

## The Effectiveness of AI Speech Recognition on Students' English Pronunciation

**Woro Kusmaryani<sup>1</sup>, Ramli<sup>2</sup>, Winarno<sup>3</sup>**  
 Universitas Borneo Tarakan, Indonesia  
 Email Correspondence: worokusmaryani@borneo.ac.id

### **Background:**

Pronunciation is probably one of the most problematic and the conventional teaching approaches do not tend to offer the immediate, personal feedback, which is required to achieve effective pronunciation. But AI speech recognition applications can provide a real-time corrective response that may be used to overcome these drawbacks in teaching.

### **Methodology:**

The research utilized a pre experimental one group pre-test post-test study design whose qualitative observations were supported by 30 first-semester students purposely selected to participate in the study. The information was gathered by means of pre- and post-teaching tests, observation forms, and AI-based feedback. The quantitative analysis was based on descriptive statistics and paired sample t-test, and instead, thematic coding based on observation was applied to qualitative data.

### **Findings:**

The results showed that the pronunciation of students has been significantly improved, and the average scores have increased by 68.4 to 81.7 ( $p < 0.05$ ). The qualitative observations revealed the enhancement of the accuracy in the production of segmental characteristics, including the interdental sounds (/θ/, /ð/), voicing contrasts (/v/ vs. /f/), and vowel length differences, and advances in supra segmental characteristics, which included word stress, rhythm, and intonation. Besides, self-confidence, motivation, and independence of the students in practicing pronunciation using AI-supported learning also improved.

### **Conclusion:**

AI speech recognition technology is a priceless aid in enhancing the English pronunciation. Feedback that was given to the learners was both regular, individualized, and timely and this resulted to a rise in accuracy and self-regulated learning.

### **Originality:**

The study provides classroom-based evidence concerning the application of AI speech recognition in learning English pronunciation. It shows the possibilities of using AI in teaching pronunciation and increasing the level of learner autonomy and positive results of learning.

<b>Keywords</b>	: AI Speech Recognition; Effectiveness; English Pronunciation
<b>DOI</b>	: 10.24903/sj.v11i1.2282
<b>Received</b>	: December 2025
<b>Accepted</b>	: April 2026
<b>Published</b>	: April 2026
<b>How to cite this article (APA)</b>	: Kusmaryani, W., Ramli., & Winarno. (2026). The Efferctiveness of AI Speech Recognition on Students' English Pronunciation. <i>Script Journal: Journal of Linguistic and English Teaching</i> , 11(1), 77-98. <a href="https://doi.org/10.24903/sj.v11i1.2282">https://doi.org/10.24903/sj.v11i1.2282</a>
<b>Copyright Notice</b>	: <b>Researchers retain copyright and grant the journal right of first publication</b> with the work simultaneously licensed under a <b>Creative Commons Attribution 4.0 International License</b> that allows others to share the work with an acknowledgement of the work's researchers ship and initial publication in this journal.



## 1. INTRODUCTION

The role of pronunciation in second language acquisition (SLA), especially in English as a second language (ESL), is essential (Aryanti & Santosa, 2024). The ability to pronounce words correctly, or rather to pronounce them in a way that they are intelligible to both native and non-familiar listeners, is often the fourth criterion of speech quality (Derwing & Munro, 2005). In the current scenario, where global communication is mainly through speakers from different linguistic backgrounds, getting the correct pronunciation is indeed a key factor for successful communication. However, it is through pronunciation classes that the majority of learners acquire the least, with their training primarily focusing on grammar and vocabulary (Khurshedjonovna, 2025; Jenkins, 2000). Additionally, conventional pronunciation training methods, e.g., teacher feedback and practicing with classmates, may not always provide students with the kind of consistent, instant, and personalized feedback they need. This update in pronunciation teaching methods has led to a higher demand for tech-assisted learning tools, particularly those that use AI to support English pronunciation improvement (Junining et al., 2020; Levis, 2018).

The AI technology-based speech recognition tools have completely transformed the process of learning languages by giving each student instant feedback that is customized to them (Miladiyenti et al., 2022; Saeed, 2020). The way these systems of technology work is by applying very advanced algorithms to the spoken language data and then going through the whole process of transcription into text, followed by evaluation of many factors of pronunciation, which include segmental sounds (consonants and vowels) and supra segmental features (stress, rhythm, and intonation) (Dutta & Arora, 2021). AI-assisted applications such as ELSA Speak, Google Speech-to-Text, and SpeechAce are now essential in the language-learning process, as they identify learners' pronunciation mistakes and instantly provide correction suggestions (Saeed & Sharma, 2020). The self-learning trend driven by AI tools aligns with modern theories of second language acquisition, which emphasize learners' independence and personalized practice as essential factors in overcoming challenges and, therefore, making language learning successful (Benson, 2011).

The segmental and supra segmental aspects of pronunciation have been recognized as significant areas in which AI speech recognition tools are practical, and numerous studies have supported this. For example, Dutta and Arora (2021) investigated AI in segmental pronunciation. They reported marked improvement in learners' production of the most difficult English sounds, particularly the most problematic consonants for ESL learners. In addition, the

tools based on AI not only give feedback on mispronunciations instantly but also guide the students in rectifying their errors as they occur, and thus make it possible for the students to practice alone even outside the classroom (Levis, 2018; Sun, 2024). Moreover, research such as that by Saeed and Sharma (2020), Abimanto and Sumarsono (2024) and Dennis (2024) emphasizes that the use of AI technologies offers students the opportunity to enhance their phonological awareness, which is a basic requirement for accurately perceiving and producing correct pronunciation.

On the other hand, although AI technology holds great promise, it still has drawbacks. Research has shown that AI speech recognition has not yet reached its full potential because it may still struggle with regions, different speaking styles, and the conversion of words into sound, which can ultimately lead to incorrect feedback (Indari, 2023; Sun, 2023; Vassallo et al., 2019). Furthermore, researchers have found that the extent of the positive impact of AI applications varies according to the learner's mother tongue, educational phase, and the specific AI system in question (Dja'far & Hamidah, 2023; Dutta & Arora, 2021; Vančová, 2024). Therefore, the use of artificial intelligence in teaching pronunciation will require careful attention to these aspects (Noviyanti, 2022).

At first, the main emphasis of pronunciation teaching was provided to the segmental ones, however, the supra segmental ones (intonation, stress, rhythm) have been receiving more and more attention since then. Such characteristics give naturalness and intonation to speech, thus influencing comprehensibility and fluency (Cutler, 2018). The results of AI tools based on supra segmental-level feedback have contributed more to improving intonation and the rhythm of speech, which is required to sound less foreign when a person speaks a second language (Derwing & Munro, 2009). According to Levis (2018) it is frequently more complex to control supra segmental than segmental, and AI tools can be a useful addition to conventional classroom teaching in this respect.

A key factor in this analysis is the learner's opinion on the use of AI tools for pronunciation practice (Sardegna & McGregor, 2022). It is not only for the sake of the future of language education to determine the extent to which students accept AI feedback as more or less effective than traditional methods, but also for the quality of the education provided. Studies by Saeed and Sharma (2020) and Saeed (2020) show that learner involvement can be a deciding factor in the success of AI-assisted pronunciation practice. If students enjoy the AI engagement and find it personalized, their pronunciation will get better. On the other hand,

students' acceptance of AI is influenced by factors such as their tech-savviness, perceived user-friendliness, and the reliability of the AI's feedback (Vassallo et al., 2019).

Moreover, although AI tools can provide helpful feedback, some research still holds that teachers' or peers' feedback is the most important (in this case, the teacher or peer giving feedback) because such feedback offers contextual understanding and sociolinguistic sensitivity that AI lacks (Derwing & Munro, 2005). Hence, the present research will compare the quality and user experience of AI-generated and teacher-driven feedback to determine which is more advantageous for pronunciation enhancement.

On the one hand, the efficiency of AI tools has been studied in various fields of language learning; on the other hand, none of the studies investigated the role of AI tools, especially in segmental and supra segmental pronunciation in the ESL setting. Also, much of the debate surrounding these studies lies in the fact that some of those researchers, rightly or wrongly, base their argument on the technology aspect whereas those researchers, who usually overlook this aspect, base their argument on the educational aspect and are primarily concerned with how the students feel about the technology or how effective the technology will be in the long run. This study will help resolve the issue by looking at the effect of the AI tool on the pronunciation improvement of students as done by pre-test and post-test tests, and subjective perceptions of the AI tool by learners. In this manner, the study will contribute to the better understanding of practical and pedagogical implications of AI in language learning.

Learners often have difficulties with English pronunciation, and the traditional methods might lack enough customization and feedback, particularly in the case of non-native students. It is important to understand whether AI speech recognition can possibly close the gap and result in successful learning experiences. The premise is that AI speech recognition will not only give the students correct and immediate feedback but also customized practice, which will eventually result in quantifiable recognition of their pronunciation and fluency progress. In case AI speech recognition proves to be very useful, first, the rules of its integration into the mainstream language-learning systems will be prepared, and second, the instructions of the use of AI speech recognition in the teaching plans of teachers will be provided. After the research is conducted, the AI tools will be enhanced in accordance with the academic results, and additional enhancements will be offered so that the pronunciation advantages can be maintained in the long-term perspective.

In these days, the traditional methods are still the heart of the English pronunciation training and development, classroom training, language laboratory, and pre-recorded audio

material. Even though such techniques can be useful, they do not always provide personal feedback and real-time correction. Other more sophisticated applications, such as pronunciation apps or language-learning software (e.g., Duolingo, Rosetta Stone), are based on simple speech recognition, but focus more on vocabulary and grammar than on individual, nuanced pronunciation problems. The speech recognition technology based on AI has gained traction in linguistics, language learning, and speech therapy. The effective creation of Google Assistant, Siri created by Apple and voice-activated transcription services has enhanced speech recognition much, yet their opportunities to be used in the process of language learning are still not fully utilized. The current research on AI speech recognition development is primarily centered on the accuracy of speech-to-text. Meanwhile, very few studies were focused on its use in educational purposes, in particular, training the pronunciation of students.

The current study is the pioneer that employs a neural network-based speech recognition system as a specific aid in the English pronunciation of non-native speakers in a classroom environment. The given scheme allows not only obtaining a phonetic transcription of the speech of a learner but also real-time feedback on the characteristics of phonemes voicing and length. It also highlights the main differences between traditional and AI-based feedback methods. In addition, the AI-based solution enables students to rehearse difficult sounds more frequently and more extensively. Another variable that makes the approach new is the variety of educational situations involved in this study, as well as the possibility to provide the student with the individual learning experience, based on the particular pronunciation issues faced by that student. This study can result in the development of tailor-made and scaled AI speech recognition applications that would seamlessly fit into a language learning platform, thus enhancing access and services to students in all parts of the world. This research gives fresh insights into the way of making the AI useful not only as a speech recognition tool but also as an active partner in personalized pronunciation training, therefore, offering the state-of-the-art language-learning solution combining AI with the most effective teaching techniques.

Even though the possible effect of AI-based tools in pronunciation practice has been reported in the previous research, there is still a lack of evidence in classrooms, especially in terms of enhancement of both segmental and supra segmental aspects in the real-life EFL settings. Moreover, not many studies have matched performance results with the results of classroom practice that can be observed. Thus, this research analysed the pronunciation increase in a natural classroom environment with the help of pre-test and post-test scores with

the help of the observational data. The paper will add classroom-based research evidence regarding the application of AI speech recognition in learning the English pronunciation.

## 2. METHODOLOGY

The research design used in the study was a pre-experimental one-group pretest-posttest design supported by qualitative observation (Creswell, 2014). The respondents were 30 purposely chosen first-semester students. The participants were told about the intention and process of the study and were involved in it with consent. Ethics were considered during the process of conducting research. These tools were pronunciation pre-test and post-test, classroom observation sheets, and AI-created pronunciation feedback of the learning applications. The intervention was a six week intervention. Ethical issues were considered through the process of explaining the study procedures to the participants and getting their consent before joining the study. The pre-test and post-test scores were obtained through human assessment by applying the pronunciation rubric, and the AI-generated scores were generated through practice sessions and were utilized as a descriptive measure of the learner progress only. Therefore, the human-rated scores served as the basis of the quantitative analysis, whereas the classroom observations and AI-feedback served to facilitate the interpretation of the improvement patterns. The performance of students in pronunciation was measured in three areas: pronunciation accuracy, stress, and intonation, and the results were given on a 0100 scale. To explain the general level of pronunciation, performance categories were used to interpret the ranges of scores. Human test scores are the scores received on the pre-test and post-test speaking assessment, and AI-generated scores are the scores received on the app-based feedback received during the practice sessions. The latter were to be employed descriptively to depict the learning progress, but not as the major dependent measure.

Three primary tools were utilized to gather the data, namely, pre-test and post-test evaluation of pronunciation accuracy, stress, and intonation before and after the intervention; observation sheet to note learning behaviours, frequent mistake in pronunciation, and engagement during AI-assisted practice among the students. It was conducted using a series of data collection procedures, which involve pre-test administration, pronunciation training using applications like ELSA Speak and Google Speech-to-Text within a period of six weeks and monitoring of the performance of the students throughout the treatment process and finally post-test administration with the same grading criteria. Quantitative data were analysed using descriptive statistics and paired-samples t-test to determine whether improvement in scores was

significant and qualitative data were obtained through observations and data were reduced, displayed, and conclusions were drawn to identify pattern of improvements. The two data sets were later merged in the interpretation stage to validate and enhance the results hence rendering the conclusion on the impact of AI Speech Recognition on pronunciation in students holistic and substantial.

### **3. FINDINGS**

#### **3.1 The Treatment Process of Using AI Speech Recognition in Pronunciation Learning**

The implementation of AI Speech Recognition in teaching pronunciation to first-semester English Education Department students of Universitas Borneo Tarakan was conducted in five sequential steps, i.e. preparation, guided practice, independent training, feedback, and reflection. The preparation phase allowed the students to get acquainted with AI-based applications (ELSA Speak and Google Speech-to-Text). The lecturer showed how one could record a voice, check the accuracy scores, and identify phonetic errors with the help of the visual feedback offered by the apps. Moreover, students were provided with the checklist of the most frequently mispronounced pronunciations including /θ/, /ð/, /v/, /f/, /ɪ/, and /i:/.

In the guided practice, the lecturer provided short pronunciation tasks on sample words and phrases, including “think,” “that,” “very,” “live,” and “leave.” The AI system examined the clarity of phonemes and accuracy and offered real-time feedback and indicated the areas where pronunciation was not in accordance with native rules. Students had seen the insignificance of mouth movements in the score of AI. The students received the independent practice stage during which they underwent home-based 10-15 minutes a day AI pronunciation training. The applications provided visualization of a text and audio feedback. Most students claimed that with repeated practice, they were able to correct themselves without necessarily involving the lecturer.

For example, a particular student first interpreted “think” as /tɪŋk/, but after constant practice successfully changed it to /θɪŋk/ with the correct dental fricative /θ/. Another boy said “very” as /peri/ at the start because the Indonesian sound /p/ was on his mind, but he eventually used AI feedback to correctly pronounce the voiced /v/as in /'ver.i/. In the feedback and reflection stage, the professor reviewed students' pronunciation reports generated by the AI tool. The professor pointed out the troublesome words and demonstrated the correct pronunciation in class. The students indicated that AI technology was a significant factor in raising awareness of their individual pronunciation weaknesses, as they had rarely noticed such things before using the app. Furthermore, based on the observations made, it was concluded

that the students gained remarkable improvement in the production of vowel and consonant sounds (individual sounds) and word stress correctness after six weeks of regular exposure to AI Speech Recognition.

a. Enhancement of Individual Sounds (Segmental Features)

Interdental sounds (/θ/ and /ð/): /t/ was produced in place of the sound /θ/. Participants could not pronounce /θ/ in words like think and thought at first. Following AI-aided training, almost all the students (23 out of 30) reached the standard pronunciation of the sound.

Example:

**Before AI Training**

*“I t’ink it is good.”*

**After AI Training**

*“I think it is good.”*

Voiced–voiceless contrast (/v/ vs. /f/): The learners at first mixed up the sound’s /v/ and /f/ in unaccented words like very/ferry, live/life. But later, a great deal of the subjects could correctly tell the two apart:

**Before AI Training**

*“I want to have a good li[f]e here.”*

**After AI Training**

*“I want to have a good li[v]e here.”*

Vowel distinction (/æ/ versus /ɛ:/): The variety improved greatly in the clarity of their vowel distinctions. This is an example:

**Before AI Training**

*“He is si[t]ting in the liv[i:]ng room.”*

**After AI Training**

*“He is sitting in the living room.”* (corrected  
/ɪ/ and /i:/)

b. Increase in Awareness of Stress

Students seemed to gain a heightened awareness of syllable stressed patterns. For example:

**Before AI Training**

*“comPUter”* → stress misplaced on  
second syllable

**After AI Training**

*“COMputer”* → correct primary stress  
on the first syllable

c. Enhancement in Speaking Skill

The students showed quite a difference in their production of sentences after the training by having a more natural rhythm and better articulation. The following are some instances:

**Before AI Training**

*"I tink English is ferry difikult."  
"Dis movie is berry fun."  
"My life is fill of hapiness."*

**After AI Training**

*"I think English is very difficult."  
"This movie is very fun."  
"My life is full of happiness."*

The ability of students to correctly stress the intonation of emotional sentences was confirmed through post-training interviews and AI practice recordings. The AI's average pronunciation score for the group rose from 68% (moderate accuracy) to 87% (good accuracy), thus showing significant development, nonetheless these numbers were only descriptively used to back up qualitative interpretation.

### 3.2 The Effectiveness of AI Speech Recognition on Students' Pronunciation

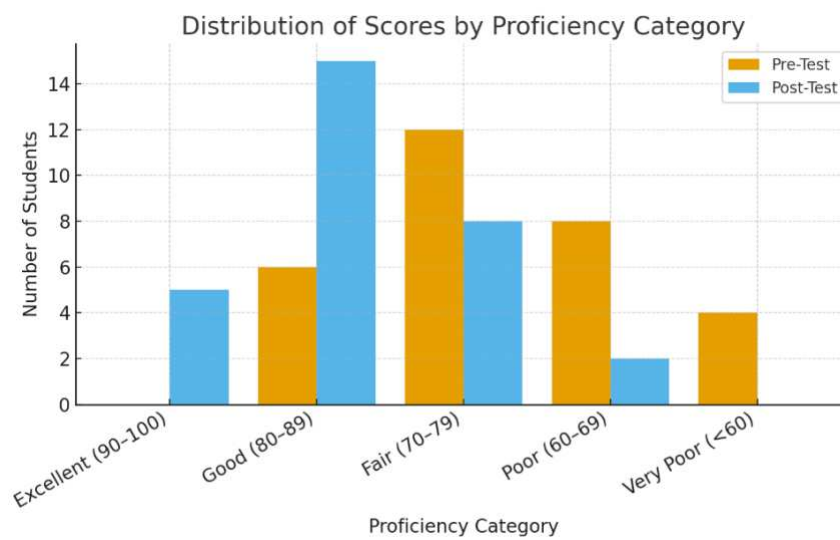
First, the pronunciation skill of students was assessed through the administration of a pre-test, which consisted of a reading passage and a word pronunciation task. The pronunciation performance of the students was evaluated on three levels: accuracy, stress, and intonation, and the scores were given in a 0100 scale. To describe the level of pronunciation in general, the performance categories were used to interpret the score ranges. The human test scores are defined as the results of the pre-test and post-test speaking test, and the AI-generated scores are the results of the app-based feedback that is shown during the practice sessions. The latter was applied descriptively as the illustration of the progress in learning, rather than serving as the dependent measure. The pre-test results showed that the mean score was 68.4, the maximum score was 78 and the minimum score was 58. Most of the students, which is 60%, fell into the fair category and their errors included sounds /θ/, /ð/, /v/ and diphthongs. They also showed inconsistent stress and intonation patterns. The post-test mean score after four weeks of pronunciation practice with AI speech recognition tools (such as ELSA Speak and Google Speech-to-Text) was 81.7. The highest score was 90, while the lowest was 72. The 13.3 points gained overall meant a significant improvement in students' pronunciation skills. Moreover, the notes taken during observation revealed the presence of higher confidence and more consistent pronunciation patterns after the practice with AI.

Table 1 Score Distribution

Score Distribution of Pre-Test and Post-Test

Score Range	Proficiency Category	Pre-Test (f)	Post-Test (f)
90 – 100	Excellent	0	5
80 – 89	Good	6	15
70 – 79	Fair	12	8
60 – 69	Poor	8	2
< 60	Very Poor	4	0
<b>Total</b>		<b>30</b>	<b>30</b>

Chart 1 Distribution of score by Proficiency Category



In the pre-test, the majority of students (12 students) were in the Fair category, followed by 8 students (26.7%), 6 students (20.0%), and 4 students (13.3%) in the Poor category. No students achieved the Excellent category. In the post-test, the score distribution shifted to the higher categories. The majority of students (15 students) were in the Good category, followed by 8 students (26.7%), 5 students (16.7%), and 2 students (6.7%) in the Poor category. There were no longer any students in the Very Poor category. The distribution of scores indicated an improvement in learning outcomes after the treatment. In the post-test, the number of students in the Good and Excellent categories increased, while the number of students in the Poor and Very Poor categories decreased.

Table 2 Descriptive Statistics

Variable	N	Mean	Std. Deviation	Std. Error Mean
Pre-test	30	68.40	7.238	1.322
Post-test	30	81.70	6.109	1.115

The average student score increased from 68.40 in the pre-test to 81.70 in the post-test. Descriptively, there was an average increase of 13.30 points.

Table 3 Tests of Normality (Shapiro-Wilk)

Data	Statistic	df	Sig.
Pre-test	0.847	30	0.001
Post-test	0.954	30	0.218
Difference (Post - Pre)	0.962	30	0.353

Decision:

- Pre-test: Sig. = 0.001 < 0.05 → not normal
- Post-test: Sig. = 0.218 > 0.05 → normal
- Difference score: Sig. = 0.353 > 0.05 → normal

Although the pre-test distribution was not normal, the difference scores were normally distributed. Because the paired t-test assumed normality of the difference scores, the paired t-test was still appropriate.

Table 4 Test of Homogeneity of Variance (Levene's Test)

Levene Statistic	df1	df2	Sig.
1.421	1	58	0.238

Decision:

Sig. = 0.238 > 0.05, so the variant data was homogeny.

Table 5 Paired Samples Statistics

Pair	Mean	N	Std. Deviation	Std. Error Mean
Pre-test	68.40	30	7.238	1.322
Post-test	81.70	30	6.109	1.115

Table 6 Paired Samples Correlation

Pair	N	Correlation	Sig.
Pre-test & Post-test	30	0.918	0.000

Table 7 Paired Samples Test

Pair	Mean Difference	Std. Deviation	Std. Error Mean	95% CI Lower	95% CI Upper	t	df	Sig. (2-tailed)
Post-test - Pre-test	13.300	2.926	0.534	12.207	14.393	24.896	29	0.000

Decision:

Significant value (2-tailed) =  $0.000 < 0.05$ , therefore,  $H_0$  was rejected and  $H_a$  was accepted.

There was a significant difference between the pre-test and post-test scores. Thus, AI Speech Recognition effectively improved students' English pronunciation.

### N Gain Test

Formula:

$$N \text{ Gain} = \frac{\text{Post-test} - \text{Pre-test}}{100 - \text{Pre-test}}$$

Average calculation results:

Mean Pre-test = 68.40

Mean Post-test = 81.70

Mean N-Gain = 0.432

N-Gain (%) = 43.15%

Table 8 N Gain Category

Range	Category
$g < 0.30$	Low
$0.30 \leq g < 0.70$	Moderate
$g \geq 0.70$	High

The average N-Gain value was 0.432, making it fall into the moderate category.

This means that the use of AI Speech Recognition provided a moderate level of effectiveness in improving students' pronunciation skills. The analysis results showed that the average student score increased from 68.40 in the pre-test to 81.70 in the post-test. The Shapiro-Wilk normality test on the difference score showed a significance value of  $0.353 > 0.05$ , so the data met the normality assumption for the paired t-test. Levene's homogeneity test also showed a significance value of  $0.238 > 0.05$ , which means the data variance was homogeneous. Furthermore, the paired t-test results showed a value of  $t = 24.896$  with Sig. (2-tailed) =  $0.000 < 0.05$ , so there was a significant difference between the pre-test and post-test scores. Thus, the use of AI Speech Recognition has proven effective in improving students' English pronunciation skills. In addition, the N-Gain result of 0.432 or 43.15% was in the moderate category, which indicated that the improvement in learning outcomes was at a moderate level of effectiveness.

The findings suggest that the use of AI speech recognition was associated with improvement in students' pronunciation performance during the intervention period. Given the one-group design, the results should be interpreted as indicating improvement following the intervention rather than proving causal effectiveness. The main quantitative outcome was the change in human-rated pronunciation scores from pre-test to post-test. Observational findings were then used to explain patterns of improvement in segmental and supra segmental features, while AI-generated scores were presented only as descriptive practice feedback.

#### **4. DISCUSSION**

The use of AI Speech Recognition (AISR) in pronunciation training for first-semester students in the English Education Department at Universitas Borneo Tarakan was a well-planned, interactive process comprising five stages: preparation, guided practice, independent practice, feedback, and reflection. The technology-enhanced process followed a cycle in harmony with the Computer-Assisted Pronunciation Training (CAPT) and Technology-Enhanced Language Learning (TELL) frameworks (Derwing & Munro, 2015; Zainuddin, 2024). In the first phase of the research, students were introduced to the AI tools (ELSA Speak and Google Speech-to-Text) in a preparatory way. The teacher's discussion about pronunciation properties, phonetic symbols, and how to interpret AI feedback provided support for the learners. This phase aligns with the "modelling and awareness-raising" stage of (Derwing & Munro, 2015) in which students are guided to specific pronunciation targets and then to independent production. The first explicit pronunciation instruction is essential for learners to identify articulatory positions and sound contrasts that differ from those of the native language, according to (Kenworthy, 1987). Likewise, Thomson and Derwing (2015) suggested that pre-instruction modelling could help learners avoid fossilizing errors. In this research, such scaffolding provided students with the basic knowledge essential for a proper understanding of AI feedback and its practical application during practice.

The professor utilized the AISR in the classroom during the guided practice stage, where he gave words and short phrases to the pronunciation exercises. The AI application gave live feedback, emphasizing the mispronounced sounds whereas the lecturer gave the corrections and illustrations. This scenario conforms to the notions of communicative pronunciation instruction (Celce-Murcia et al., 2010), that believe in the combination of explicit teaching with significant practice. The presence of both AI feedback and lecturer support helped to make the students more inclined to noticing and monitoring themselves, which are essential requirements of the pronunciation development. Research has found that

guided practice with AI produces increased awareness in learners and increased engagement. Indicatively, [Nguyen et al. \(2025\)](#) administered a survey where ASR trained drills resulted in better articulation and motivation of the learners due to instant and personalized feedback loop. In line with this, [Kholis \(2021\)](#) found that guided pronunciation training using ELSA Speak led to higher accuracy in students on complex sounds of the English language, including /θ/ and /ð/, that were also the same sounds that were better in this study.

The independent learning stage, during which students practiced daily at home with AI tools, marked a transition from teacher-led instruction to independent, self-regulated learning. The students used the app multiple times to improve their pronunciation scores, reflect on mistakes, and track their progress. This stage aligns with the autonomous learning model proposed by [Benson \(2013\)](#), which emphasizes learner control over the pace, content, and evaluation. The possibility of practicing without teacher supervision so often aligns with self-regulated learning theory, in which students plan, monitor, and evaluate their learning activities ([Zimmerman, 2002](#)). Similarly, empirical studies show equivalent results. [Clarke-Jones \(2021\)](#) found that students who used CAPT applications developed greater autonomy and self-confidence, while [Zainuddin \(2024\)](#) noted that mobile-based language tools enhance student engagement and self-directed learning. In light of this research, students' constant practice with AI tools can be seen as a source of feeling responsible for their pronunciation progress, which aligns with the mentioned studies.

The final stage of the procedure was feedback and reflection, during which, together with the lecturer's comments, the AI-generated reports served as the basis for evaluating progress. The students were performing their self-analysis, identifying difficulties they encountered, and comparing the results obtained pre- and post-practice. This phase is a direct manifestation of [Schmidt \(1990\)](#) noticing hypothesis, which postulates that taking notice of and being aware of the feedback provided in language is an essential process in language acquisition. The triangulation of AI feedback (visual, auditory, and numerical) and lecturer feedback (providing qualitative explanation) constituted a multimodal way of noticing. Using a pronunciation pedagogy approach, researchers found that reflection was an integral part of learning improvement. According to [Derwing and Munro \(2015\)](#), reflective practice after pronunciation training fosters learners' ability to take feedback and make the necessary changes in their pronunciation habits. [Hidayatullah et al. \(2025\)](#) also found that students who learned pronunciation with ASR developed enhanced metacognition, became more aware, and were

able to distinguish between their strengths and weaknesses in learning pronunciation using AI-generated performance analytics.

The general implementation in this study denotes collaboration between the feedback of AI and the facilitation of the lecturer. The AI software offered instant, factual corrections, whereas the teacher was able to confirm the cognition of the learner and provide emotional encouragement. This is an integrated process that aligns with the Hybrid CALL pedagogy, which implies the combination of automated feedback and human mediation to reach the balanced pronunciation competencies (Nursyafida & Putri, 2025; Hsu, 2023). Wen and Li (2023) emphasized that technology cannot be the only developer of communication based on pronunciation, modelling and teacher feedback are necessary to cover supra-segmental characteristics like rhythm and intonation. Therefore, the contribution of the instructor in the interpretation of AI outcomes and providing tailored instruction was a significant factor when it comes to overall pronunciation enhancement in the students.

The purpose of the research was to measure the efficacy of an AI speech recognition intervention in mastering English pronunciation with EFL learners, which was conducted with the help of a one-group pre-test-post-test (pre-experimental) research design. The outcomes not only indicated a mean pre-test score of 68.4 but also a mean post-test score of 81.7 which showed an improvement of pronunciation performance which was statistically and practically significant after the intervention. The findings in the current discussion have been interpreted in the light of available theory and research, implications on practice discussed, limitations addressed, and future research directions recommended. The massive difference between the average scores of 68.4 and 81.7 is not only an indication but also a reflection of a greater pronunciation performance, which, most likely, came about due to the immediate and personal feedback that is a salient feature of AI speech recognition systems. According to meta-analytic evidence, Automatic Speech Recognition (ASR) tools in ESL/EFL pronunciation training have yielded a medium overall effect size of  $g = 0.69$ , and their impact on the segmental (vowel/consonant) features is significantly greater when compared with that on the supra-segmental (intonation/stress) ones, especially when explicit corrective feedback is given. The size of the absolute change in this research indicates that the AI tool was instrumental in the development of awareness among learners of mispronunciation and later correction of articulatory characteristics. The current consistency with the previous studies substantiates the assertion that the intervention facilitated the phonological accuracy of the learners.

The implication for teaching is that AI-based feedback could bolster the traditional teacher-led pronunciation instruction by giving students more opportunities to be corrected individually and frequently. [Neri et al. \(2008\)](#) stress the importance of personalized feedback, since the teacher cannot provide equal input to each student in class due to time constraints. Moreover, while progress is evident, researchers have warned that developments in the supra-segmental area will always lag those in the segmental area, meaning that in subsequent iterations of the project, stress and intonation might have to be the focus of the intervention to enable further improvements. As scores rose, qualitative observations during the research indicated that learners' confidence was boosted and that the practice they did alone increased, consistent with the notion of learner autonomy. Research shows that when learners engage in pronunciation practice through online resources or computer-assisted language learning (CALL) tools, their autonomy and self-regulation are enhanced, and these, in turn, lead to better pronunciation outcomes. For instance, [Kruk \(2012\)](#) concluded that online resources were instrumental in developing learners' autonomy, and that these resources led to positive pronunciation gains.

Likewise, [Clarke-Jones \(2021\)](#) found that pronunciation instruction strategies that promote learner autonomy were more effective at sustaining learners' participation. In this research, the AI tool plays a vital role in establishing a more autonomous practice setting where students can perform pronunciation practice outside the classroom, receive instant feedback, and thus feel more in charge of their learning path. It is a direct consequence of instructional design: letting learners work alone, at their own pace, with instant feedback might lead to more time spent on practice, as reflected in pronunciation mistakes. It might be worthwhile for teachers to arrange pronunciation modules so that they gradually withdraw from providing AI tools, thus merging the running of the class with independent, tool-supported practice. The literature reveals that the two factors which have the most significant influence on the outcome of the ASR-based interventions are the period of treatment and the adherence to the planned course of treatment. According to the review by [Ngo et al. \(2023\)](#), moderate and long-term practice times with ASR yield the greatest learning and short-term interventions may not have different effects than non-AASR conditions. The key positive feedbacks in this study were that the number of sessions and duration was well outlined, which is an indication that the intervention offered a dose adequate to initiate change. However, additional investigation is required with indicators of usage (i.e., number of sessions, amount of practice, frequency, etc.) to be more aligned with best practice.

The other important aspect is the category of errors, i.e. segmental and supra-segmental, i.e. fricatives and stop consonants. Such studies as [Spring and Tabuchi \(2022\)](#) discovered that ASR training made more progress in vowel problems compared to supra segmental problems in learners. The knowledge of the most enhanced error types in your study would promote a better understanding of the functionality of the AI tool. As an illustration, in case the performance of the students were higher in segmentals, it would be in line with the available literature. On the other hand, in case the supra segmental gains were also meaningful, it would mean that the tool was particularly adapted to those features.

The findings of this study have certain teaching implications. Firstly, the teaching of pronunciation should take advantage of the different technological tools in order to complement the traditional one. AI speech recognition systems offer personalized feedback that is reliable and stable, which is why it is simpler to correct pronunciation mistakes. The researchers [Martin \(2023\)](#) and [Nasir et al. \(2023\)](#), state that computer-assisted pronunciation training (CAPT) enhances the motivation of learners and gives them more autonomy. Next, teachers are to combine practice with AI tools in the following way: teacher-mediated input (modelling, phoneme drills), AI-mediated independent practice, and peer interaction (which meta-analysis may help to increase the effect size). Then, it is necessary to train students to understand and respond to AI feedback. It is not sufficient that learners receive feedback, they should interpret the corrective information (e.g., what phoneme to modify, how to alter the articulatory position) and think over their reactions. Lastly, it is quite important to track the use, and the mentality of the psychologists; it is desirable to enforce practice on a regular and consistent basis, instead of intermittent, because the practice is what will help.

Although the results were positive, there are a number of limitations to the research that should be identified. First of all, the one-group pre-test post-test design does not include a control group, which is extremely hard to use to identify the impact of the AI intervention alone (along with other variables, including maturation and exposure to the English language outside of the treatment). It will be more convincing that future studies employ quasi-experimental or randomized controlled design to prove their causal claims. Secondly, the study fails to give the breakdowns by type of error as well as retention of gains over time (i.e., delayed post-test), but the findings show aggregate improvement. According to the literature, the long-term maintenance and the effect of pronunciation acquisition has not been fully explored in the ASR research. Thirdly, the usage indicators (sessions, time) were tracked, yet no particular qualitative learner engagement and feedback utilization (i.e., the way students reacted to AI

feedback) were analyzed. Next-generation studies ought to record larger volumes of information regarding interaction of learners with the tool, qualitative assessment, and patterns of usage. Moreover, additional research can examine the way proficiency level, L1 background, and motivation can mediate the efficacy of AI speech recognition interventions.

The observed improvement in mean scores, together with classroom observations of better pronunciation awareness and increased confidence, indicates that AI-supported practice may be a useful supplement to pronunciation instruction. Nevertheless, a control group is lacking and restricts more decisive conclusions.

## 5. CONCLUSION

The study has revealed that AI-based speech recognition can be utilized as a feasible tool in training English pronunciation in EFL learners in the university. The possible choice of methods that was conducted in a pre-experimental one group pre-test and post-test design and supported by qualitative observational data and 30 first-term students enabled it to provide feedback that is consistent, individualized, and immediate, which is extremely difficult to provide in a traditional classroom. It did so with the help of AI applications, in particular, ELSA Speak and Google Speech-to-Text. The quantitative findings showed a significant change in pronunciation performance, as the mean scores of students increased significantly in the pre-test (68.4) to the post-test (81.7) and the statistically significant paired-samples t-test also supported this. The improvements in segmental features, including those expressed by interdental sounds (/θ/, /ð/), voicing contrasts (/v/ vs. /f/), and vowel differences, and supra-segmental features, including word stress, rhythm, and intonation, were also validated by qualitative comments. The systematic treatment process made the learners more independent, since they practiced independently, and they checked their progress using AI feedback and became more conscious about their pronunciation problems. The findings are consistent with previous studies on CALL and CAPT that emphasize the utility of AI feedback in pedagogical terms because it promotes accuracy, motivation, and self-regulated learning. Conversely, the absence of a control and short-term test suggests that larger research is needed to establish the sustainability of these gains and to examine how various types of errors and student traits affect the result. In general, the research provides a strong empirical basis for the inclusion of AI speech recognition in pronunciation instruction. It provides viable recommendations to teachers who want to enhance the speaking ability of their students in technology-based and feedback-filled learning settings. The research adds to the evidence of AI speech recognition applications in learning English pronunciation to the classroom.

## 6. REFERENCES

- Abimanto, D., & Sumarsono, W. (2024). Improving English pronunciation with AI speech-recognition technology. *Acitya: Journal of Teaching and Education*, 6(1), 146–156. <https://doi.org/https://doi.org/10.30650/ajte.v6i1.3810>
- Aryanti, R. D., & Santosa, M. H. (2024). A systematic review on artificial intelligence applications for enhancing EFL students' pronunciation skill. *The Art of Teaching English as a Foreign Language (TATEFL)*, 5(1), 102–113. <https://doi.org/https://doi.org/10.36663/tatefl.v5i1.718>
- Benson, P. (2011). Autonomy in language learning. *Language Teaching*, 44(2), 149–171. <https://doi.org/https://doi.org/10.1017/S026144481000040X>
- Benson, P. (2013). *Teaching and researching autonomy (2nd ed.)*. Routledge.
- Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). *Teaching pronunciation: A course book and reference guide (2nd ed.)*. Cambridge University Press.
- Clarke-Jones, L. (2021). Engaging learners in pronunciation: Developing learner autonomy via an action research approach. *Language Teaching*, 54(3), 434–437. <https://doi.org/https://doi.org/10.1017/S0261444821000112>
- Creswell, J. W. (2014). *Research Design Qualitative, Quantitative, and Mixed Methods Approaches Fourth Edition*. SAGE Publication, Inc.
- Cutler, A. (2018). Suprasegmentals: Prosody in English pronunciation teaching. *Language Teaching*, 51(2), 228–246. <https://doi.org/https://doi.org/10.1017/S0261444817000361>
- Dennis, N. K. (2024). Using AI-Powered Speech Recognition Technology to Improve English Pronunciation and Speaking Skills. *IAFOR Journal of Education*, 12(2). <https://doi.org/https://doi.org/10.22492/ije.12.2.05>
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379–397. <https://doi.org/https://doi.org/10.2307/3588486>
- Derwing, T. M., & Munro, M. J. (2009). Put on your listening ears: The role of pronunciation in ESL teaching. *Language Teaching*, 42(4), 482–504. <https://doi.org/https://doi.org/10.1017/S0261444809990144>
- Derwing, T. M., & Munro, M. J. (2015). Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research. *Language Learning & Language Teaching*, 42. <https://doi.org/https://doi.org/10.1075/llt.42>
- Dja'far, V. H., & Hamidah, F. N. (2023). Improving English pronunciation skills through AI-based speech recognition technology. *Ethical Lingua: Journal of Language Teaching and Literature*, 11(2). <https://doi.org/https://doi.org/10.30605/25409190.747>
- Dutta, P., & Arora, P. (2021). Exploring AI-based tools for improving pronunciation skills in ESL learners. *Journal of Applied Linguistics*, 42(2), 231–248. <https://doi.org/https://doi.org/10.1080/17501229.2021.1887984>
- Hidayatulllah, M., Zainuddin, Z., & Hamdani, B. (2025). Exploring speech-recognition technology on English pronunciation skills: A qualitative study. *Jo-ELT*, 12(1). <https://doi.org/https://doi.org/10.33394/jo-elt.v12i1.15230>

- 
- Hsu, L. (2023). Integrating AI-based feedback and human mediation in pronunciation instruction: A hybrid CALL approach. *Language Learning & Technology*, 27(2), 48–65. <https://www.lltjournal.org/27.2/hsu>
- Indari, A. (2023). The detection of pronunciation errors in English speaking skills based on artificial intelligence (AI). *Jurnal Serunai Bahasa Inggris*, 15(2). <https://ejournal.stkipbudidaya.ac.id/index.php/jd/article/view/1007>
- Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.
- Junining, E., Alif, S., & Setiarini, N. (2020). Automatic speech recognition in computer-assisted language learning for individual learning in speaking. *JEES (Journal of English Educators Society)*, 5(2), 219–223. <https://doi.org/https://doi.org/10.21070/jees.v5i2.867>
- Kenworthy, J. (1987). *Teaching English pronunciation*. Longman.
- Kholis, A. (2021). ELSA Speak App: Automatic Speech Recognition (ASR) for supplementing English pronunciation skills. *Pedagogy: Journal of English Language Teaching*, 9(1). <https://doi.org/https://doi.org/10.32332/joelt.v9i1.2723>
- Khurshedjonovna, O. S. (2025). AI technologies for assessing students' pronunciation. *Web of Humanities: Journal of Social Science and Humanitarian Research*, 3(2), 163–167. <https://webofjournals.com/index.php/9/article/view/3275>
- Kruk, M. (2012). Using online resources in the development of learner autonomy and English pronunciation: The case of individual learners. *Journal of Second Language Teaching & Research*, 1(2), 113–142. <https://pops.uclan.ac.uk/index.php/jsltr/article/view/28>
- Levis, J. M. (2018). Technology and pronunciation. In *Cambridge Handbook of English Pronunciation* (pp. 487–504). Cambridge University Press. <https://doi.org/https://doi.org/10.1017/9781316688973.033>
- Martin, I. A. (2023). Motivation in computer-assisted pronunciation training: Online and face-to-face environments. *Language Learning & Technology*, 27(1), 1–21. <https://doi.org/https://doi.org/10.64152/10125/73526>
- Miladiyenti, F., Rozi, F., Haslina, W., & Marzuki, D. (2022). Incorporating mobile-based artificial intelligence to English pronunciation learning in tertiary-level students: Developing autonomous learning. *International Journal of Advanced Science Computing and Engineering*, 4(3), 220–232. <https://doi.org/https://doi.org/10.62527/ijasce.4.3.92>
- Nasir, M., Nazar, F., Abbas, M. K., Waheed, A., & Ahmad, J. (2023). Enhancing pronunciation skills of intermediate students through computer assisted language learning (CALL). *Al-Qantara*, 9(3), 142–160. <https://www.alqantarajournal.com/index.php/Journal/article/view/300>
- Neri, A., Cucchiari, C., & Strik, H. (2008). The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. *ReCALL*, 20(2), 225–243. <https://doi.org/https://doi.org/10.1017/S0958344008000724>
- Ngo, T. T. N., Chen, H. H. J., & Lai, K. K.-W. (2023). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 36(1), 4–21. <https://doi.org/https://doi.org/10.1017/S0958344023000113>

- Nguyen, T. S., Nguyen, T. D. T., & Hoang, N. Q. N. (2025). How AI-powered voice recognition has supported pronunciation competence among EFL university learners. *Computer-Assisted Language Learning Electronic Journal*, 26(3), 64–83. <https://doi.org/https://doi.org/10.54855/callej.252634>
- Noviyanti, S. D. (2022). Artificial intelligence (AI)-based pronunciation checker: An alternative for independent learning in pandemic situation. *ELT Echo: The Journal of English Language Teaching in Foreign Language Context*, 5(2). <https://doi.org/https://doi.org/10.24235/eltecho.v5i2.7246>
- Nursyafida, & Putri, R. E. (2025). Exploring technology use in English pronunciation instruction: A systematic review of trends and tools. *SALEE: Study of Applied Linguistics and English Education*, 6(2), 414–436. <https://doi.org/https://doi.org/10.35961/salee.v6i2.414-436>
- Saeed, A. (2020). The impact of speech recognition technology on language learning. *International Journal of Educational Technology in Higher Education*, 17(1), 1–15. <https://doi.org/https://doi.org/10.1186/s41239-020-00209-w>
- Saeed, A., & Sharma, K. (2020). AI in ESL classrooms: Improving pronunciation through speech recognition tools. *Journal of Language Teaching and Research*, 11(6), 967–975. <https://doi.org/https://doi.org/10.17507/jltr.1106.14>
- Sardegna, V. G., & McGregor, A. (2022). Classroom research for pronunciation. In J. M. Levis & T. M. Drawing (Eds.), *Second Language Pronunciation: Bridging the Gap Between Research and Teaching*. John Wiley & Sons, Inc.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/https://doi.org/10.1093/applin/11.2.129>
- Spring, R., & Tabuchi, R. (2022). The role of ASR training in EFL pronunciation improvement: An in-depth look at the impact of treatment length and guided practice on specific pronunciation points. *Computer-Assisted Language Learning Electronic Journal*, 23(3), 163–185. <https://callej.org/index.php/journal/article/view/417>
- Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. *Frontiers in Psychology*, 14. <https://doi.org/https://doi.org/10.3389/fpsyg.2023.1210187>
- Sun, Y. (2024). The application of intelligent speech recognition in the teaching of spoken English in colleges and universities. *Applied Mathematics and Nonlinear Sciences*, 9(1). <https://doi.org/https://doi.org/10.2478/amns-2024-2125>
- Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, 36(3), 326–344. <https://doi.org/https://doi.org/10.1093/applin/amu076>
- Vančová, H. (2024). AI and AI-powered tools for pronunciation training. *Journal of Language and Cultural Education*, 11(3), 12–24. <https://doi.org/https://doi.org/10.2478/jolace-2023-0022>
- Vassallo, S., Boulos, M. N. K., & Goller, P. (2019). AI in language learning: Challenges and opportunities for speech recognition. *Computers & Education*, 137, 1–12. <https://doi.org/https://doi.org/10.1016/j.compedu.2019.04.007>

- Wen, X., & Li, Y. (2023). Beyond accuracy: Teacher mediation in AI-assisted pronunciation learning. *ReCALL*, 35(1), 88–106.  
<https://www.cambridge.org/core/journals/recall/issue/E9F26F8863E0DDA0DF264E2A888D518F>
- Zainuddin, N. (2024). Technology enhanced language learning research trends and practices: A systematic review (2020–2022). *Electronic Journal of E-Learning (EJEL)*, 21(2).  
<https://doi.org/https://doi.org/10.34190/ejel.21.2.2835>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70. [https://doi.org/https://doi.org/10.1207/s15430421tip4102\\_2](https://doi.org/https://doi.org/10.1207/s15430421tip4102_2)