Teaching English as a Foreign Language, Linguistics, and Literature Journal
Vol. 1(2), 2025
ISSN 3063-8879

AI Based Applications to Foster Pronunciation Accuracy

Ahsin Fahmi Mubarok[1], Mochammad Abdullah Salim[2], Zhenita Deliany[3]

[1] *Politeknik Negeri Malang. e-mail: ahsinfahmi@polinema.ac.id*

[2] *Institut Agama Islam Tribakti. e-mail: abdullahsalim@uit-lirboyo.ac.id*

[3] *Universitas Jember. e-mail: zhe.deliany@gmail.com*

ARTICLE INFO

Essence, *Teaching English as a Foreign Language, Linguistics, and Literature Journal,* Vol(2) 1, 2025

DOI:
http://dx.doi.org/10.21093/Essence.vxxixx.xxx

ABSTRACT

Accurate pronunciation is fundamental to effective communication. Meanwhile, many EFL learners struggle due to limited exposure and first language interference. This study investigates the effectiveness of two AI-based applications ElsaSpeak and ElevenLabs in improving English pronunciation accuracy among Indonesian vocational students. Employing a comparative descriptive method, 20 first-semester students were divided into two groups based on equal pre-test scores. Each group received three pronunciation practice sessions using one of the tools. Post-test results showed significant improvement in both groups, with ElsaSpeak yielding slightly higher gains due to its real-time, phoneme-level feedback. Learner responses confirmed positive experiences with both tools, although ElsaSpeak was preferred for its interactive features. These findings suggest that AI-assisted pronunciation tools can serve as effective supplements to traditional instruction by offering individualized, flexible learning. The study highlights the value of integrating such technologies into EFL pedagogy to enhance pronunciation development and promote learner autonomy.

Keywords:
English Pronunciation, Artificial Intelligence, ElsaSpeak, ElevenLabs, EFL Learners, Pronunciation Accuracy

How to cite:
Mubarok, A. F., Salim, M. A., & Deliany, Z. (2025). AI-based applications to foster pronunciation accuracy. Essence, Teaching English as a Foreign Language, Linguistics, and Literature Journal, 2(1), 2025.

Introduction

Possessing high English pronunciation accuracy is crucially needed for administering collaboration and communication which become two of four skills required in the 21st century. By having that accuracy, someone is able to minimize misunderstanding and enhance clarity in performing complex communication as it ensures intended message by speaker is received precisely by the listener with no or minimal ambiguity. Several positive effects for having this ability also facilitate someone to be successful in many areas such as in academic context where accurate pronunciation helps students to understand lectures better and in workplace setting where accurate pronunciation is indispensable for executing team presentations, contributing in successful negotiations, and building rapport with clients and colleagues (Poposka, 2017; Ho, 2018). Although accurate pronunciation has huge impact on someone's life development, numerous learners still have difficulty in achieving the targeted accuracy level especially those who learn English as their second language such as in Indonesia.

When attempting to attain excellent English pronunciation accuracy in a variety of contexts, Indonesian learners frequently encounter difficulties. The first thing they find difficult is the fact that there are fewer vowel sound variants in Indonesian than in other languages (Sirait, Hutagaol, Simanjuntak & Sitanggang 2023). This makes it hard to distinguish one English vowel from another, like /ɪ/ (used in "bit") and /i:/ (used in "beat") or /ae/ (used in "cat") and /ʌ/ (used in "cut"). Furthermore, there are numerous consonant clusters in English (such as "str-," "spl-," and "thr-") that are rare or nonexistent in Indonesian (Younis, 2022). Because of this, pronouncing these clusters might be challenging. Not only that, final consonants in English words also complicate Indonesian learners to reach English pronunciation accuracy since in Indonesian, many words end in vowels. In the non-contrastive analysis aspects, many Indonesian students, particularly those who reside in rural areas, have little exposure to English-speaking environments, which also limits their access to real resources like English audio and video. The majority only use Indonesian English that they hear from other Indonesian speakers, which might lead to mistakes being repeated. Seeing this obstacle, an effort to foster Indonesian accuracy in pronouncing English is necessary to be taken into account.

Artificial Intelligence (AI) based learning tools are gaining their popularity in recent years due to its conveniences for learners in all disciplines including pronunciation learning (Mohammadkarimi, 2024; Kristiawan, 2024; Anh, 2024). AI-assisted English pronunciation instruction offers the benefit in the form of flexibility, as it may be used at any time and from any location to suit learners' needs, including after school. Additionally, AI can give learners proper pronunciation examples of words, phrases, or sentences in a variety of models and dialects. Additionally, some AI systems provide comments on student recordings, highlight pronunciation faults in certain portions, and offer activities to fix them. Numerous researches have demonstrated the efficacy of using AI to acquire English pronunciation such as done by Abimanto and Sumarsono (2024) and Rusmiyanto et al., (2023). However, the majority of these studies don't focus on the type of AI utilized; instead, they just look at how scores improve before and after AI is applied in learning. This study therefore compares the efficacy of Elsa Speak and ElevenLabs in enhancing the accuracy of English pronunciation and determining the user experience following pronunciation instruction using two distinct AI types.

Method

Using a comparative descriptive approach, this study compares the accuracy of English pronunciation using Elsa Speak and ElevenLabs. Following treatment, participants are given a questionnaire to complete in order to record their experiences using the two tools. Students enrolled in the D3 mechanical engineering study program at the State Polytechnic of Malang served as the study's sample. Participants were 1st semester students who were taking the English engineering 1 course. 20 students participated in the study; ten of them used Elsa Speak to improve their pronunciation accuracy, while ten used ElevenLabs. A pretest measuring pronunciation accuracy was administered to 60 students (shown by Table 1), from which 20 participants were chosen (shown by Table 2). The pretest results served as the foundation for dividing groups A and B with an average pronunciation accuracy that was equivalent.

Table 1

Score Gained from Pre-test

Student's Number & Score	Student's Number & Score	Student's Number & Score	Student's Number & Score	Student's Number & Score	Student's Number & Score
1 = 75	11 = 72	21 = 73	31 = 60	41 = 65	51 = 70
2 = 65	12 = 68	22 = 70	32 = 72	42 = 72	52 = 68
3 = 77	13 = 67	23 = 75	33 = 67	43 = 75	53 = 65
4 = 61	14 = 70	24 = 67	34 = 65	44 = 77	54 = 70
5 = 65	15 = 65	25 = 65	35 = 75	45 = 65	55 = 67
6 = 78	16 = 70	26 = 77	36 = 72	46 = 60	56 = 60
7 = 60	17 = 75	27 = 75	37 = 78	47 = 64	57 = 65
8 = 66	18 = 71	28 = 66	38 = 76	48 = 70	58 = 78
9 = 70	19 = 65	29 = 63	39 = 75	49 = 72	59 = 78
10 = 72	20 = 75	30 = 68	40 = 72	50 = 75	60 = 73

Table 2

Group Distribution

Group A (ElsaSpeak)		Group B (ElevenLabs)	
58 = 78	35 = 75	59 = 78	39 = 75
6 = 78	43 = 75	37 = 78	50 = 75
3 = 77	21 = 73	26 = 77	60 = 73
1 = 75	10 = 72	17 = 75	11 = 72
23 = 75	32 = 72	27 = 75	42 = 72

This research lasted for 2 months starting from February-March 2025. The pretest was conducted on February 1st then the results were analyzed on February 2nd. The research process was continued by providing 3 treatment meetings to 2 groups with 2 different tools on February 6th, 12th, and 18th. The post-test to determine the impact of learning with the 2 tools on students' pronunciation accuracy was conducted on February 24th. The post-test results were analyzed using a t-test to determine whether there was a significant difference in accuracy values from learning pronunciation with 2 different tools. The pretest material given to the participants was a procedure text for operating a SMAW welding machine. The same text was also used for treatment 3 times to the participants by dividing 1 complete text into 3 parts. The text was then retested to determine the increase in pronunciation accuracy.

Findings and Discussion

Findings

To obtain two groups with the same pronunciation ability, a pre-test to obtain initial pronunciation ability was conducted on 60 students. From that number, 20 students were selected purposively to obtain 2 groups with the same average score. Table 3 shows the statistical data of the 20 selected students and Table 4 shows the significance of the difference in the average between the two groups.

Table 3

Statistics for Selected Students for Two Groups

	Group	N	Mean	Std. Deviation	Std. Error Mean
Pre-test	A	10	75.00	2.211	.699
	B	10	75.00	2.211	.699

Table 3 shows the same average value, that is 75, so there is no significant difference in the average as indicated by the sig value above 0.05 (1.00) in Table 4.

Table 4

Mean Difference of The Two Groups for Pre-Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
			t	df	Sig. (2-tailed)	Mean difference	Std. Error Difference	95% Confidence Interval of the Difference	
	F	Sig.						Lower	Upper
Equal variances assumed	.000	1.000	.000	18	1.000	.000	.989	-2.077	2.077
Equal variances not assumed			.000	18.000	1.000	.000	.989	-2.077	2.077

After using ElsaSpeak and ElevenLabs three times for three weeks, the pronunciation accuracy of both groups was retested to determine whether or not there

was an increase in accuracy. Tables 5 and 6 are statistical data of the scores obtained by Group A who learned pronunciation by using ElsaSpeak before and after treatment while tables 7 and 8 are the scores obtained by Group B who learned pronunciation by using ElevenLabs before and after treatment.

Table 5

Mean Comparation of Group A Before and After Treatment

A group	N	Mean	Std. Deviation	Std. Error Mean
Pre-test	10	75.00	2.211	.932
Post-test	10	79.70	2.946	.699

The difference in the average pre-test and post-test scores of 4.7 was tested using a t-test in SPSS to determine whether or not the difference was statistically significant. The significant value in Table 6 shows a value of 0.00 (below 0.05) so that this test proves that there is a significant increase in pronunciation accuracy before and after learning pronunciation using ElsaSpeak.

Table 6

Mean Difference of Group A for pre-test and post-test

	t	df	t-test for Equality of Means			
			Sig. (2-tailed)	Mean difference	95% Confidence Interval of the Difference	
					Lower	Upper
Pre-test	107.265	9	.000	75.000	73.42	76.58
Post-test	85.557	9	.000	79.700	77.59	81.81

The mean difference was also tested in group B who learned pronunciation using ElevenLabs. The difference of 3.5 was proven to be statistically different as shown by the sig value of 0.00 in Table 8.

Table 7

Mean Comparation of Group B Before and After Treatment

A group	N	Mean	Std. Deviation	Std. Error Mean
Pre-test	10	75.00	2.211	.699
Post-test	10	78.50	2.991	.946

Table 8

Mean Difference of Group B for pre-test and post-test

	t	df	t-test for Equality of Means			
			Sig. (2-tailed)	Mean difference	95% Confidence Interval of the Difference	
					Lower	Upper
Pre-test	107.265	9	.000	75.000	73.42	76.58
Post-test	83.003	9	.000	78.500	76.36	76.58

After proving that there was an increase in pronunciation accuracy from the pre-test to the post-test, Table 9 shows the difference in the average pronunciation accuracy scores after learning pronunciation by using ElsaSpeak and ElevenLab in the post-test.

Table 9

Independent t-test to measure mean difference between group A and B for post-test

Post-test	**Levene's Test for Equality of Variances**		**t-test for Equality of Means**						
			t	df	Sig. (2-tailed)	Mean difference	Std. Error Difference	95% Confidence Interval of the Difference	
	F	Sig.						Lower	Upper
Equal variances assumed	.042	.839	.904	18	.378	1.200	1.327	-1.589	3.989
Equal variances not assumed			.904	17.996	.378	1.200	1.327	-1.589	3.989

From the value indicated by the sig value of 0.839 (more than 0.05), it can be concluded that there is no significant difference in learning outcomes after learning pronunciation by using the two AI-based software.

After the post-test results were analyzed, 2 sets of questionnaires were given to groups A and B to capture the learning experience felt by students after learning pronunciation using ElsaSpeak and ElevenLabs. There were 10 questions asked using a Likert scale of 1-5 (1 strongly disagree, 2 disagree, 3 neutral, 4 agree, 5 strongly agree) the results of which are listed in Tables 10 and 11 below:

Table 10

Ten Questions for Students Who Learn Pronunciation by Using ElsaSpeak

No	Questions	Results
1	Using ElsaSpeak has made me feel that my accuracy in pronouncing English words has improved.	4.1
2	I feel more confident that I am pronouncing English sounds correctly after using ElsaSpeak.	1

3	ElsaSpeak has helped me to identify and correct specific pronunciation errors that I used to make.	4.1
4	I feel that my pronunciation now sounds more like a native English speaker due to my practice with ElsaSpeak.	4
5	Using ElsaSpeak has increased my awareness of the small differences between similar English sounds, leading to more accurate pronunciation.	4.1
6	I feel that the feedback I received from ElsaSpeak helped me to focus on and improve my pronunciation accuracy.	4.3
7	Practicing with ElsaSpeak has made me feel more precise in the way I form and pronounce English words.	3.9
8	I believe that ElsaSpeak has contributed significantly to the overall accuracy of my spoken English.	4.1
9	I feel that using ElsaSpeak has helped me to pronounce English words with the correct stress and intonation, leading to more accurate delivery.	4
10	Overall, I feel satisfied with how ElsaSpeak has helped me improve the accuracy of my English pronunciation.	4

Table 11

Ten Questions for Students Who Learn Pronunciation by Using ElevenLabs

No	Questions	Results
1	Using ElevenLabs has made me feel that my accuracy in pronouncing English words has improved.	4
2	I feel more confident that I am pronouncing English sounds correctly after using ElevenLabs.	3.9
3	ElevenLabs has helped me to identify and correct specific pronunciation errors that I used to make.	3.8
4	I feel that my pronunciation now sounds more like a native English speaker due to my practice with ElevenLabs.	4
5	Using ElevenLabs has increased my awareness of the small differences between similar English sounds, leading to more accurate pronunciation.	3.9
6	I feel that the feedback I received from ElevenLabs helped me to focus on and improve my pronunciation accuracy.	4
7	Practicing with ElevenLabs has made me feel more precise in the way I form and pronounce English words.	4
8	I believe that ElevenLabs has contributed significantly to the overall accuracy of my spoken English.	4
9	I feel that using ElevenLabs has helped me to pronounce English words with the correct stress and intonation, leading to more accurate delivery.	3.9
10	Overall, I feel satisfied with how ElevenLabs has helped me improve the accuracy of my English pronunciation.	4

Discussion

The Effectiveness of ElsaSpeak in Enhancing Students' Pronunciation Accuracy

The effectiveness of ElsaSpeak in fostering students' pronunciation accuracy can be attributed to its tailored design, which specifically addresses common learner difficulties through real-time feedback and phoneme-level analysis. One key strength lies in its AI-powered error detection system, which enables learners to recognize mispronunciations as they occur and immediately receive corrective input. This form of direct feedback is especially helpful for Indonesian learners, who often struggle with English sounds not found in their first language, such as the distinction between /ɪ/ and /i:/ or /θ/ and /t/. Previous research has shown that when learners are given timely and individualized feedback, their ability to self-monitor improves significantly (Liakin, Cardoso, & Li, 2017; McCrocklin, 2019).

In addition to identifying errors, ElsaSpeak provides clear visual cues and breakdowns of individual phonemes. These visual aids guide learners in adjusting their articulation, thereby supporting both auditory and kinesthetic awareness. This dual-channel input is important for learners who may not have access to native-speaker models or sufficient pronunciation instruction in traditional classroom settings. Studies by Foote and Trofimovich (2016) and Kristiawan (2024) emphasize the value of focused practice on problematic phonemes, especially when paired with adaptive technological support.

What sets ElsaSpeak apart is its integration of suprasegmental training, including emphasis on intonation, rhythm, and word stress. These elements are frequently overlooked in formal education but are essential to sounding natural and comprehensible in real-life interactions. Derwing and Munro (2015) argue that prosody—the musicality of speech—contributes significantly to perceived fluency, often more than correct articulation alone. By offering interactive activities that target these areas, ElsaSpeak gives learners a more holistic experience of pronunciation learning.

Moreover, the students' responses to the post-treatment questionnaire suggest a high level of satisfaction, particularly in how the application helped them recognize subtle sound differences and build confidence in their speech production. Similar observations were noted by Wang and Young (2015), who found that learner engagement increased notably when digital tools were perceived as responsive and goal-oriented.

The Role of ElevenLabs in Improving Students' Pronunciation Accuracy

Although ElevenLabs is not designed as a pronunciation learning tool, its application in this study yielded notable improvements. Its strength lies primarily in the quality of its audio output, which offers learners a consistent and accurate model of native-like pronunciation. For many learners, especially those in non-English speaking environments, repeated exposure to clear and natural speech can lead to better phonological awareness and mimicry (Saito, 2013).

Another contributing feature is ElevenLabs' customizable voice settings and adjustable playback speed. These allow learners to slow down speech, repeat specific sections, and practice in a way that suits their learning pace. According to Kormos and Dénes (2004), slowing down auditory input helps learners detect phonemic and prosodic

features more effectively, which is essential for those still developing an ear for English phonology.

The platform's text-to-speech functionality also proved useful during the treatment sessions. By converting the same procedural text into natural speech, learners were able to engage in shadowing and repetition exercises—techniques known to promote fluency and intonation control (Murphey, 2001). These features made ElevenLabs a flexible tool that supported pronunciation practice even though it lacked direct corrective mechanisms.

Participants' feedback reinforces these findings. Most reported feeling that their pronunciation had improved and that the tool increased their awareness of stress and intonation, even in the absence of structured guidance. These responses align with Bui and Huang (2022), who suggest that high-quality TTS systems, when paired with learner autonomy, can lead to meaningful gains in oral language skills.

Identified Limitations in Using ElevenLabs for Pronunciation Training

Despite its utility, ElevenLabs has limitations that merit attention. One of them is the absence of immediate, diagnostic feedback—a feature that ElsaSpeak excels at. Without real-time correction, learners must rely on self-assessment, which can be difficult for those who have not developed a sensitive ear for English phonemes. Hincks (2003) points out that feedback, especially when provided immediately after an error, plays a crucial role in helping learners adjust their articulation patterns effectively.

This limitation likely explains why students in the ElevenLabs group, while showing improvement, did not experience gains as pronounced as their ElsaSpeak counterparts. As noted by Neri et al. (2008), the absence of visual and auditory feedback can lead to repetition of errors, which may go uncorrected without expert or AI support.

In this context, ElevenLabs is best viewed as a complementary tool rather than a standalone pronunciation solution. While it provides valuable input and listening practice, its lack of interactivity constrains its effectiveness for learners who require explicit guidance and correction.

Conclusion

This study confirms that both ElsaSpeak and ElevenLabs effectively enhanced students' English pronunciation accuracy. ElsaSpeak's specialized feedback and phoneme-focused training yielded slightly higher improvement due to its targeted features. In contrast, ElevenLabs supported pronunciation development through high-quality audio modeling and flexible TTS features, though its lack of interactive feedback limited deeper correction.

Despite no significant statistical difference between the two tools, learners reported positive experiences with both, indicating that AI-based applications can meaningfully support EFL learners' oral proficiency. Future studies should involve larger samples, longer durations, and consider classroom integration to explore long-term impacts and broader applicability.

References

Abimanto, D., & Sumarsono, W. (2024). Improving English Pronunciation with AI Speech-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/10.30650/ajte.v6i1.3810

Anh, N. H. (2024). Harnessing AI-based tools for enhancing English speaking proficiency: Impacts, challenges, and long-term engagement. *International Journal of AI in Language Education, 1*(2), 18–29. https://doi.org/10.54855/ijaile.24122

Bui, T., & Huang, J. (2022). The effects of TTS tools on EFL learners' pronunciation and listening comprehension. *International Journal of Language and Linguistics, 10*(2), 89–97. https://doi.org/10.11648/j.ijll.20221002.12

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins Publishing.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363–406.

Foote, J. A., & Trofimovich, P. (2016). Pronunciation instruction: Recent research and implications for teaching. *TESL Canada Journal, 33*(1), 1–22. https://doi.org/10.18806/tesl.v33i1.1235

Hincks, R. (2003). Speech technologies for pronunciation feedback and evaluation. *ReCALL, 15*(1), 3–20. https://doi.org/10.1017/S0958344003000211

Ho, P. V. P. (2018). Fluency as successful communication. In *Proceedings of the 1st National Conference on English Language Teaching Upgrade: A Focus on Fluency (CELTU)*.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*(2), 145–164. https://doi.org/10.1016/j.system.2004.01.001

Kristiawan, D. Y. (2024). Artificial intelligence in English language learning: A systematic review of AI tools, applications, and pedagogical outcomes. *The Art of Teaching English as a Foreign Language (TATEFL), 5*(2), 207–218. https://doi.org/10.36663/tatefl.v5i2.912

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

Liakin, D., Cardoso, W., & Li, S. (2017). The effect of text-to-speech technology on L2 French pronunciation. *The Canadian Modern Language Review, 73*(3), 277–306. https://doi.org/10.3138/cmlr.3316

McCrocklin, S. M. (2019). ASR-based dictation: A tool for pronunciation feedback and improvement. *Computer Assisted Language Learning, 32*(8), 849–874. https://doi.org/10.1080/09588221.2018.1531033

Mohammadkarimi, E. (2024). Exploring the use of artificial intelligence in promoting English language pronunciation skills. *LLT Journal, 27*(1), 98–115. https://doi.org/10.24071/llt.v27i1.8151

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System, 34*(4), 520–531. https://doi.org/10.1016/j.system.2006.09.004

Murphey, T. (2001). Exploring conversational shadowing. *Language Teaching Research, 5*(2), 128–155. https://doi.org/10.1177/136216880100500203

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2008). The pedagogy–technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning, 21*(5), 441–467. https://doi.org/10.1080/09588220802447843

Poposka, V. P. (2017). A study of proper pronunciation as a factor of successful communication. In *Proceedings of the Central European Conference on Teaching and Learning* (pp. 778–783). http://dx.doi.org/10.12955/cbup.v5.1024

Rusmiyanto, H., Huriati, N., Fitriani, N., Tyas, N. K., Rofi'i, A., & Sari, M. N. (2023). The role of artificial intelligence (AI) in developing English language learners' communication skills. *Journal on Education, 6*(1), 750–757.

Saito, K. (2013). Reexamining effects of form-focused instruction on L2 pronunciation development. *Studies in Second Language Acquisition, 35*(1), 1–29. https://doi.org/10.1017/S0272263112000666

Sirait, M. I., Hutagaol, Y. A., Simanjuntak, P. P., & Sitanggang, A. (2023). A contrastive analysis of vowel in Indonesian and English. *PIJAR: Jurnal Pendidikan dan Pengajaran, 1*(3), 400–406. https://doi.org/10.58540/pijar.v1i3.392

Wang, Y., & Young, S. S. C. (2015). A study of the design and evaluation of an online interactive speaking system using automatic speech recognition technology for EFL learners. *Computer Assisted Language Learning, 28*(6), 535–555. https://doi.org/10.1080/09588221.2014.881386

Younis, R. Z. (2022). A study of consonant clusters in English compound words. *Journal of Tikrit University for the Humanities, 29*(6), 59–82. http://dx.doi.org/10.25130/jtuh.29.6.2022.25