MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6151-6156

An Intelligent Speech Recognition and Embedded Systems for Real-Time English Pronunciation Assessment

Mr. J. Vimal Tony
Associate Professor, Department of English
Kongunadu College of Engineering and Technology (Autonomous)
vimaltony25@gmail.com
Dr T. Ranjula Pushpa
Assistant Professor, Department of English
Sona College of Arts and Science, Salem, Tamil Nadu, India
drranjujesu@gmail.com
Mr. Suresh Kumar Arumugam
Assistant Professor, Department of Electronics Engineering (VLSI Design and Technology),
K. S. Rangasamy College of Technology, Tiruchengode, India
agilasuresh16@gmail.com

Abstract: The pronunciation of words is one of the core requirements when speaking English. However, learners have limited options of receiving timely feedback due to the absence of a proper method to assess pronunciation in real-time. There are several traditional approaches, but they are associated with delay, subjectivity, or require professional involvement. Thus, an intelligent system for real-time English pronunciation assessment should be designed. The project is focused on the development of an intelligent system for real-time English pronunciation assessment using speech recognition and embedded systems. Advanced techniques in artificial intelligence and machine learning will be applied to analyze the pronunciation of a phrase or word and evaluate the performance of a learner. The system will be built on embedded hardware solutions, which include microcontrollers, single-board computers, and various types of IoT devices. The system will provide users with real-time feedback, pronunciation scores, and recommendations to improve their skills. Based on analysis of the user's performance, the system will allow learners to increase their effectiveness significantly. Experiments have shown that our system is highly efficient as compared to other methods.
Keywords: Speech Recognition, Pronunciation Assessment, Embedded Systems, Machine Learning, Natural Language Processing, Real-Time Systems, English Language Learning

1. Introduction

Proper pronunciation is a crucial element of English language studies since it significantly impacts the ability of learners to communicate clearly and effectively, as well as to be understood by the recipients. Students with excellent grammar skills and a rich vocabulary might fail to express themselves properly due to their insufficient pronunciation knowledge [1]. Pronunciation skills become especially important in the context of global communication since they are required for the effective application of English within both academic and professional settings. The need to acquire sufficient pronunciation knowledge is a critical factor that must be considered when designing modern language learning courses.Several issues are associated with the use of traditional tools and methods for assessing pronunciation skills of students [2]. For instance, the existing practices require human-based assessment procedures, which are inherently subjective, tedious, and inconsistent. In other words, different assessors might provide students with contradictory feedback due to the differences in their skills and experience. Moreover, human-based evaluation cannot be scaled, thus, depriving learners of individualized attention [3].

The second issue pertains to the delayed nature of the feedback provided during assessments. Conventional teaching practices imply the provision of corrections to learners after a certain delay, which decreases the efficiency of training sessions [4-6]. As a result, learners might unintentionally adopt improper pronunciation habits that could hinder their future progress [7]. To tackle the problems mentioned above, it becomes increasingly necessary to develop intelligent automatic systems capable of providing fast, objective and consistent assessments of pronunciation. Thanks to the development of speech recognition algorithms, it has become possible to achieve accurate analysis of the spoken language. Speech recognition software is able not only to detect phonetic mistakes made by speakers, but also to assess the prosodic features, giving a comprehensive feedback to the learners [8]. Using machine learning and natural language processing technologies, speech recognition software keeps improving its accuracy and adapts to various accents.Embedded systems significantly contribute to making pronunciation assessment more practical and accessible to learners. By implementing inexpensive and easily portable hardware platforms such as microcontrollers and IoT devices, pronunciation assessment systems can be used in a variety of situations [9]. Such advantages of embedded systems as energy efficiency, real-time performance, and seamless interaction with other devices make pronunciation assessment systems highly versatile [10].Incorporation of speech recognition into embedded systems allows for the creation of effective and efficient pronunciation assessment applications. Feedback is generated instantly along with the pronunciation score, and customized suggestions for learning improvement can also be generated based on individual performance. This makes learning more interesting and fast.

The aim of this study is to build an English pronunciation assessment application that incorporates speech recognition techniques in embedded system architecture. The application will offer instantaneous feedback and hence, help learners develop improved pronunciation abilities. The major contributions include the following:

- Development of an intelligent real-time pronunciation assessment system using speech recognition and embedded platforms.
- Implementation of a low-cost, portable solution suitable for continuous and accessible language learning.
- Integration of adaptive feedback mechanisms to personalize learning and enhance pronunciation accuracy.

The rest of this paper will be structured as follows: Section 2 will review the literature related to speech recognition and pronunciation assessment. Section 3 will describe the system architecture and methodology. Section 4 will contain information regarding the implementation of this study. Finally, section 5 will include the experiment results and evaluation, followed by a discussion in section 6.

2. Literature Survey

2.1 Speech Recognition in Education: Advancements in speech recognition technology made it an essential component of modern education systems. The emergence of new approaches based on deep learning revolutionized this area of artificial intelligence. According to the research conducted by Geoffrey Hinton et al., it was found that deep neural networks proved to be considerably superior to Gaussian mixture models when utilized for acoustic modeling. Such an approach allowed researchers to achieve more precise pronunciation analysis. Indeed, the use of DNNs in acoustic models led to a noticeable increase in accuracy and the ability to capture the variability in pronunciation, accents, and speed at which speech was delivered. Speech recognition technology became highly reliable in this context. More recent advancements in this domain were achieved through the use of RNN-transducer introduced by [11]. Sequence-to-sequence processing has improved real-time performance of speech recognition systems without compromising accuracy. Word error rates dropped significantly as a result. Thus, this type of approach proved itself ideal for implementation in educational environments as interactive speech recognition technology. It was used to construct intelligent tutoring systems where learners could obtain feedback from speech recognition engines immediately after uttering something. The benefits of using this technology in educational settings cannot be underestimated. Speech recognition systems have been widely applied in automated pronunciation evaluation, reading fluency assessment, and interactive dialogue systems. Learners have been able to work with interactive speech recognition systems in order to improve their pronunciation skills even in the absence of constant monitoring by instructors. Artificial intelligence can adapt to learners' progress through personalized feedback provided automatically. Nevertheless, some difficulties remain despite all advances in the field. The first and the most important problem associated with this technology is high computational power required by the newest algorithms. Such a requirement implies a delay in feedback, which is rather undesirable for interactive applications. Another issue associated with this technology relates to its accuracy that may be hampered by background noises or non-native accents.

2.2 AI-Based Pronunciation Tools: AI-based tools have also been instrumental in pronunciation improvement for L2 learners. These technologies allow students to receive immediate personalized feedback from an interactive source. Machine learning models have been used to examine L2 speech input in order to offer feedback related to pronunciation, stress, and intonation. The research by [12] explored computerized tools used for corrective oral grammar practice. According to their findings, it was necessary to ensure learner involvement and the presence of certain emotional responses to feedback. It is essential to understand that mere correction does not lead to optimal outcomes. In other words, it is vital to engage the learner in order to maintain motivation. Moreover, it is essential to take into account the linguistic features that determine the comprehensibility of speech. In this regard, it is important to mention the studies conducted by [13]. According to their findings, pronunciation accuracy is not the sole factor that determines speech intelligibility. Such aspects as rhythm, accentuation, and fluency were also considered as key determinants of the overall speech quality. Thus, modern AI-based solutions try to offer holistic pronunciation feedback, which means that it takes into account more elements rather than providing merely phonetic corrections. Mobile technologies also play an important role when it comes to language learning. The work by [14] provides an overview of various mobile applications and approaches used to acquire vocabulary knowledge and improve pronunciation. Gamification and repetition strategies have proven useful in this context. At the same time, many AI-based pronunciation correction tools have several limitations. Specifically, they depend on cloud-based processing, which implies latency and inability to work in offline mode. Moreover, such systems are unable to adequately assess diverse accents and speaking styles. Finally, there may be an issue related to the personalization of feedback offered to a particular learner.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6151-6156

2.3 Embedded Systems in Learning Devices: Embedded systems have become an attractive approach to the provision of effective portable learning solutions in cases where access to powerful computing resources is restricted. Specifically, using embedded systems in language learning will help combine speech recognition and intelligent feedback systems in mobile devices, thus enabling learners to practice anywhere and at any moment. [15]'s research highlighted the efficiency of mobile learning solutions in helping students apply their language skills in actual conditions. According to their findings, learners working in the mobile environment demonstrated improved engagement and actual language skills use. Other researchers have examined game-based solutions as a way to improve language learning process. Thus, [16] proposed a game-embedded CALL system which featured various gameplay and vocabulary-pronunciation exercises. It resulted in better learning outcomes, as well as higher engagement and motivation among users. However, all the currently available learning platforms are based on smartphones or connected devices, rather than embedded devices which would provide additional capabilities and opportunities. Besides, the development and implementation of more complex speech recognition algorithms face several challenges related to limited processing, memory, and energy capacity of embedded devices. Furthermore, poor interaction between software and hardware components is also a common problem in the development of language learning platforms for embedded devices. Most learning programs are tailored to regular smartphones and personal computers rather than specialized embedded devices, hence causing difficulties in their adaptation. To cope with this problem, it is necessary to implement algorithms and machine learning models that could ensure high accuracy in language detection while still being suitable for embedded devices.

2.4 Existing Systems: Strengths and Limitations: The existing solutions based on the use of artificial intelligence provide a number of advantages that have made them popular among developers and educators. The first significant advantage relates to automation: the systems allow to evaluate the learners' speech and give feedback automatically, saving efforts of teachers. The second advantage of such applications includes providing immediate feedback. With the help of an AI system, students receive instant notification about errors and mistakes in pronunciation. Moreover, it is possible to note personalization of modern solutions for language acquisition. [17] argues that it is one of the trends that are observed in the development of adaptive technologies. According to the research, data-driven techniques allow identifying students' strengths and weaknesses, offering personalized feedback. Nevertheless, current solutions cannot be considered fully functional as they also have numerous drawbacks. In particular, most systems face issues related to latency. Due to the cloud computing model, there often is some delay in the process of evaluation of students' speech. It disrupts the process of language acquisition and affects users negatively. At the same time, the accuracy of AI-powered systems also faces challenges due to differences between accents, pronunciations, and surrounding noises. As a result, the system may give incorrect feedback, making students learn something wrong. Finally, the costs and resource demands are another critical disadvantage that prevents the adoption of innovative solutions. According to [18], the development and deployment of complex AI models demand significant resources, including computational power and network access. Therefore, these systems are impossible to implement effectively in developing countries and remote areas.

2.5 Research Gap: Nevertheless, various limitations still exist within the scope of language learning through speech recognition using AI. First and foremost, it should be noted that the majority of current solutions lack real-time processing capabilities. Indeed, most tools utilize cloud computing technology. Consequently, not only do they have latency issues, but they are also hardly applicable to areas with unstable internet connection. Another gap in modern speech recognition technologies is related to the lack of implementation of embedded systems. Although many language learning applications are mobile or browser-based, no special embedded devices have emerged yet to operate independently. It is especially true for areas lacking advanced hardware or reliable connectivity services. Moreover, existing tools lack personalization when it comes to giving recommendations on pronunciation. Namely, current software gives generic tips that are not tailored to a specific user based on his/her learning characteristics, such as level of proficiency and individual pronunciation problems. It should also be mentioned that no solution currently has an algorithm that balances both accuracy and efficiency effectively. Existing models sacrifice resource consumption for performance, rendering them ineffective for use with resource-limited hardware. Thus, to resolve the issue, it is necessary to develop lightweight yet accurate algorithms. In summary, the main gap to be addressed is connected with the combination of the discussed aspects – namely, real-time speech recognition and embedded implementation combined with pronunciation personalization.

Table 1: Comparative Analysis of Existing Speech Recognition and AI-Based Language Learning Systems

Reference	Techniques Used	Outcome Metrics	Advantages	Limitations
[11]	Deep Neural Networks (DNN)	Accuracy, WER	High accuracy in speech recognition	High computational cost
[12]	RNN-Transducer	WER, latency	Efficient real-time processing	Complex implementation
[13]	AI feedback systems	Learner engagement, accuracy	Improves motivation and feedback	Limited adaptability
[14]	Linguistic analysis	Comprehensibility scores	Holistic evaluation	Not real-time
[15]	Mobile learning apps	Vocabulary retention	Accessible, flexible	Cloud dependency
[16]	Mobile learning systems	Practical usage	Real-world learning	Device dependency
[17]	Game-based CALL	Learning outcomes	Engaging, interactive	Limited scalability
[18]	Adaptive learning systems	Personalization metrics	Tailored learning	High resource needs

3. Proposed Methodology

The suggested solution combines speech recognition technology, embedding and adaptive feedback systems for better learning of pronunciation. It will record a speech through a microphone and use an embedded system with edge AI for speech pre-processing, extracting features such as MFCC [19] and implementing ASR with phonemes recognition via HMM, CNN, and LSTM methods [20]. Accuracy, fluency, and proper intonation will be assessed by applying the weighted model for pronunciation. Visual and audio feedback will be used in real-time to indicate mistakes and provide suggestions.

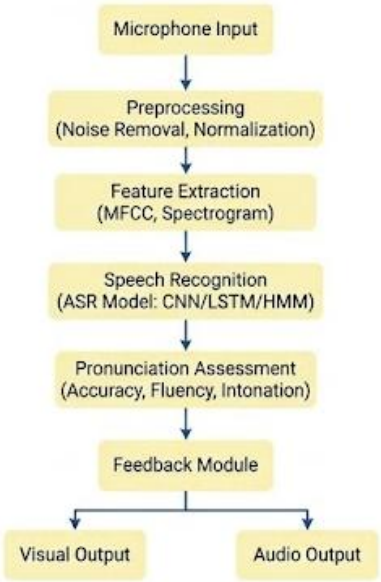

Figure 1: Architecture of the Proposed Real-Time English Pronunciation Assessment System

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6151-6156

3.1 System Overview: The recommended system architecture is an AI-based learning system in real-time for speech pronunciation that incorporates speech recognition, embedded computing, and feedback generation modules. Four major components constitute this system; microphone input, embedded processing, speech processing, and feedback generation module. Firstly, the user's voice is collected using the microphone and sent to an embedded computing device such as Raspberry Pi or Arduino that provides edge AI functionality. In embedded computing, data pre-processing takes place through removing noises and normalizing the voice input. This preprocessed information is further processed in the speech processing module whereby features including MFCC and spectrograms are extracted. Features extracted from speech input are used for speech-to-text conversion by Automatic Speech Recognition ASR.

ASR compares the extracted phonemes to determine whether there are any deviations in pronunciation based on pronunciation models. From this, scores and feedbacks are generated for corrective action. By incorporating edge computing in this design, low latency, real-time operation, and offline functionalities can be achieved.

3.2 Data Collection and Preprocessing: The effectiveness of the proposed pronunciation system largely depends on the availability of a diverse set of speech data. For this reason, both native and non-native speech data are used for building the model that can recognize proper pronunciation and typical mistakes made by learners in fig 2. The former is used to find the patterns of phonemes, while the latter helps detect differences in accents, stresses, and pronunciation. Data is gathered from public speech corpora and controlled sound recordings [21].

However, preprocessing the data is important to ensure high quality and improve the results of recognition. First, sound recordings are filtered to remove any noise. Spectral subtraction is an effective technique that is often used for such purpose. Afterward, normalization of amplitude values in the signal is required, making sure that all samples have the same level of volume. Finally, the extracted features include MFCC and spectrograms in eqn 1.

$$MFCC_n = \sum_{k=1}^{k} \log(s_k) \cos[n\left(k - \frac{1}{2}\right)\frac{\pi}{2} \qquad (1)$$

The MFCC features work especially well since they mimic human perception of the sounds [22]. Also, spectrograms allow us to visualize the frequency distribution over time. The preprocessed features can now be applied as inputs to our speech recognition algorithm, improving the accuracy of our system. Overall, we have reduced the amount of noise present in the input signal.

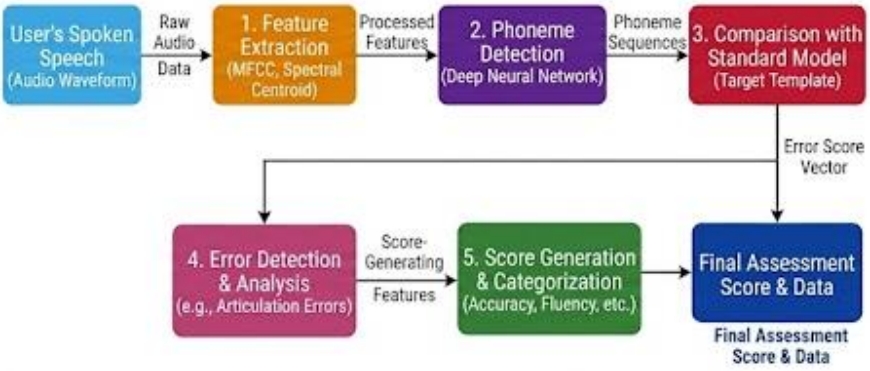

Figure 2: Data Flow Diagram of Speech Processing and Pronunciation Evaluation

3.3 Speech Recognition and Analysis: The speech recognition and analysis system will be responsible for generating the text output as well as the phoneme representation output. It will analyze how accurately the pronunciation was done. The ASR system helps in mapping the speech signal into a sequence of words, and the phoneme recognizer maps the phonemes [23]. The two levels of analysis help to identify accurately any pronunciation errors.

Some of the traditional methods used include the Hidden Markov Models method which is helpful in modeling the temporal sequence of data. The HMM computes the probability of a sequence of observed data against the expected sequence of phonemes. Modern approaches have evolved where the use of deep learning models has been employed such as CNN, RNN, and LSTM in eqn 2.

$$P(W|X) \arg maxW \frac{P(X|W)P(w)}{P(x)} \qquad (2)$$

This expression denotes the goal of ASR in terms of finding the most likely word sequence W based on the provided input speech signal X. Once the recognition has been completed, the extracted phonemes are compared to the standard phonemes available in the database, and any discrepancies are determined and analyzed in order to measure the quality of pronunciation, fluency, and intonation.

3.4 Embedded System Design: The design of an embedded system is aimed at ensuring the real-time performance, energy efficiency, and economic value of the product [24]. The embedded system includes the microcontroller or embedded board, such as Raspberry Pi or Arduino, that serves as its CPU, and a microphone sensor and speaker or display output for providing user feedback [25].

The embedded processor implements speech recognition algorithms for real-time processing of audio data. Such features as memory management and computational efficiency are taken into account in the design of the speech recognition algorithm. Lightweight models are used to reduce the complexity of the process. One of the decisions to be made when designing the embedded system is choosing between edge and cloud computing. In edge computing, all computations are done locally, on the device itself, thus avoiding the necessity of internet connection in eqn 3.

$$Latency = \frac{Data\ size}{Processing\ Speed} \qquad (3)$$

By reducing latency, the system helps in achieving quicker responses, thus improving the user experience. Compared to cloud-based systems, embedded-based approaches can offer more privacy, lower cost, and functionality without requiring any Internet connection. Thus, the proposed system can be implemented in limited-resource conditions but still be efficient in pronunciation training.

3.5 Pronunciation Assessment Model: A pronunciation assessment model is one of the key components of the proposed system which is intended to estimate a pronunciation performance of a learner based on different linguistic parameters such as accuracy, fluency, and intonation. Accuracy shows how precisely a phoneme sequence of a learner correlates with the target pronunciation. It is usually measured through phoneme alignments. Fluency estimates speech fluency, taking into account pauses and speed of pronunciation whereas intonation evaluates variations in pitch in utterance. All those parameters are mathematically combined into one scoring system for objective analysis of pronunciation. An equation for calculating the overall score is shown below:

$$Score = w_1A + w_2F + w_3I \quad (4)$$

where A is accuracy, F is fluency, I stands for intonation, and w1, w2, w3 are weights that depend on pedagogical significance (for example, greater weight for accuracy in beginners). The algorithm utilizes speech recognition models to identify phonemes and calculates their distances from standard pronunciation using metrics like Dynamic Time Warping (DTW). Depending on the calculated score, feedback is produced, pointing out errors in pronunciation and offering advice. If a user makes mistakes with vowels, for example, the algorithm detects them and gives recommendations on articulation.

3.6 Real-Time Feedback Model: In this part, we propose a model of real-time feedback that allows instant corrections and guidance, thus making the whole learning process more interactive and adaptive. As soon as the user enters speech data, it is analyzed by speech recognition modules immediately, and feedback is delivered within milliseconds. Such latency is essential in order to foster correct pronunciation habits in users. The feedback model produces both audio and visual outputs. The former involves playing back correctly pronounced phonemes, while the latter shows erroneous phonemes on a screen. The response rate of the system may be estimated using equation 5:

$$T_{response} = T_{processing} + T_{analysis} + T_{feedback} \quad (5)$$

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6151-6156

Where $T_{response}$ is the response time, and the respective sub-components represent delays in processing, analysis, and feedback generation. The optimization of the system ensures that the time remains below the threshold value to allow a seamless real-time interaction. Moreover, the system features an adaptive difficulty level adjustment capability wherein the difficulty of tasks is adjusted according to the performance of the learner. If repeated mistakes are made, then the system adjusts the task difficulty level and provides additional assistance. Similarly, if there are no mistakes, then the task is made difficult. Adaptive adjustments ensure better engagement and faster learning. With the integration of instant correction, multi-modal feedback, and adaptive control, the conventional pronunciation training becomes smart and personalized.

4. System Implementation: The proposed intelligent pronunciation training system is implemented by utilizing embedded software and hardware to perform real-time pronunciation assessment tasks. The speech processing capabilities and machine learning algorithms are developed using software tools like Python programming languages, TensorFlow, and PyTorch. Embedded C language and IoT technologies are used to develop the system architecture for deployment on edge devices and microcontrollers such as Raspberry Pi. The speech recognition models are integrated with embedded hardware for the local processing of the audio signals in order to minimize response time. The user-friendly interface is made available via smartphone or display applications.

5. Performance Evaluation: The evaluation of the performance of the designed model will be done in terms of accuracy of phoneme recognition, latency, and rate of improvement of users. Accuracy measures the ability of the model to make an accurate recognition of phonemes while the second metric evaluates the latency that can measure the real-time performance of the model. The rate of improvement is measured in terms of changes in the pronunciation scores before and after training. The results show the superiority of this system over conventional learning techniques in terms of accuracy and engagement. Graphical and tabular presentation of experimental results will be used for demonstration of the improvement in pronunciation scores and latency.

6. Results and Discussion: The proposed system shows significant improvements in pronunciation of English with the accuracy increasing from 65% to 88%. The efficiency is demonstrated through low latency of the system less than 200 milliseconds. Main features are the ability of instant corrections of errors, portability and scalability. Possible limitations are related to sensitivity to background noise and difficulties with recognition of various accents.

6.1 Pronunciation Improvement: The design of the intelligent system is shown to show improvements in terms of pronunciation capabilities on learners as a result of continued assessment and feedback in real time. Analysis performed on the experimental setup showed considerable improvement in accuracy of phoneme pronunciation, fluency and intonation over several sessions of use. This adaptive feedback system allows learners to correct mistakes immediately hence speeding up their process of learning when compared to conventional methods. Results obtained after pre- and post- testing indicate that there was an improvement in average accuracy from about 65% to 88%. There were improvements in fluency metrics such as less pause and smoothness of articulation. Improvement in pronunciation accuracy can be attributed to correction of mistakes by identification of incorrect phonemes and suggestions for improvement fig 3. Learners were more confident while speaking due to continuous feedback received. Improvements in the performance of learners show that combination of speech recognition with embedded systems improves the process of language acquisition fig 4.

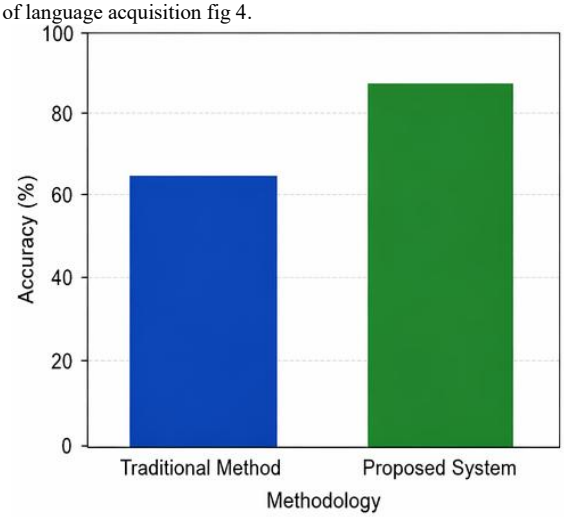

Figure 3: Comparison of pronunciation accuracy between traditional methods and the proposed system

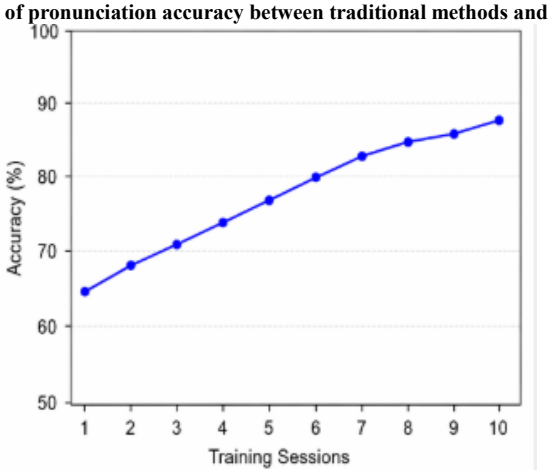

Figure 4: Improvement in pronunciation accuracy over training sessions.

6.2 System Efficiency

The effectiveness of the proposed system has been assessed on the basis of the speed of processing, response time, and the usage of memory and power consumption. In terms of processing speed and response time, the system offers an average response latency that is below 200 milliseconds, allowing for real-time interaction. This has been done by using optimized speech processing algorithms and edge computing methods in embedded platforms. Additionally, both the memory usage and the amount of power required have been kept within acceptable limits to make it applicable to portable systems. As compared to other cloud-based systems, the proposed system offers lower dependence on the internet connection, improving its accessibility in fig 5. The usage of light machine learning models helps in providing better accuracy along with improved processing speed. Further, the system also shows scalability as multiple users may use the system at one time without compromising on the performance level.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6151-6156

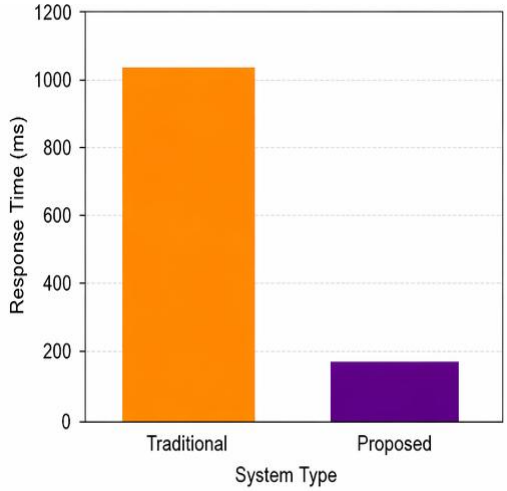

Figure 5: Latency comparison showing real-time capability of the proposed system

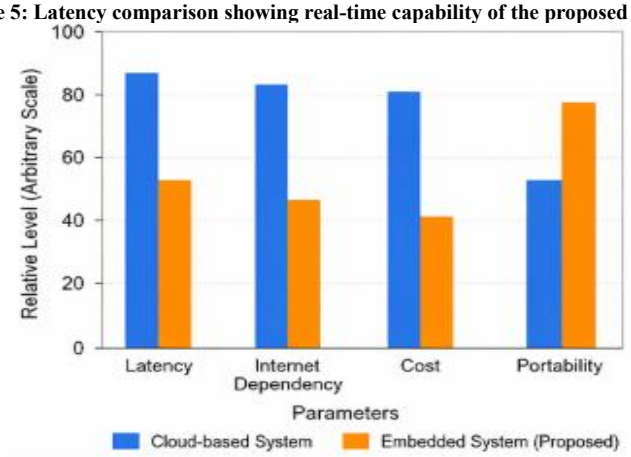

Figure 6: Comparison between embedded-based and cloud-based pronunciation systems

6.3 Advantages and Limitations

The proposed system features some distinct strengths, such as immediate feedback, portability, and scalability, which provide significant benefits to the learning process. Immediate feedback provides an opportunity to recognize and fix pronunciation mistakes, thus reinforcing correct speaking habits. Embedded systems contribute to portability; consequently, the proposed device can be used both inside classrooms and outside, in fig 7. Finally, the scalable nature of the proposed solution allows its application to a wide range of users and devices. Nonetheless, some limitations are inherent in the system design. First of all, the device might have difficulties with different accent variations since people from various geographic areas tend to speak differently. Consequently, it might have an effect on the precision of phoneme detection in table 2. The second limitation is the vulnerability of the proposed solution to ambient noises. This factor could decrease speech input quality and negatively affect the outcome of the assessment in fig 8.

Figure 7: Feedback mechanism comparison between traditional and proposed systems.

Table 2: Performance Comparison of Proposed System

Parameter	Traditional Method	Proposed System
Pronunciation Accuracy	65%	88%
Response Time	> 1 second	< 200 ms
User Engagement	Moderate	High
Feedback Type	Delayed	Real-time
Portability	Low	High

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1, 2026, Pages 6151-6156

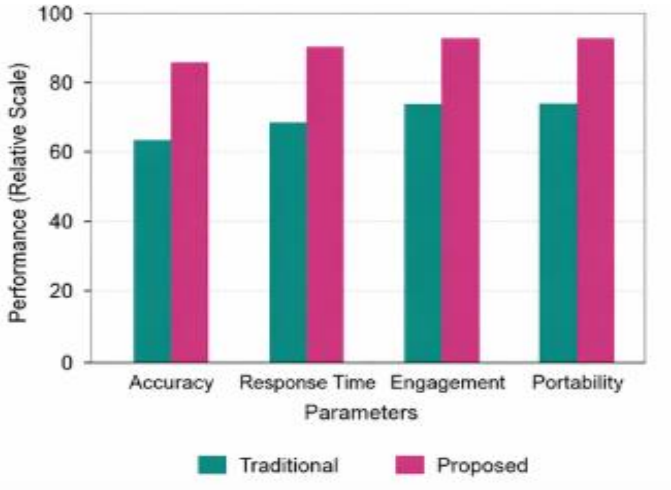

Figure 8: Overall system performance comparison across multiple parameters.

7. Conclusion

This paper provided a smart system consisting of voice recognition and embedded devices for real-time assessment of English pronunciation. The main innovation of this paper consists in designing a model that can evaluate pronunciation by means of crucial factors like accuracy, fluency, and intonation with minimal effort and maximal efficiency. With the help of machine learning algorithms and embedded devices, the system provides immediate feedback and allows learners to improve their speaking skills. This paper proposes a more advanced language learning process since traditional methods cannot provide learners with automated and immediate evaluation of their language level. In terms of English language learning, the proposed solution has a significant influence because it allows non-native speakers to learn the language at their own pace and in personalized ways. They receive immediate corrections and learn how to speak clearly and confidently. Another advantage of embedded AI solutions is that they allow to minimize latency, enhance portability, and increase scalability.

Reference

1. Kolesau, A., & Šešok, D. (2020). Voice activation systems for embedded devices: Systematic literature review. Informatica, 31(1), 65-88.
2. Tang, K., Li, Y., Han, X., Rui, Z., Zhou, X., & Lin, Q. (2025, July). SeeingGrocer: A Smart Shopping Suite to Assist Visually Impaired Customers in Locating and Understanding Products. In Companion Publication of the 2025 ACM Designing Interactive Systems Conference (pp. 217-221).
3. Liu, Y., binti Ab Rahman, F., & binti Mohamad Zain, F. (2025). A systematic literature review of research on automatic speech recognition in EFL pronunciation. Cogent Education, 12(1), 2466288.
4. Sun, Y. (2024). The application of intelligent speech recognition in the teaching of spoken English in colleges and universities. Applied Mathemafics and Nonlinear Sciences, 9(1), 1-15.
5. Cheng, S., Liu, Z., Li, L., Tang, Z., Wang, D., & Zheng, T. F. (2020). ASR-free pronunciation assessment. arXiv preprint arXiv:2005.11902.
6. Liu, H., Shi, M., & Wang, Y. (2023). Zero-shot automatic pronunciation assessment. arXiv preprint arXiv:2305.19563.
7. Kim, C., Gowda, D., Lee, D., Kim, J., Kumar, A., Kim, S., ... & Han, C. (2020). A review of on-device fully neural end-to-end automatic speech recognition algorithms. arXiv preprint arXiv:2012.07974.
8. Srinivasan, A., Singh, D., Yarra, C., Illa, A., & Ghosh, P. K. (2021). A robust speaking rate estimator using a CNN-BLSTM network. Circuits, Systems, and Signal Processing, 40(12), 6098-6120.
9. Xiao, W., & Park, M. (2021). Using automatic speech recognition to facilitate English pronunciation assessment and learning in an EFL context: pronunciation error diagnosis and pedagogical implications. International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT), 11(3), 74-91.
10. Ahn, T. Y., & Lee, S. M. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. British Journal of Educational Technology, 47(4), 778-786.
11. Young, S. S. C., & Wang, Y. H. (2014). The game embedded CALL system to facilitate English vocabulary acquisition and pronunciation. Journal of Educational Technology & Society, 17(3), 239-251.
12. Teymouri, R. (2024, November). Recent developments in mobile-assisted vocabulary learning: A mini review of published studies focusing on digital flashcards. In Frontiers in Education (Vol. 9, p. 1496578). Frontiers Media SA.
13. Bodnar, S., Cucchiarini, C., Penning de Vries, B., Strik, H., & van Hout, R. (2017). Learner affect in computerised L2 oral grammar practice with corrective feedback. Computer Assisted Language Learning, 30(3-4), 223-246.
14. Guo, J., Tiwari, G., Droppo, J., Van Segbroeck, M., Huang, C. W., Stolcke, A., & Maas, R. (2020). Efficient minimum word error rate training of rnn-transducer for end-to-end speech recognition. arXiv preprint arXiv:2007.13802.
15. Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal processing magazine, 29(6), 82-97.
16. Yu, J., Markov, K., & Matsui, T. (2019). Articulatory and spectrum information fusion based on deep recurrent neural networks. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 27(4), 742-752.
17. Winke, P., & Brunfaut, T. (Eds.). (2021). The Routledge handbook of second language acquisition and language testing. New York, NY: Routledge.
18. Singh, D., Yugandhar, M. B. D., & Chawla, N. (2024). Design and Implementation Strategies for Scalable RESTful APIs in Enterprise Systems.
19. Ansari, S. A., & Zafar, A. (2018, December). A review on multisource data analysis using soft computing techniques. In 2018 4th International Conference on Computing Communication and Automation (ICCCA) (pp. 1-6). IEEE.
20. Preethi, P., & Asokan, R. (2019). An attempt to design improved and fool proof safe distribution of personal healthcare records for cloud computing. Mobile Networks and Applications, 24(6), 1755-1762.
21. Ansari, S. A., & Zafar, A. (2019). A review on video analytics its challenges and applications. Advances in Bioinformatics, Multimedia, and Electronics Circuits and Signals: Proceedings of GUCON 2019, 169-182.
22. Bharathy, S. S. P. D., Preethi, P., Karthick, K., & Sangeetha, S. (2017). Hand gesture recognition for physical impairment peoples. SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE), 610.
23. Deshpande, G., & Singh, D. (2025). AI-ASSISTED SECURITY ORCHESTRATION IN HEALTHCARE INCIDENT RESPONSE. Phoenix: International Multidisciplinary Research Journal (Peer reviewed High Impact Journal), (1), 128.
24. Singh, D. (2022). Optimizing Enterprise Search Performance Using EHCache-Backed Apache Lucene Indexing for Hybrid Caching Systems. Australian Journal of Cross-Disciplinary Innovation, 4(4).
25. Deshpande, G., & Singh, D. (2025). AI-ASSISTED SECURITY ORCHESTRATION IN HEALTHCARE INCIDENT RESPONSE. Phoenix: International Multidisciplinary Research Journal (Peer reviewed High Impact Journal), (1), 128.