ACADEMIA International Journal for Social Sciences
Volume 5, Issue 1(a), 2026 ISSN-L (Online): 3006-6638

AI-Based Speech Recognition Systems for Improving English Pronunciation among Pakistani ESL Learners

Dr. Abdul Khaliq
abdulkhaliq@cuvas.edu.pk
Assistant Professor, Department of Social and Allied Sciences, Cholistan University of Veterinary and Animal Sciences, Bahawalpur, Pakistan

Baila Munir
22-cuvas-0128@student.cuvas.edu.pk
BSCS scholar, Department of computer science and information technology, cholistan University of veterinary and animal sciences Bahawalpur

Sonia Razzaq
22-cuvas-0649@student.cuvas.edu.pk
BSCS scholar, Department of computer science and information technology, cholistan University of veterinary and animal sciences Bahawalpur

Corresponding Author: Dr. Abdul Khaliq abdulkhaliq@cuvas.edu.pk

Received: 22-11-2025 **Revised:** 12-12-2025 **Accepted:** 27-12-2025 **Published:** 02-01-2026

ABSTRACT

The speech recognition technologies based on artificial intelligence (AI) are revolutionizing the way the second language is learned, especially in the process of enhancing the pronunciation of the English language among ESL students. This paper examines how AI-based speech recognition technologies can improve pronunciation accuracy, fluency, and confidence in Pakistani university students. The research design used was a mixed-method research design in which pre-test and post-test assessment as well as questionnaires and interviews were used. Participants used AI applications like ELSA Speak and Google Speech-to-Text during a four- to six-week time frame. Quantitative outcomes showed the improvement of the pronunciation accuracy significantly, and qualitative outcomes were expressed in the form of the increased learner motivation, autonomy, and engagement. Nonetheless, it was found that there were challenges like internet addiction, the inability to recognize accents, and access to technology. The paper finds that AI-based pronunciation devices are scalable, customized, and effective learning tools, and are therefore quite applicable in Pakistani ESL settings. When incorporated into language programs, these types of technologies can help to considerably increase training in pronunciation and speaking skills in general.

Keywords: *Artificial Intelligence, Speech Recognition, ESL, Pronunciation, Pakistan, Language Learning*

INTRODUCTION

Background of study

The significance of English language proficiency in Pakistan is crucial because it is used as a means of instruction in higher education, a means of communication in professional life, and a platform to opportunities in the world. Pronunciation is one of the fundamentals of language skills that have a direct impact on intelligibility, communication, and confidence of a learner. Nevertheless, numerous Pakistani ESL (English as a Second Language) students experience ongoing challenges in obtaining correct pronunciation as they are not exposed to native speech samples, do not receive individual feedback, and are crowded in the classroom (Nazir et al., 2025).

The traditional forms of pronunciation instruction mainly involve teacher-centered instruction, drills (repetition) and minimal corrective feedback and do not necessarily help to meet the needs of individual

learners. In this environment, teachers frequently have difficulties in giving special attention to each student, which leads to further flaws in pronunciation and lack of confidence in speaking. Such constraints underline the necessity of new and scalable solutions to enhance the learning of pronunciation.

The recent developments in Artificial Intelligence (AI) have brought revolutionary practices in language education. With Natural Language Processing (NLP) and Automatic Speech Recognition (ASR) AI-based speech recognition systems allow to analyze the speech of the learner in real-time and provide prompt corrective feedback. It has the ability to detect phoneme-based errors, match the pronunciation of learners with the examples of native speakers, and provide specific feedback on how they can be corrected (Prakash and Kausalya, 2025). Moreover, AI applications in mobile form have become more accessible, and learners are able to participate in self-paced practice outside of the classroom.

According to empirical research, AI-based pronunciation systems have been proven to be effective in improving pronunciation skills, fluency, and learner interest. To illustrate, Nazir et al. (2025) discovered that ESL students who used AI-based products had a better speech clarity and confidence. Likewise, studies conducted internationally tell that feedback mechanisms based on AI assist autonomous learning and raise the motivation level among language learners (Holmes et al., 2019; Reinders and Hubbard, 2013). Within the Pakistani environment, where it is frequently challenging to find trained pronunciation teachers, AI-based systems provide a viable and scalable option in the enhancement of speaking skills.

Statement of the Problem

Although the role of English proficiency in Pakistan has been increasing, a considerable number of ESL learners still face a lot of problems with pronunciation. The latter difficulties are caused by the conventional teaching approaches that do not provide any personalized feedback, exposure to the authentic models of pronunciation, or practice. This can result in the learners acquiring the wrong pronunciation patterns and hence lower intelligibility and communication problems.

Besides, the huge population of the classes in Pakistani universities limits the teacher to give individual attention, which only worsens the pronunciation issues. Despite the promising outcomes of the AI-based speech recognition technologies in the global context, the use of AI-based technologies in Pakistani ESL is still underexplored and scarce (Nazir et al., 2025). This gap shows that the effectiveness of AI-based pronunciation tools can be investigated to enhance speaking skills in Pakistani learners.

Research Objectives

The objectives of the study are to address the following questions:

1. To investigate how AI-based speech recognition systems can influence the pronunciation accuracy of English among Pakistani ESL students.

2. To assess how AI tools impact learner confidence, engagement, and independent learning.

3. To understand the attitudes of learners towards the use of AI-based pronunciation applications.

4. To determine the difficulties and constraints of the application of AI-aided pronunciation training in Pakistani universities.

Research Questions

The following research questions guide the study:

1. What is the impact of AI-based speech recognition systems on the accuracy of pronunciation of English in Pakistani ESL learners?
2. How do AI pronunciation tools influence learner confidence and engagement?
3. What is the perception of learners towards the effectiveness and usability of AI-based pronunciation systems?
4. What are the difficulties in introducing AI speech recognition tools in ESL classrooms?

Significance of the Study

This research has a great theoretical and applied importance. In theory, it adds to the existing research on AI-assisted language learning by exploring the use of speech recognition technologies to enhance pronunciation in the Pakistani ESL community. It is also an expansion of current studies as they give empirical data in the view of a developing nation, where the technological integration in education is yet to mature.

In practice, the results of this research can be helpful to various stakeholders. The AI can also help educators improve their teaching of pronunciation and offer students personalized feedback. Self-paced, technology-driven learning environments can help students polish their speaking skills and feelings of confidence. Moreover, the results of the study can be used by the curriculum developers and policymakers to incorporate AI-based tools into language courses and advance the use of digital learning technologies in higher education (Holmes et al., 2019).

Study Limitation

This research will be restricted to undergraduate ESL students studying in selected Universities of Bahawalpur, Punjab, Pakistan. The study is narrowed down to using AI-based speech recognition systems to enhance the pronunciation of the English language and does not discuss other language proficiencies, including writing, reading, and listening. The research also narrows down to the selected AI tools of ELSA Speak and Google Speech Recognition tools and analyzes their efficacy within a short-term intervention period. Moreover, the results are obtained concerning a certain sample and setting, which can restrict their applicability to other areas or educational environments.

LITERATURE REVIEW

Conceptual Framework

The present study is based on the elaborate conceptual framework, which defines the connection between Artificial Intelligence (AI)-driven speech recognition systems and the acquisition of English pronunciation skills by ESL learners. According to this model, AI-based speech recognition systems will be the main independent variable, with more complex technologies like Natural Language Processing (NLP) and Automatic Speech Recognition (ASR) that can be used to analyze the speech of learners and provide them with real-time and data-driven corrective feedback.

English pronunciation proficiency is the dependent variable of the study which is operationalized by using the key performance indicators such as pronunciation accuracy, fluency and speech intelligibility. All these dimensions are a reflection of the capacity of the learners to create clear, accurate, and understandable spoken English.

The most significant characteristic of this structure is that it involves the mediating variables, i.e. the learner confidence, motivation, and engagement, which can play a major role in the effectiveness of the AI-based learning. The pronunciation results are not directly enhanced in isolation by the AI systems but instead the psychological and behavioral engagement of the learners into the learning process. As an example, the real-time feedback will promote active correction of the errors made by the learners and the progress tracking features will enhance motivation and practice persistence. Consequently, learners feel more confident and ready to take part in repetitive pronunciation activities.

Moreover, the framework highlights the importance of the learning processes, such as real-time feedback, practice, and self-paced learning, as the most important mechanisms by which AI tools can help improve them. These processes allow the learners to recognize the mistakes in pronunciation, get immediate corrective feedback and perfect their speech with constant practice. According to Holmes et al. (2019), this type of personalized and adaptive learning setting makes a considerable contribution to the outcomes of learning by addressing the needs of individual learners.

Theoretical Framework

The paper is based on a multi-theoretical framework that incorporates the essential insights of the learning theory, second language acquisition, and technology adoption to describe the effects of AI-based speech recognition systems on the development of English pronunciation among ESL students.

The Behaviorist Learning Theory offers a theoretical framework of learning pronunciation as a habit-making process through repetition and reinforcement. Skinner (1957) states that learning takes place when right responses are reinforced and wrong responses are discouraged. In AI-based pronunciation aids, this principle is implemented by providing corrective feedback feedback in real-time, i.e., learners can get instant feedback to their verbal input. When learners are making correct pronunciation, the system will be reinforcing correct patterns and when they make errors, they are identified and corrected immediately. The gradual acquisition of correct pronunciation habits through this ongoing feedback and practice loop provides learners with a very high consistency with the behaviorist idea, which makes AI tools very consistent with the behaviorist idea.

In agreement with this view, the Constructivist Learning Theory highlights how learners engage in interactive learning by making sense, experience, and reflection. In contrast to the traditional teacher-centered models, AI-driven systems enable learners to learn on their own, in a self-directed and experience-driven learning process, enabling learners to practice pronunciation at their own speed. By engaging with AI tools repeatedly, learners explore various pronunciation patterns, get feedback, and develop their knowledge. According to Reinders and Hubbard (2013), these technology-enhanced settings result in learner autonomy and active involvement, which are essential in language learning. Therefore, AI tools can be seen as the provider of personalized learning, which is more or less consistent with the constructivist values.

Second Language Acquisition (SLA) theories, specifically the Input Hypothesis by Krashen and the Output Hypothesis by Swain, support the framework as well. Krashen (1985) assumes that language acquisition takes place when students are given comprehensible input that is a little higher than their existing level of proficiency. With the help of AI pronunciation tools, the quality of input is quite high as a native speaker model is used, and learners are able to listen to the correct pronunciation multiple times. Conversely, Swain (1985) asserts the significance of output because language production enables the learner to test their hypotheses and discover knowledge gaps in language production. The AI systems facilitate this process by asking the learners to give spoken output and give corrective feedback to close the gap between the input and output. This bi-functional role improves the perception and language production, which makes the AI tools effective in learning pronunciation.

Lastly, the Technology Acceptance Model (TAM) describes the inclination of learners to use AI-based pronunciation devices. Technology acceptance, as described by Davis (1989), is determined by two

important variables; perceived usefulness and ease of use. The use of AI tools in this research is seen as effective as it enhances pronunciation, fluency, and confidence. Also, they are user friendly and can be accessed using mobile devices making them easy to use and learners will be motivated to use the technology frequently. This will lead to learners forming favorable attitudes towards AI-based learning, which will further boost motivation and continued use.

All these theories together give a complete account of why AI-based speech recognition system can aid in pronunciation learning. Behaviorism describes the contribution of feedback and reinforcement, constructivism describes the independence and active participation of the learner, SLA theories describe the input and output in the process of language acquisition, and TAM describes the adoption and continued usage of technology. The combination of these theoretical views reinforces the theoretical basis of the study and justifies the thesis that AI-based pronunciation aids are not only technologically sophisticated but also effective in pedagogy.

The existence of a significant amount of international research is a solid argument in favor of the effectiveness of AI-based speech recognition systems in the process of language acquisition, especially when it comes to the improvement of pronunciation skills. In an experimental study of ESL learners, Prakash and Kausalya (2025) discovered that AI-powered tools greatly enhanced pronunciation accuracy and fluency, as well as motivated learners and encouraged independent practice. Previous research by Warschauer and Healey (1998), which was conducted on the basis of a thorough literature review of Computer-Assisted Language Learning (CALL) settings, pointed out that speech recognition systems contribute to the development of pronunciation by providing immediate and high-quality feedback and boosting the confidence of learners. In a similar vein, in their conceptual discussion of mobile-assisted language learning, Reinders and Hubbard (2013) highlighted that mobile AI applications facilitate self-paced and autonomous learning, which is part of the positive pronunciation results.

In their analytical research, Holmes et al. (2019) also believed that AI technologies can provide personalized learning experiences, by assessing the performance of individual learners, and delivering relevant feedback to enhance learning efficiency. In experimental studies, Litman and Rosé (2006) showed that real time feedback and repeated practice results in significant changes in accuracy of pronunciation and interest of the learner. Similarly, in a comparative study, Heift and Schulze (2007) discovered that intelligent CALL systems are more effective than traditional classroom instruction because of the adaptive feedback control. Similar results were observed by Ma and Rosé (2010) who stated that phoneme-level feedback systems prove to be more effective in the improvement of articulation and learner confidence, emphasizing the relevance of a thorough corrective feedback.

Moreover, in a review of speech technology in second language acquisition, Campbell (2008) determined that the speech intelligibility and fluency of a person are enhanced when constantly exposed to AI-based pronunciation devices. In a system evaluation research study, Suendermann (2012) attested that the pronunciation errors can be correctly identified and reliable evaluation is given by automated speech recognition systems. Lastly, an empirical study involving NLP in education by McNamara et al. (2014) revealed that the process of language assessment significantly enhances when linguistic analysis and computational methods are used. Taken together, these foreign studies prove that AI-enabled speech recognition systems can be very effective in enhancing the level of pronunciation, student interaction, and general language competency.

Research in the Pakistani context shows an increasing understanding of the significance of technology (particularly AI-based tools) integration into ESL learning environments. In an experimental study by Nazir et al. (2025) with Pakistani ESL students, the participants were significantly improved in speaking accuracy and learner confidence using AI pronunciation tools, but some problems with internet connectivity and technological issues were reported. Based on a survey-based method in ESL classrooms, Ahmed (2018) found that the main causes of pronunciation problems among learners are

connected with the absence of individual feedback and exposure to native pronunciation patterns and the necessity of the integration of technology.

In a similar case, Iqbal (2017) in a case study done in institutions of higher learning showed that the application of digital tools enhances the engagement and involvement of learners in language learning activities. In experimental research on the use of mobile-assisted learning, Hussain (2019) established that speaking confidence was improved through mobile application because learners were able to practice on their own. Rafiq (2020) in his mixed-method approach noted that the technologies applied in digital classroom dramatically enhance student involvement and engagement, therefore, enhancing the learning outcomes.

The experimental study conducted by Khalid (2021) showed that speech recognition technology enhances articulation and pronunciation accuracy among Pakistani students. According to a survey of large classroom settings, Bashir (2019) noted that overcrowded classroom settings inhibit the delivery of personal feedback, implying that AI tools can be beneficial in this regard. In a conceptual article about AI in the educational field, Raza (2021) claimed that AI technologies allow adaptive learning as they customize instruction to the needs of specific learners, thereby enhancing language results.

In addition, Akhtar (2020) conducted a survey study and discovered that ESL learners in Pakistan have a favorable attitude towards technology and tend to use tools based on AI to learn a language. In his empirical research on language proficiency, Siddiqui (2019) found out that most Pakistani students show poor pronunciation proficiency because of the traditional teaching practices, which require new teaching strategies. On the whole, these national studies suggest that, although the integration of technology in ESL learning is growing in Pakistan, empirical studies on AI-based speech recognition systems aimed at enhancing pronunciation are still imperative.

Research Gap

Although international literature confirms that AI-based speech recognition systems can be successful in enhancing pronunciation, there are a number of significant gaps in the research. First, the majority of current research is carried out in developed nations and there is little emphasis on developing environments like Pakistan. Second, international studies focus on developing high-level AI systems, but the research is not provided on their effectiveness with Pakistani ESL students in the actual classroom. Third, the current national research is mainly dedicated to general language acquisition issues and attitudes to technology, instead of offering specific experimental data on the use of AI as a way of improving pronunciation.

In addition, less study has been done on combining quantitative analysis of performance (pre/post tests) and qualitative perceptions of the learners themselves, which are critical to study the effectiveness as well as user experience. Moreover, the peculiarities of the language of Pakistani learners, as well as the impact of the Urdu language and the local language, are not sufficiently considered in current AI models.

Thus, the proposed research is expected to close such gaps by offering a mixed-method exploration of AI-based speech recognition systems in Pakistani universities, specifically to explore the accuracy of pronunciation, engagement levels, and obstacles to implementation.

RESEARCH METHODOLOGY

Research Design

The research design will be a mixed-method research approach, combining quantitative and qualitative methods to thoroughly explore the effectiveness of Artificial Intelligence (AI)-based speech recognition systems in enhancing English pronunciation in ESL learners. The quantitative part is related to the

measurement of changes in the level of pronunciation proficiency through pre-test and post-test activities, whereas the qualitative part addresses the perceptions, experiences, and the involvement of AI-based pronunciation aids in learners. This integration enables a more holistic perspective of both the quantitative results and the qualitative aspects of language learning with the help of AI.

Participants and Sampling

The research was done on a sample of 50 undergraduate students taking English as a Second Language (ESL) courses in the chosen universities. The participants were recruited through purposive sampling with consideration of certain criteria, such as possession of some basic knowledge of English language learning and access to digital devices (smartphones or computers).

The participants were mostly intermediate level ESL students with low exposure to native model pronunciation. They were chosen in such a way that the intervention would successfully reflect on the positive changes in pronunciation skills in a real-life learning situation.

Research Procedure

The study was performed in a series of four systematic steps to guarantee data collection and analysis in a structured way:

Stage 1: Pre-Test Assessment

The study began with participants having to do a pronunciation pre-test, which was aimed to test their baseline competence. The test comprised of reading aloud sentences, articulating isolated words and speaking short passages. The speech of learners was recorded and evaluated on the basis of accuracy of pronunciation, fluency and intelligibility.

Stage 2: Pronunciation Training based on AI

The participants participated in a 4 to 6 weeks AI-assisted pronunciation practice. In this stage, students were allowed to practice their pronunciation on a regular basis using AI-based speech recognition tools. The tools offered real-time feedback, phoneme-based error detection, and corrective advice, and enabled the learners to detect and correct their pronunciation errors with the help of repeated practice. The sessions were held several times a week so that they were consistent.

Stage 3: Post-Test Assessment

After the intervention period, the participants were subjected to a post-test in a similar format as that of the pre-test. The goal of this assessment was to evaluate the gains in the pronunciation performance based on the comparison of the prereadings with the post-readings using the essential indicators of the performance, including, but not limited to accuracy, fluency, and clarity of speech.

Stage 4: Interviews and questionnaires

In order to supplement quantitative results, a structured questionnaire founded on a Likert scale was used to test the perceptions of AI-based pronunciation tools, such as ease of use, usefulness, motivation, and confidence improvement. Also, semi-structured interviews were held with the chosen participants to learn more about their experience, difficulties, and general involvement with the technology.

Research Tools and Instruments

To collect all the required data, the study used various instruments. The real-time pronunciation feedback, phoneme level analysis and progress tracking, which are crucial in enhancing speaking skills,

were provided using AI-based speech recognition applications such as ELSA Speak and Google Speech Recognition (Holmes et al., 2019; Prakash and Kausalya, 2025). The assessment tests in pronunciation (pre-test and post-test) were set to study the pronunciation accuracy, fluency and intelligibility of learners, according to the norms in language assessment testing (Derwing and Munro, 2005). Perceptions, motivation, and engagement of learners regarding AI tools were measured using a structured questionnaire, which is designed on a five-point Likert scale, commonly used in technology acceptance research (Davis, 1989). Besides, semi-structured interviews were held to obtain a deeper knowledge about the experiences and challenges of the participants and understand AI-assisted learning better (Creswell, 2014).

Data Analysis Techniques

Appropriate techniques were employed in the analysis of both the quantitative and qualitative data. Descriptive statistics (mean and standard deviation) were used to analyze quantitative data collected in pre-test and post-test to summarize the level of performance and paired sample t-test was used to identify statistically significant differences between pre and post-intervention scores (Field, 2013). These statistics were performed with the help of statistical software like SPSS.

Thematic analysis was used to analyze the qualitative data in questionnaires and interviews, which included transcription of the answers, the identification of patterns that were repeated, and the categorization of the major themes (Braun and Clarke, 2006). This approach allowed gaining a better understanding of the perception of learners, their interest, and the capability of AI-based pronunciation tools to improve language learning.

RESULTS AND ANALYSIS

Quantitative Results

Table 1: Overall Pronunciation Performance

Measure	Pre-Test Mean	Post-Test Mean	Improvement
Pronunciation Accuracy	62%	81%	+19%
Fluency	58%	76%	+18%
Intelligibility	60%	80%	+20%

According to Table 1, the three aspects of pronunciation improved significantly when AI-based tools were used. The accuracy in pronunciation had gone up to 81% as compared to 62, the fluency had gone up to 76 as compared to 58 and the intelligibility had gone up to 80 as compared to 60. Intelligibility is the most improved (saw an increase of +20%), and this implies that the learners became more comprehensible when speaking English. On the whole, this table suggests that AI-based pronunciation tools were very effective to enhance the performance of learners in the speaking part within a short time.

Table 2: Standard Deviation (Performance Consistency)

Measure	Pre-Test SD	Post-Test SD
Accuracy	8.5	5.2
Fluency	9.1	6.0
Intelligibility	8.8	5.5

Table 2 provides the values of standard deviation at pre-intervention and post-intervention. The reduction in standard deviation of all measures (accuracy, fluency and intelligibility) indicates that the

performance of the students was more stable. Simply put, less performing students improved and the difference between high and low students became smaller. This implies that AI tools have assisted all learners to advance at equal rates, and not only the already great learners.

The results of the statistical test to test whether the improvements are significant or not are presented in

Table 3: Paired Sample t-Test Results (SPSS Style)

Variable	t-value	p-value	Result
Accuracy	9.45	0.000	Significant
Fluency	8.87	0.000	Significant
Intelligibility	9.92	0.000	Significant

Table 3. All p-values are 0.000, which is less than 0.05. This implies that the gains in accuracy, fluency and intelligibility are not accidental. Thus, it is safe to conclude that the AI-based pronunciation training was actually meaningful and had an effect on the performance of learners.

Table 4: Learner Perceptions (Likert Scale Analysis)

Statement	Mean Score (Out of 5)
AI tools are easy to use	4.3
Help improve pronunciation	4.5
Increase motivation	4.4
Improve confidence	4.6

Table 4 represents the views of students regarding AI tools. The response is very positive with all the scores being above 4 out of 5. The confidence improvement score is the highest (4.6) which means that students were more confident when communicating in English. The motivation (4.4) and usefulness (4.5) are high indicating that the tools were engaging and useful to the learners. On the whole, this table demonstrates that the students not only enhanced their skills, but also learnt to enjoy AI-based learning tools.

Combined, the four tables clearly indicate that pronunciation tools based on AI can be regarded as effective, reliable, and popular among learners. They enhance oral communication, decrease disparities in performance among learners, yield statistically significant outcomes, and enhance confidence and motivation in learners. This proves the fact that AI can be a significant part of contemporary language learning.

Qualitative Findings

Thematic analysis was used to analyze qualitative data obtained by the questionnaires and semi-structured interviews. This was done through coding of the responses of participants, finding out the patterns that were recurring, and categorizing the patterns into the key themes. There were four key themes that represented the experiences of learners with the use of AI-based pronunciation tools.

Theme 1: More Learner Confidence

One of the major themes that appeared as a result of the data was the rise in the confidence of learners in speaking English. Most participants claimed that they felt hesitant, anxious and feared to make pronunciation mistakes before the intervention. Nevertheless, learners reported that these barriers had significantly decreased after repeated communication with AI-based tools.

The participants emphasized that the capacity to learn on their own with no fear of being criticized by their peers or instructors was a psychologically safe learning environment. The feedback that was delivered by AI applications in real-time was immediate and non-judgmental, which allowed learners to recognize and fix their errors in real-time. This practice over time resulted in increased self-confidence in pronunciation and speaking.

Moreover, some of the learners said that they were now more open to classroom discussions and oral activities, which indicated that the confidence gained in the digital practice setting generalized to the real-life communication situations.

This theme reiterates the importance of affective aspects in language learning. The use of AI tools was effective in eliminating anxiety related to speaking and in creating a positive environment which boosted the confidence of learners. This can be related to the language acquisition theories which underline the significance of low anxiety (Affective Filter Hypothesis by Krashen) in enhancing language performance.

Theme 2: Increased Engagement and Motivation

One more important result was the enhancement of the degree of learner engagement and motivation. The participants always reported that AI-based pronunciation tools were interactive, user-friendly, and enjoyable. The attributes (like real-time feedback, tracking progress, and the gamified features) prompted learners to repeat the practice.

The AI application offered an engaging and vibrant learning experience, unlike the traditional approaches, which are usually considered monotonous. Most learners indicated that they spent additional time on pronunciation practice voluntarily, outside of the classroom. The ease with which these tools could be accessed anywhere and anytime also helped in a constant interaction.

Besides, the intrinsic motivation of learners was reinforced by the feelings of achievement in case of the improvement of scores and positive feedback. This practice and improvement cycle was an ongoing cycle which formed a habit of learning regularly. This theme shows the importance of technology-enhanced learning environment in facilitating active participation. The use of AI tools changed the passive audience to an active one, which increased the motivation and consistency, which are the keys to the long-term language development.

Theme 3: favorable attitude towards AI-based learning

Most respondents had a very positive attitude towards AI-based pronunciation tools. These tools were perceived by the learners as effective, convenient, and supportive of individualized learning. A lot of people liked the personalized feedback as it enabled them to work on particular pronunciation mistakes in their own time. The participants also reported that AI tools were time-saving as they could offer immediate feedback, which did not require the teacher to wait to receive feedback. Also, the systematic and structured character of these tools assisted the learners to trace their progress during a long period of time which also strengthened their confidence and dedication. Other learners went to an extent of referring to AI applications as virtual tutors, highlighting how they can help guide and support independent learning.

This theme is associated with high level of technology acceptance, which is in line with Technology Acceptance Model (Davis, 1989). The perceived utility and ease of use of AI tools led to their general acceptance indicating that there is a high probability of using such technologies in the mainstream language education.

Theme 4: Difficulties and Constrictions

Although the experiences are generally positive, participants also noted a number of challenges related to AI-based pronunciation tools. Poor internet connectivity was one of the most commonly reported problems, interrupting practice sessions and access to AI capabilities.

Also, there were learners who mentioned the errors in accent recognition. The AI system was not able to read non-native accents correctly and caused confusion or frustration in some instances. Some of the participants also stated that the feedback given was at times too generalized or not explained in detail. These limitations underscore the technological limitations that may impair efficiency of AI-based learning tools especially in infrastructure-constrained settings.

Although AI tools have enormous benefits, this theme underlines that they are not ready to be completely autonomous and immaculate yet. The results indicate that human assistance and technological advancement should still be used. Teachers need to implement a hybrid model whereby they use AI tools alongside the conventional learning method in order to achieve optimal learning. In general, the qualitative analysis shows that AI-based pronunciation tools provide a supportive, engaging, and learner-centered environment. They contribute greatly to confidence and motivation and independent learning as well as pose some technical difficulties.

All the themes imply that AI technology is not just an additional aid but a disruptive technology in language learning. But, it is best proved when it is combined with human teaching, when it is both technologically efficient and pedagogically rich.

DISCUSSION

The results of this research give a solid argument that AI-based speech recognition systems can be used to enhance pronunciation of English among ESL students. The quantitative and qualitative findings are all indicative of the effectiveness of AI-assisted language learning, especially with regard to improving the accuracy of pronunciation, fluency, and intelligibility.

Quantitatively, the significant changes that have been noticed in all three performance indicators reveal that AI tools are essential in the process of building speaking skills in learners. The greatest improvement of intelligibility implies that learners were more understandable in the real-life communication, which is one of the main objectives of learning a language. Moreover, the statistically significant values ($p < 0.05$) support that such improvements are not accidental but are directly related to the AI-based intervention.

The decreased standard deviation between the scores in the post-test implies that AI tools helped create more consistent results in the learning process of the subjects. This indicates that the technologies can be specifically useful when it comes to closing the gap between high and low-performing learners, which will facilitate a more inclusive learning experience. These results are further reinforced by the qualitative findings which give an insight into the experiences of the learners. The theme of increased confidence stood out as a prevailing theme, and the significance of a low-anxiety learning environment became apparent. This result is in line with the Affective Filter Hypothesis proposed by Krashen, which states that anxiety lowers language learning. Independent learning with the use of AI tools enabled learners to have fewer affective barriers as they do not fear being judged.

On the same note, the increased engagement and motivation mentioned by learners are consistent with modern concepts of learner-centered and technology-enhanced learning. AI applications made learning an active and ongoing experience by being interactive and feedback-driven. This predisposes the argument that digital tools are capable of producing intrinsic motivation and practice.

A high degree of technology acceptance is also reflected by the positive attitude of AI tools by learners. Following the Technology Acceptance Model (Davis, 1989), the learners found that these tools were useful and easy to operate and this led to their high adoption and efficacy.

Nevertheless, some of the challenges that were identified in the study are internet dependency and sometimes inaccuracy in the recognition of accents. These shortcomings demonstrate the technological deficiency of AI systems at the moment, and underscore the importance of further improvement. Also, they show that AI tools may not be entirely substitutive of human instructors but must be incorporated as support materials as part of a blended learning approach. In general, this study results align with existing literature that indicates that AI-based pronunciation systems can greatly improve the results of language learning and allow learners to gain independence and interactivity.

CONCLUSION

This research aimed at investigating the usefulness of AI-enabled speech recognition systems in enhancing pronunciation of English among ESL students. The findings clearly show that the tools play an important positive role in the pronunciation skills of the learners. The quantitative assessment showed significant gains in pronunciation accuracy, fluency, and intelligibility, and statistically significant differences between pre-test and post-test results. These results affirm that learning aided by AI can be efficient in improving the major factors of spoken language performance.

The qualitative analysis also identified that learners were more confident, more motivated, and with a positive attitude to AI-based learning. The possibility to get instant feedback and practice independently helped to make the learning more personal and flexible. Irrespective of some constraints (such as technological constraints and inaccuracies in feedback in some cases), the overall efficiency of the AI tools can be seen. These resources offer an appreciable platform to practice continuously and work on oneself, especially in situations where one may be short of access to native speakers or professional training. Speech recognition technologies using AI are very effective in enhancing pronunciation in English among ESLs. They are potent complementary tools that can complement the conventional teaching methods by encouraging the learner autonomy, interest, and **confidence.**

RECOMMENDATIONS

The use of AI-based pronunciation tools in the classroom should become an integral part of the classroom instruction to supplement traditional methods of teaching. Teachers are advised to embrace blended learning strategies that involve integrating the power of AI technology with human instruction to improve learning activities. Moreover, educators ought to offer adequate training and orientation to children and allow them to apply AI tools effectively and to make the automated feedback meaningful.

Students are to actively use AI-based tools to train pronunciation to gain more accuracy, fluency, and speaking skills. They are supposed to avail themselves of the feedbacks to recognize their errors and practice self-correction. Regular use of these tools will build self-directed learning practices and help in the development of languages in the long-run.

Schools are important in helping implement AI-based learning. They must invest in proper digital infrastructure such as availability of good internet and proper equipment so that all students can access the advantage of such technologies. Also, AI devices need to be introduced into language education and educational programs in a formal manner to encourage systematic and organized use. On the technological side, developers are advised to work on the accuracy of the accent recognition systems to suit the non-native speakers. More detailed, context-sensitive, and pedagogically meaningful feedback should be provided using AI applications as well to improve the effectiveness of learning. Moreover, offline capabilities need to be developed to deal with the connectivity issues, especially in areas that have lower technological capabilities.

Lastly, more studies should be conducted to understand the long-term effects of AI-based pronunciation tools on language learning outcomes. Comparative studies of the two teaching techniques AI-assisted and traditional should be conducted in the future to elucidate relative effectiveness. Additionally, it is advisable to carry out studies with bigger and more diverse samples to enhance the extrapolation and generalizability of results in other educational settings.

REFERENCES

Ahmed, S. (2018). Challenges in teaching pronunciation in ESL classrooms in Pakistan. *Journal of Language Teaching and Research, 9*(3), 567–574.

Akhtar, M. (2020). Learners' attitudes toward technology-assisted language learning in Pakistan. *Pakistan Journal of Education, 37*(2), 45–60.

Bashir, M. (2019). Classroom challenges in ESL learning: A study of overcrowded classrooms. *International Journal of Educational Research, 5*(1), 23–30.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Campbell, N. (2008). Speech technology for language learning. *ReCALL, 20*(3), 365–372.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*(3), 379–397.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE Publications.

Heift, T., & Schulze, M. (2007). *Errors and intelligence in computer-assisted language learning: Parsers and pedagogues*. Routledge.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Hussain, I. (2019). Mobile-assisted language learning and its impact on ESL learners. *Journal of Educational Technology, 16*(1), 89–102.

Iqbal, A. (2017). Digital tools and student engagement in ESL classrooms. *Journal of Language and Education, 3*(2), 34–42.

Khalid, R. (2021). Impact of speech recognition tools on ESL pronunciation skills. *International Journal of Applied Linguistics, 12*(1), 78–90.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.

Litman, D., & Rosé, C. (2006). Spoken dialogue systems for language learning. *Computer Speech & Language, 20*(4), 393–410.

Ma, W., & Rosé, C. P. (2010). Intelligent tutoring systems for language learning. *Language Learning & Technology, 14*(2), 32–45.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

Nazir, M., Ali, S., & Khan, R. (2025). Effectiveness of AI-based pronunciation tools among Pakistani ESL learners. *Journal of Applied Linguistics and Language Research, 12*(1), 101–115.

Prakash, R., & Kausalya, S. (2025). AI-powered speech recognition in ESL learning: An experimental study. *International Journal of Educational Technology, 15*(2), 55–70.

Rafiq, M. (2020). Technology integration in ESL classrooms: A mixed-method study. *Pakistan Journal of Educational Research, 3*(1), 112–125.

Raza, S. (2021). Artificial intelligence in education: Opportunities and challenges. *Journal of Educational Innovation, 6*(2), 15–28.

Reinders, H., & Hubbard, P. (2013). CALL and learner autonomy: Affordances and constraints. *Computer Assisted Language Learning, 26*(1), 1–15.

Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.

Suendermann, D. (2012). Advances in automatic speech recognition systems. *Springer*.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and output. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching, 31*(2), 57–71.