improve their speaking skills. Additionally, they can view a progress report to track their improvement over time. Administrators manage the system by maintaining the available languages and tracking usage data to ensure effective operation and system optimization. This system efficiently combines speech analysis and user interaction to assist students in improving their pronunciation skills.

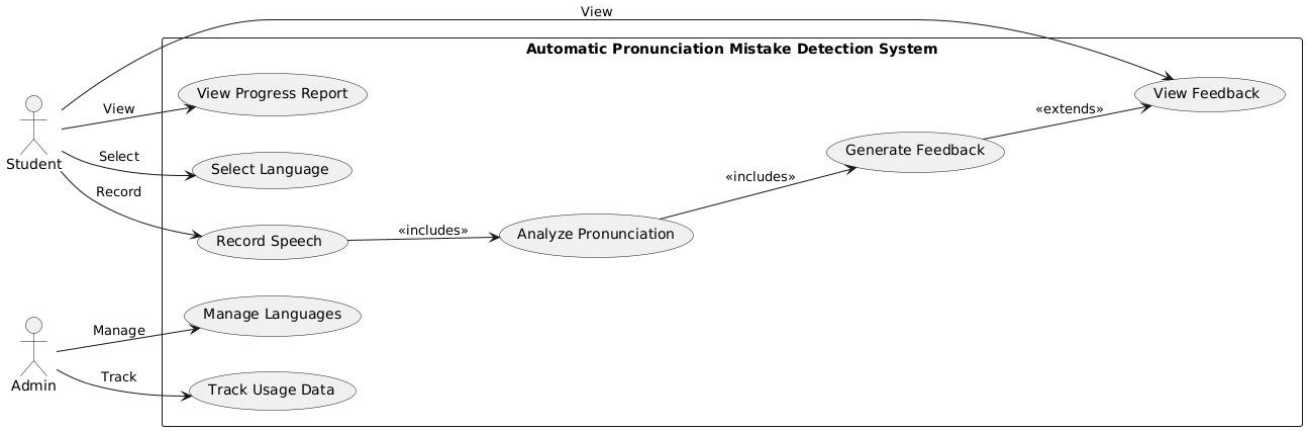


Figure 5.1: Use Case Diagram

3.2Sequence diagram:

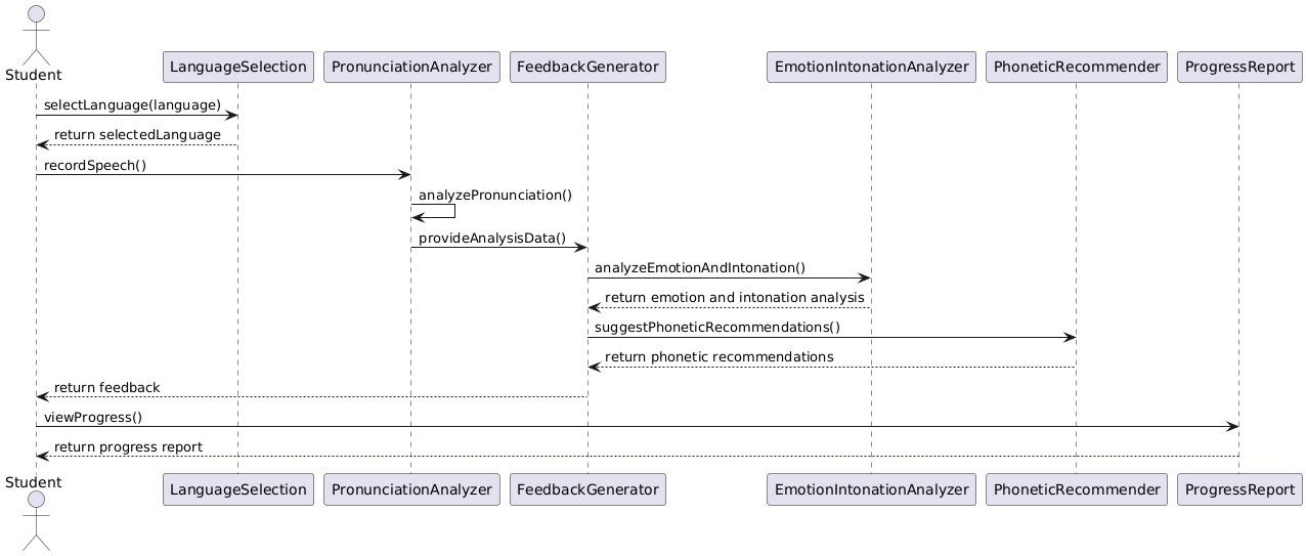


Figure 5.2: Sequence Diagram

The sequence diagram represents the workflow of an Automatic Pronunciation Mistake Detection System, showcasing the interactions between a student and the system's components. The process begins with the student selecting a language through the Language Selection module, which returns the selected language for further operations. The student then records their speech, which is sent to the Pronunciation Analyzer for analysis. This module evaluates the pronunciation and provides detailed analysis data. The data is further processed by the Emotion Intonation Analyzer, which assesses the emotional tone and intonation in the speech and returns the analysis results. Subsequently, the data is passed to the Phonetic Recommender, which suggests specific improvements to enhance the phonetic aspects of the speech. The feedback from these analyses is consolidated by the Feedback Generator, which provides actionable feedback to the student. Finally, the student can track their progress over time by viewing their progress report through the Progress Report module. This system offers a comprehensive and interactive approach to improving pronunciation by integrating linguistic, emotional, and phonetic analysis into a cohesive feedback mechanism.

3.3 State diagram:

The diagram depicts the states and transitions of a pronunciation learning application. It begins in an "Idle" state. The user initiates the process by starting "Language Selection," which transitions the application to the "Language Selection" state. Once a language is selected, the application moves to the "Speech Recording" state, where the user records their speech. Upon successful recording, the application proceeds to "pronunciation Analysis" to analyse the recorded speech. The analysis triggers the "Feedback Generation" state, where feedback based on the analysis is generated. This feedback is then displayed in the "Display Feedback" state. Finally, the user can view their progress in the "Progress Viewing" state, which then loops back to the initial "Idle" state, ready for another session. Each transition is triggered by a specific event, such as "start Language Selection," "Language Selected," "Speech recorded," "Analyse pronunciation," "Generate Feedback," and "View Progress."

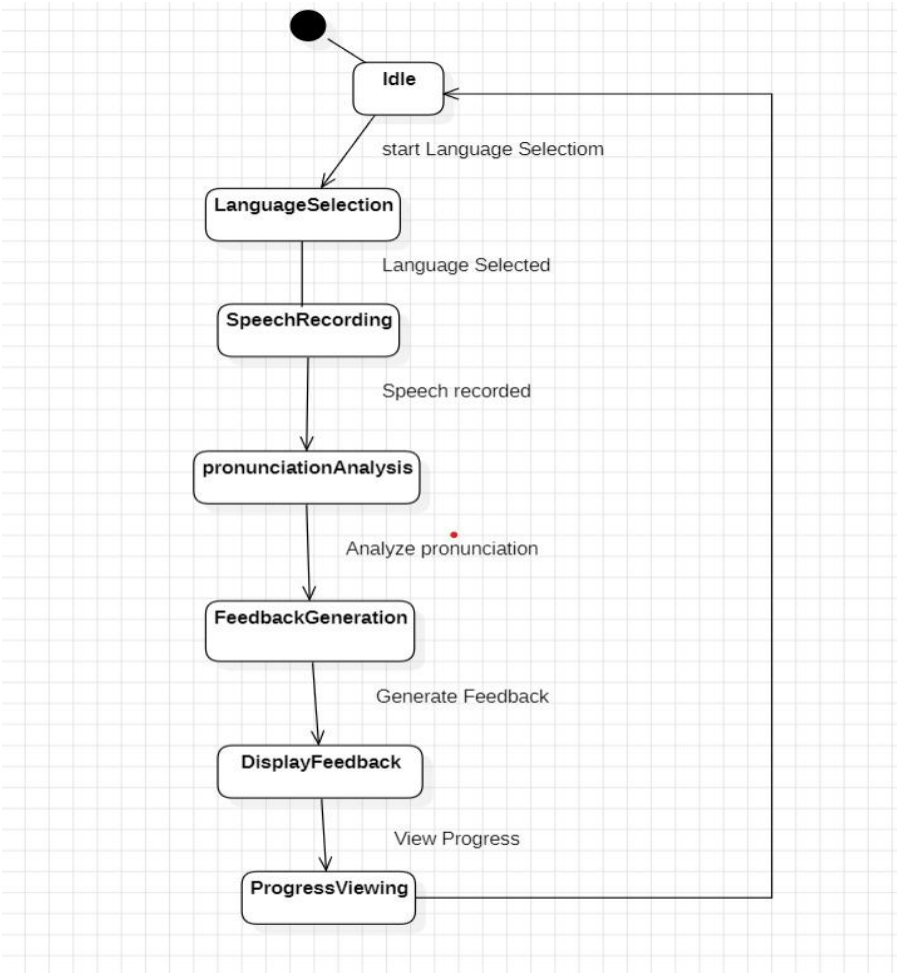
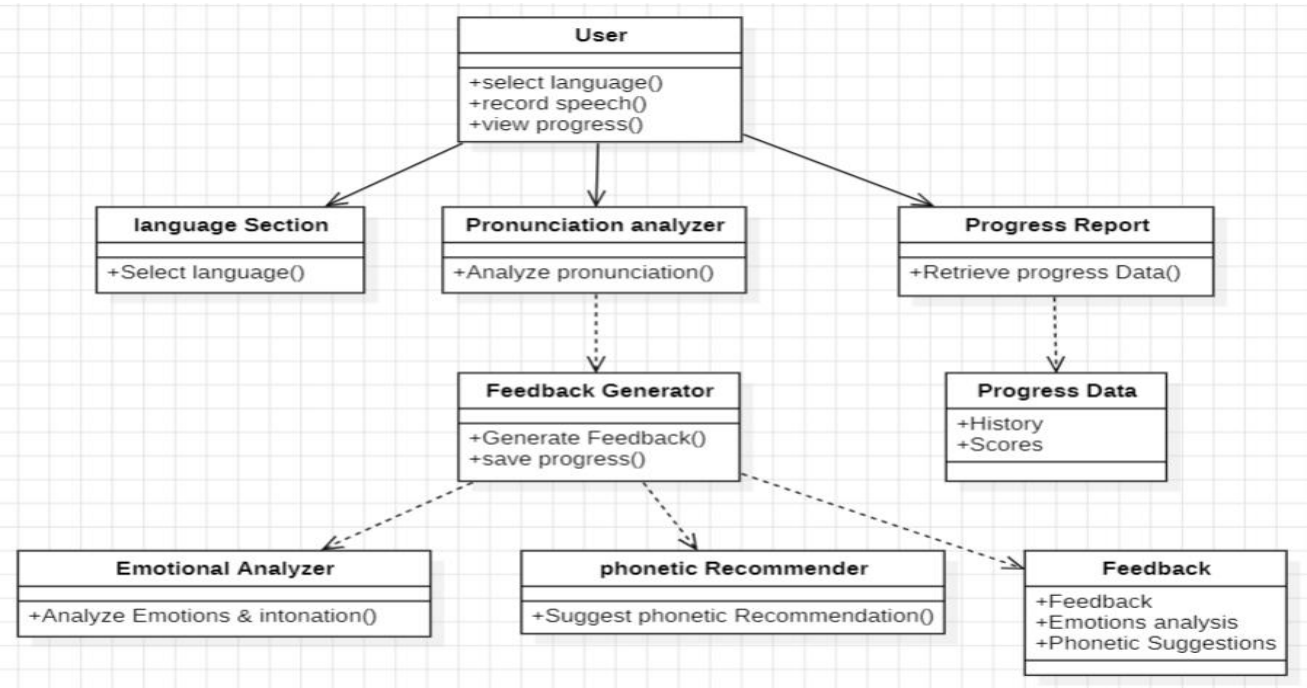


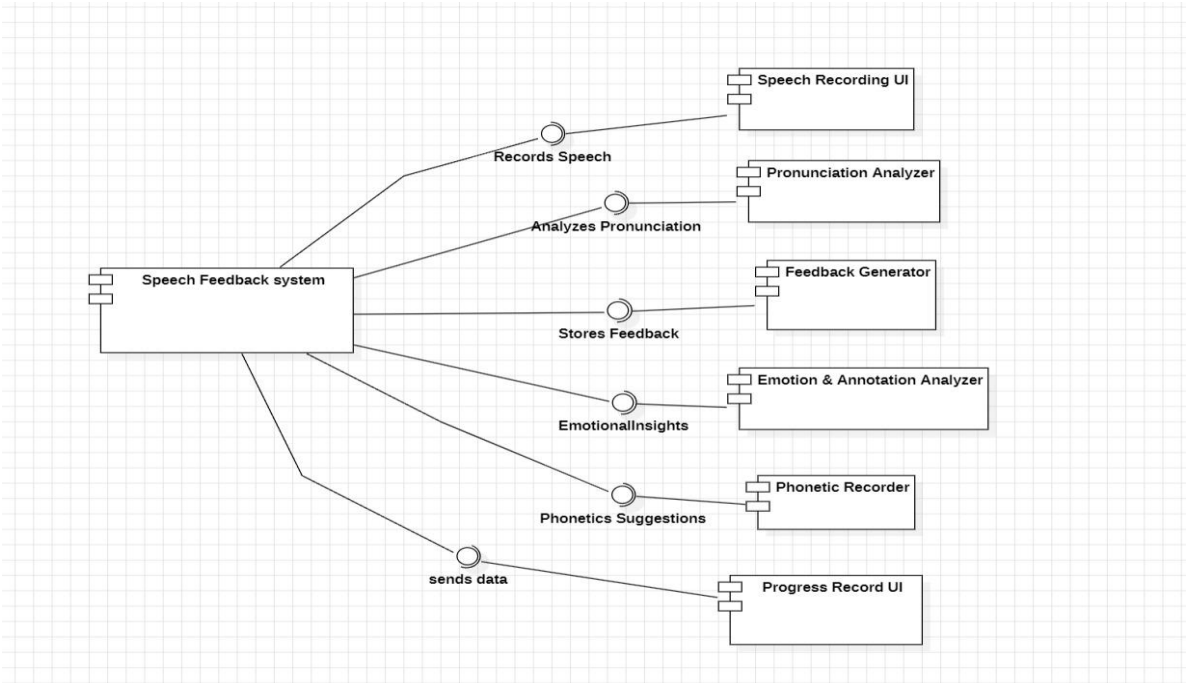
Figure 5.3: State diagram

3.4 Class Diagram



The diagram represents a system designed to assist users in language learning and pronunciation improvement. At the centre is the User, who interacts with various components of the system. The user can select a language through the Language Section, record speech, and view progress. Speech recordings are analysed by the Pronunciation Analyzer, which evaluates the user's pronunciation. Based on this analysis, the Feedback Generator creates detailed feedback, which may include suggestions for improvement, emotional tone analysis, and phonetic recommendations. The Emotional Analyzer assesses the user's emotions and intonation in their speech, while the Phonetic Recommender provides specific tips for refining pronunciation. The system tracks the user's progress through the Progress Report, which retrieves data stored in the Progress Data module. This module keeps a record of the user's learning history and scores. Feedback generated by the system is consolidated into three categories: general feedback, emotional analysis, and phonetic suggestions. By integrating these components, the system provides a comprehensive and user-friendly tool for language learning and pronunciation enhancement.

3.5 Component Diagram



The diagram represents a "Speech Feedback System" and its interconnected components designed to assist with speech analysis and improvement. The core system records speech through a "Speech Recording UI," which acts as the user interface for capturing audio. This recorded speech is analysed by the "Pronunciation Analyzer" to evaluate pronunciation accuracy. The system further processes and stores feedback using a "Feedback Generator," which provides constructive insights to improve speech. Emotional insights are derived using an "Emotion & Annotation Analyzer," capturing emotional tone and annotating key aspects of the speech. Additionally, phonetics are managed with a "Phonetic Recorder," which records and suggests phonetic improvements. The system tracks progress and sends data to a "Progress Record UI," enabling users to monitor their development over time. This interconnected structure ensures comprehensive speech evaluation and personalized feedback, helping users enhance their pronunciation, emotional tone, and overall speaking skills.

3.6 Deployment Diagram

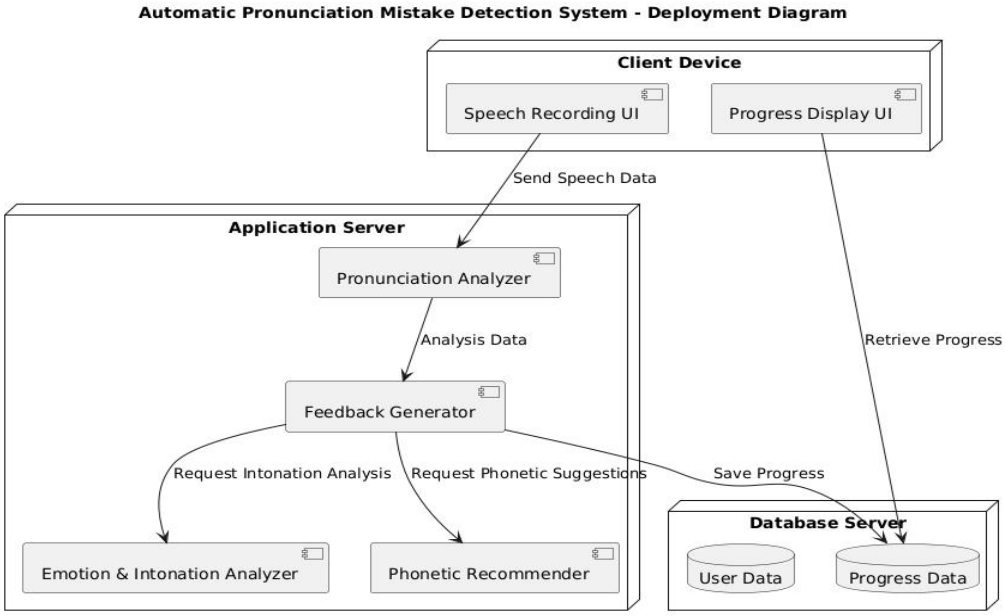


Fig:5.6 Deployment diagram

The deployment diagram illustrates the architecture of the Automatic Pronunciation Mistake Detection System, comprising the Client Device, Application Server, and Database Server. The Client Device includes the Speech Recording UI for capturing speech and the Progress Display UI for displaying feedback and progress. The Application Server, with its Pronunciation Analyzer and Feedback Generator, detects pronunciation errors and provides actionable feedback. The Feedback Generator collaborates with the Emotion & Intonation Analyzer for expressive speech insights and the Phonetic Recommender for corrective suggestions. The Database Server stores User Data and Progress Data, enabling performance tracking. Users record speech, which is analyzed and processed on the server, and the results are stored and displayed. This system provides real-time analysis, phonetic recommendations, intonation evaluation, and progress tracking for efficient pronunciation improvement.

SYSTEM REQUIREMENTS

The system requirements for the software encompass both software and hardware specifications to ensure optimal performance. On the software side, the operating system must be Windows 10 or later, with development utilizing the full MERN stack, including MongoDB, Express.js, React.js, and Node.js. Machine learning functionalities rely on libraries such as TensorFlow or PyTorch, along with NLP libraries for advanced processing. API integrations, including speech recognition and translation APIs, are essential for seamless functionality, while web technologies like HTML5, CSS3, JavaScript, and WebSockets support efficient and interactive web-based operations. On the hardware side, a server with a multi-core processor (Intel i7 or equivalent) and a client with a modern multi-core CPU (such as Intel i5 or AMD Ryzen 5 and above) are required. Additionally, a stable, high-speed internet connection, along with a webcam and high-quality microphone, is necessary for real-time processing and interaction.

Category	Requirement
Operating System	Windows 10 or later, macOS 10.15 or later, Ubuntu 18.04 or later
Programming Languages	Python 3.8 or later, JavaScript (for frontend)
Frameworks/Libraries	TensorFlow, PyTorch, Kaldi, Librosa, Flask/Django (backend), React.js/Flutter
Databases	PostgreSQL, MySQL, or MongoDB
Audio Processing Tools	Praat, SoX (Sound eXchange), Audacity
IDE/Editors	Visual Studio Code, PyCharm, Jupyter Notebook
Visualization Tools	Matplotlib, Seaborn, Plotly
Speech APIs	Google Speech-to-Text, Azure Speech Services, CMU Sphinx

Table 4.1: Software Requirements

4.2 Hardware Requirements

Category	Requirement
Processor	Intel Core i5 or AMD Ryzen 5 (minimum), Intel Core i7 or better (recommended)
RAM	Minimum 8 GB, Recommended 16 GB
Storage	Minimum 256 GB SSD, Recommended 512 GB SSD or more
GPU (Optional)	NVIDIA GPU with CUDA support for training models (e.g., NVIDIA RTX 3060 or better)
Microphone	High-quality external microphone or headset
Audio Interface	Optional external sound card for precise audio recording
Network Connectivity	Stable internet connection for accessing APIs and cloud services

Table 4.2: Hardware Requirements

5. CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

The Automatic Pronunciation Mistake Detection System provides an innovative and efficient solution to improve spoken English proficiency, particularly for non-native speakers. By utilizing advanced speech processing techniques, machine learning models, and real-time feedback mechanisms, the system enables users to identify and correct pronunciation errors effectively. It promotes a self-guided and scalable learning environment, eliminating dependency on traditional methods such as human tutors, and making language learning more accessible and convenient. The system serves as a vital tool in bridging communication gaps, fostering better comprehension, and enhancing confidence in spoken communication

5.2 Future Scope

Looking ahead, there are several avenues for further development and enhancement of the platform:

Real-time Conversational Feedback

- Extend functionality to provide pronunciation feedback during live conversations, such as video calls or group discussions.
- Integrate with platforms like Zoom, Google Meet, and Microsoft Teams to assist users in professional and social settings.

Cross-domain Integration

- Use in specialized industries such as healthcare (medical terminology pronunciation), aviation (technical jargon), or customer service (script adherence).

Phoneme-specific Practice

- Develop a module for detailed training of specific problem phonemes or syllables based on individual user weaknesses.
- Offer tailored exercises for regional and cultural differences in speech patterns.

Emotion and Tone Detection

- Implement AI models to assess emotional tone and provide suggestions for tone modulation to enhance communication.

Cloud-based Processing

- Leverage cloud computing to process large datasets for training and real-time analysis.
- Enable low-latency feedback for users with limited local hardware capabilities.

Community Learning Features

- Introduce peer-to-peer learning platforms where users can practice speaking and provide feedback to one another.
- Host challenges and competitions to foster engagement and collaborative improvement.

Offline Functionality

- Develop offline modes for regions with limited internet access, enabling audio uploads for post-session analysis.

Comprehensive Language Coverage

- Expand the database to include support for tonal languages (e.g., Mandarin, Vietnamese) and complex phonetic languages (e.g., Hindi, Arabic).
- Add regional dialects and slang recognition for localized learning experiences.

Automatic Grammatical Correction

- Integrate grammatical error detection alongside pronunciation feedback for holistic speech improvement.

Accessibility Features

- Include features for users with disabilities, such as speech impediments or hearing impairments.
- Add text-to-speech (TTS) and speech-to-text (STT) capabilities to aid learning.

REFERENCES

1) Adas, Abubakar, and Ahmet Condur. "Automatic Speech Recognition and Its Applications to Language Learning." *Journal of Language and Linguistic Studies* 12, no. 1 (2016): 57–66.
2) Baker, Rachel, and Patricia Munro. "Pronunciation Training in Second Language Learning: Trends and Innovations." *Applied Linguistics Review* 4, no. 2 (2013): 251–272.
3) Chen, Bin, and Yuan Yuan. "Speech Recognition Systems and Their Impact on English Pronunciation Training." *Computer Speech & Language* 46 (2017): 124–133. https://doi.org/10.1016/j.csl.2016.06.002.
4) Chun, Dorothy M. "Computer-Assisted Pronunciation Teaching: The State of the Art." *Studies in Second Language Acquisition* 30, no. 3 (2008): 433–465.
5) Derwing, Tracey M., and Murray J. Munro. "Accent, Intelligibility, and Comprehensibility." *Studies in Second Language Acquisition* 20, no. 1 (1998): 1–25. https://doi.org/10.1017/S0272263198001010.
6) Eskenazi, Maxine. "Using Automatic Speech Processing for Pronunciation Feedback in Second Language Learning." *Language Learning & Technology* 3, no. 2 (1999): 62–76.
7) Flege, James Emil. "Second Language Speech Learning: Theory, Findings, and Problems." In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by Winifred Strange, 233–277. Baltimore: York Press, 1995.
8) Garcia, Eduardo. "Interactive Pronunciation Feedback Using Machine Learning." *Computers and Education* 100 (2016): 19–28. https://doi.org/10.1016/j.compedu.2016.04.003.
9) Hahm, Sohee, and Mark Hasegawa-Johnson. "Acoustic Features for Pronunciation Error Detection." *Speech Communication* 85 (2017): 43–56. https://doi.org/10.1016/j.specom.2016.09.007.
10) Hardison, Debra M. "The Role of Technology in Pronunciation Training." *Journal of Second Language Pronunciation* 1, no. 1 (2015): 29–56. https://doi.org/10.1075/jslp.1.1.02har.
11) Hincks, Rebecca. "Computer Support for Learners of Spoken English." *ReCALL* 17, no. 1 (2005): 5–20. https://doi.org/10.1017/S0958344005000219.
12) Jenkins, Jennifer. *The Phonology of English as an International Language.* Oxford: Oxford University Press, 2000.
13) Jones, Daniel. *English Pronouncing Dictionary.* 19th ed. Edited by Peter Roach, James Hartman, and Jane Setter. Cambridge: Cambridge University Press, 2011.
14) Ladefoged, Peter, and Keith Johnson. *A Course in Phonetics.* 7th ed. Boston: Cengage Learning, 2014.
15) Levis, John M. "Pronunciation and the Assessment of Spoken Language." *International Journal of English Studies* 6, no. 2 (2006): 37–56.
16) Munro, Murray J., and Tracey M. Derwing. "Evaluating the Effectiveness of Accent Reduction Programs for Adult ESL Learners." *Canadian Modern Language Review* 53, no. 4 (1997): 637–660.
17) Neri, Ambra, Catia Cucchiarini, and Helmer Strik. "ASR-Based Feedback on Pronunciation: Does it Really Work?" *Computer Assisted Language Learning* 21, no. 5 (2008): 423–446. https://doi.org/10.1080/09588220802447651.

18) Patel, Rupal, and Yaser Yadegari. "Real-Time Pronunciation Feedback Using Speech Analysis Tools." *Language Testing* 29, no. 4 (2012): 522–541. https://doi.org/10.1177/0265532212436654.

19) Rossi, Roberto, and Giulia Di Gangi. "Pronunciation Modeling for Second Language Acquisition." *Journal of Educational Technology Systems* 48, no. 2 (2020): 225–243. https://doi.org/10.1177/0047239520918584.

20) Strik, Helmer, Catia Cucchiarini, and Lou Boves. "Automatic Evaluation of Dutch Pronunciation by Using Speech Recognition Technology." *Language Learning & Technology* 3, no. 1 (1999): 66–75.

APPENDICES

List of Abbreviations

AI	Artificial Intelligence
AR	Augmented Reality
ASR	Automatic Speech Recognition
CTC	Connectionist Temporal Classification
CNN	Convolutional Neural Network
GPT	Generative Pre-trained Transformer
HSV	Hue, Saturation, and Value (Color Space)
LSTM	Long Short-Term Memory
ML	Machine Learning
NLP	Natural Language Processing
NMT	Neural Machine Translation
POS	Part of Speech
RNN	Recurrent Neural Network
TM	Translation Memory
UML	Unified Modeling Language