Students' Perceptions of AI-based Speech Recognition in the ELSA Speak Application for Pronunciation Practice

Rina Marliana[1], Nasyrah Sakinah[1*], Selvi[1], Rahmat [1]

[1]*English Education, Tarbiyah and Teaching Department, Sekolah Tinggi Agama Islam Negeri (STAIN) Majene*

*E-mail: nasyrahsakinahh@gmail.com

Abstract

This study investigates students' perspectives on the effectiveness of ELSA Speak in improving pronunciation skills. Speech recognition on the ELSA Speak application can help one's pronunciation skills. Using the descriptive qualitative research method, the researchers focused on English Language Education students at STAIN Majene, specifically selecting five sixth-semester students as samples. The selection of this sample was based on the fact that the students had used the ELSA Speak application as a medium to conduct pronunciation exercises. To collect data, the researcher used a structured interview consisting of five questions. The findings showed that the majority of students found the speech recognition feature of the ELSA Speak helpful and at the same time, they also experienced challenges in it. The positive result is that they feel their pronunciation is getting better, and then they learn how to pronounce and pronounce a word correctly in English. The AI-based ELSA Speak provides corrections so that users know what percentage they lack, especially those who are introverted. They felt accuracy directly when using the voice recognition feature in the ELSA Speak. Most students found Elsa Speak's speech recognition feature useful to improve their pronunciation. Although some of them also faced some challenges, they felt the benefits of ELSA Speak more. This also creates interest in continuing to use ELSA Speak, that not only one skill but also skills that are important for English students to master can be achieved. The positive results listed in this study include improved pronunciation accuracy and a better understanding of the articulation of English words and sentences, as verified by the application's speech recognition feature.

***Keywords**: AI (Artificial Intelligence), Speech recognition, ELSA Speak, Pronunciation.*

INTRODUCTION

Teaching pronunciation is an important aspect of learning a foreign language. The ability to pronounce words and expressions in the target language correctly plays a crucial role in communicating effectively. Good pronunciation not only helps convey messages clearly but also increases learners' confidence in using the language. However, it is often a challenge for teachers to provide adequate feedback and practice for each student. Large language classes with a large number of students limit the time and attention the teacher can give to each individual. Apart from that, differences in students' initial abilities and learning styles also demand a more personal and measured approach to teaching pronunciation.

Another complicating factor is limited resources and learning aids. Teachers often rely solely on their hearing to correct students' pronunciation errors, which can be subjective and less accurate. Meanwhile, access to language laboratory facilities or specialized software for pronunciation training is limited, especially in resource-limited settings.

Therefore, innovative solutions are needed that can help overcome challenges in teaching foreign language pronunciation. A new approach utilizing advanced technology has the potential to provide an effective and affordable alternative for teachers and students to practice and correct pronunciation. With the development of artificial intelligence (AI) technology, new solutions have emerged in the form of AI-based speech recognition applications that can help students practice pronunciation independently. These apps utilize machine learning algorithms to analyze user speech and provide real-time feedback and suggestions for improvement. One of the popular applications in this field is Elsa Speak, which was developed by a technology company from Vietnam. Elsa Speak uses advanced speech recognition technology to detect pronunciation errors and help users improve their pronunciation through practice and personalized feedback (dai et al., 2020). Research shows that the use of applications such as Elsa speak can significantly improve language learners' pronunciation skills (Liakin et al., 2015).

The main advantage of the AI-based speech recognition application is that the system recognizes the user's speech while honing pronunciation skills. This system is called speech recognition. Speech recognition is one of the forms of artificial intelligence, according to Levis and Su Orov (2013) in research by Bashori, Van Flout, Strik, and Cucchiarni, speech recognition is a product of technological advances that can perform the function of translating and copying spoken speech. This technology has continued to develop rapidly in recent years, driven by advances in the field of artificial intelligence (AI).

Speech recognition is a technology that allows machines to identify and understand human speech and convert it into text (Sharma et al, 2020). It involves the processes of acoustic feature extraction, pattern matching, and speech recognition. Through speech recognition, humans can communicate with machines only through voice. Even though there is a complex system and process behind it, this technology can run very quickly so that people feel comfortable using it. This system is one of the AI-based technologies in the ELSA Speak application which is very helpful in improving the pronunciation abilities of application users. So this can train and make foreign language learning for its users more optimal and effective. Therefore, it is necessary to know the extent or opinion of the users of the ELSA speak application and have enjoyed this speech recognition system directly.The study by Quintero et al. (2019) studied students' perceptions of using a similar application and found that most participants felt it helped to improve their pronunciation skills. Meanwhile, research by Handoko et al. (2021) revealed several challenges in using speech recognition applications, such as technical problems and limited feedback provided. By studying student perceptions, we can better understand the strengths and limitations of applications like ELSA Speak, as well as identify areas for improvement. This research aims to explore students' views on the use of ELSA Speak as a tool for pronunciation practice, as well as the factors that influence their perceptions.

METHOD

This research design uses qualitative descriptive research methods. It was designed to explore students' perceptions of AI-based speech recognition in the ELSA Speak for pronunciation practice. The population in this study consisted of English education students at STAIN Majene. Meanwhile, the sample of this research were six semester students, with a total of five students that selected through purposive sampling. This

method is very useful in qualitative research, especially when the researchers aims to identify and select information-rich cases related to the phenomenon of interest (Ames et al.,2019). In purposive sampling, researchers use their expert judgment to select participants who are best suited to answer the research questions. This approach allows researchers to focus on individuals or cases that have specific experiences or knowledge relevant to the study.

One of the main strengths of purposive sampling is its ability to generate an informative sample with limited resources (Serra et al., 2023). However, it is important to remember that this method can be prone to researcher bias and may not be representative of the wider population. Therefore, the researcher focused on this purposive sample considering that students have used the ELSA Speak application as a medium to do some pronunciation practice. This purposive sampling can help researchers identify participants who have significant experience with the technology under study (Korstjens & Moser, 2021). This can lead to richer and more relevant insights than random sampling. So, the sample size is set at 5 students, which is deemed sufficient for this qualitative study to reach data saturation while allowing for in-depth analysis of each participant's experience (Braun & Clarke,2021).

In order to collected data, the researchers used a structured interview consisting of five questions. Each interview will last approximately 20-30 minutes and conducted in a quiet, comfortable setting on campus. With participants' consent, the interviews were recorded for accurate transcription and analysis. The questions asked by the researchers were based on various important aspects of using the ELSA Speak application, such as duration using the ELSA Speak application, the advantages and disadvantages of speech recognition features, and the improvement of pronunciation using this application.

Ethical approval was obtained from the university's ethics committee. Informed consent was secured from all participants, ensuring they understand the nature of the study and their rights (Roth & von Unger,2023). Confidentiality was maintained by using pseudonyms and securely storing all data. This studies also have limitations that include its small sample size and focus on a single university, which may limit the generalizability of findings (Smith & Nobles, 2020). Additionally, the study captures perceptions at a specific point in time and may not reflect changes in perceptions over extended use of the application.

This quantitative descriptive approach provided valuable insight into students' perceptions of AI-based speech recognition in ELSA Speak, contributing to our understanding of how such AI technologies are experienced in a language learning context.

RESULTS AND DISCUSSION

1.1 RESULTS

The data in this study were obtained through a structured interview consisting of 5 questions. This interview was conducted with five 6th-semester students of English education at STAIN Majene to find out the students' perceptions of the AI-based speech recognition feature in the ELSA Speak pronunciation practice application.

The interview was carried out face-to-face with some of the students and with others through online interactions, such as via mobile phones and over the phone via WhatsApp. The results were therefore mixed. Each of these questions is presented in each point. The

answers from the interviews gave a variety of results from each point of the question given by the researcher. Some explanations of these results are explained below. Length of time using the application in the interview. The researcher asked all participants how long they had been using the Elsa Speak application. Here are the answers from each participant:

R said:
"I have been using Elsa Speak for the past two weeks"

T said:
"For the past week, I have started trying to use the elsa application speak this."

R said:
"I haven't used ELSA Speak for a long time. Approximately 3 months."

N said:
"I have been using this app for the past month."

NS said:
"For the past few weeks, I have been learning English using the ELSA Speak application."

According to the result of the interviews above from the five participants, it can be concluded that the five participants are novice users of the ELSA Speak application. Starting from a week, two weeks, a month to three months. So we were not able to get maximum results because of the short time the participants used the ELSA Speak application. What's unique is that when they talked about their experience of using the application, they seemed enthusiastic and excited when using this application. Researchers concluded that, even though users are still relatively short in using Elsa speak, it has given its users a pleasant impression.

1. Do the responses and conversations given by the ELSA Speak app sound natural or good? In the interview, the author asked each participant about how long they had been using the ELSA Speak application as a learning media for pronunciation. The following are the answers from out of five students; the answers from each participant

R said:

"For free applications, the response is quite natural."

T said:

"Alhamdulillah, so far it sounds good and natural according to the pronunciation."

R said:

"I think the responses and conversations provided by the ELSA Speak application sound natural, good, and even clear. However, if you want the feed you receive to be more responsive, you should subscribe to the premium version for better results."

N said:

"yes, the response to the free application is very good."

NS said:

"It's quite good and very helpful in finding good pronunciation answers."

Based on interviews with five respondents, it can be concluded that the responses from the Elsa Speak application sound quite natural and help in improving pronunciation skills. However, one respondent suggested subscribing to the premium version for more responsive results and more comfort to use. Then the feature (voice record) is very effective for students, especially English language students, both free and premium versions, are considered good and very helpful in improving speaking skills, especially pronunciation.

2. Comments regarding the advantages and disadvantages of the speech recognition (voice record) feature in the ELSA Speak app? With different responses, respondents answered as follows:

R said:

"Because I haven't used this application for too long, in my opinion so far the application is good and helps me, in my opinion, there are no drawbacks because this app helps its users a lot."

T said:

"When we use the voice recording application Elsa Speak, helps me because I am an introverted child, so I am free to express myself without being afraid of making mistakes or anything like that because the one who provides corrections is AI. I haven't found any drawbacks yet, because I've only been using this application for about a week."

R said:

"The advantage is that we know where our pronunciation errors are (stress, intonation, etc.). However, the disadvantage is that because the Elsa Speak application uses mostly British accents, sometimes some accents from non-British users cannot be read accurately."

N said:

"Advantages, providing specific feedback. The downside is that some of the features require a paid premium."

NS said:

"The advantage is in the speech recognition feature because the feature can detect our level of pronunciation. The disadvantage is that the speech recognition feature is limited, if we want to use it further, we have to switch to premium or paid."

It can be concluded that the ELSA Speak application is rated as a positive application in helping users improve their English speaking skills. Even though some respondents have not used this application for a long time, they said that the ELSA Speak application is very effective in providing specific feedback regarding pronunciation errors.

Respondents can also express themselves freely when learning pronunciation, and then AI-based ELSA Speak provides corrections so that users know what percentage they are lacking, especially those who are introverts. However, several shortcomings were also identified, such as the application's tendency to prioritize British style accents in its corrections, because all users tend to use American style so some of them have difficulty adjusting. Then there is limited access to certain features that are only available to premium users, and the speech recognition feature can only be used in full by subscribing to the premium version. Nonetheless, the general impression from these interviews was positive, with users recognizing the significant benefits the app provided in respondents' efforts to improve their English-speaking skills.

3. Challenges in using the AI-based speech recognition feature of the ELSA Speak application. Participants gave very diverse answers regarding challenges when using the speech recognition feature in this application.

R said:

"The challenge with the speech recognition feature is more for me. I have a voice that tends to be soft and small. So when I use this speech recognition feature, the results are sometimes wrong because the feature cannot capture my voice."

T said:
"When I use the speech recognition feature, the results provided are sometimes unclear. I don't know why. I am confused as to why the feature is not working. Or maybe someone doesn't understand how to use this feature."

R said:
"I think I haven't found the Feature speech recognition challenge so far."

N said:
"The challenge is that sometimes the mic device doesn't work. So I couldn't optimally test my pronunciation method in learning a foreign language."

NS said:
"I didn't face any challenges this is my favorite feature because it can directly detect our pronunciation."

All answers from the participants above can be grouped into two parts. Three answers from participants complained about the same thing. Where there are challenges in the speech recognition feature, sometimes the results are not clear. Whether it's from the user's factors or the application features not functioning properly. This is also a note for the founder of the Elsa Speak application to organize to develop features so that they can be used optimally, especially in learning foreign languages.

4. Improving pronunciation skills before and after using this application while using the speech recognition feature. In this question, participants will explain whether there are any differences in their pronunciation skills when using the speech recognition feature in ELSA Speak.

R said:
"The results of improving my pronunciation skills don't seem to be there yet. Maybe it's because I'm still using this application for a short time."

T said:
"For me, there is a significant improvement. Both from pronunciation skills and vocabulary. Apart from using the speech recognition feature which is helpful even though there are several challenges, other features in this application also really helped me in improving these skills."

R said:
"I felt there was a change after using various features in Elsa's speech, especially speech recognition. Where before I was still lacking and missed pronunciation."

N said:
"Yes, there are changes that I feel after using the features in this application. My pronunciation improved."

NS said:
"Of course, there is. The changes are quite significant too. Initially because of the difference in the way we pronounce our language from the foreign language we are learning. But because of this feature, my pronunciation is slowly becoming much more similar to native speakers."

Of the explanations from the sources above, 4 out of 5 answers gave a positive response. Where the use of the speech recognition feature in Elsa speaks has resulted in changes in their speaking skills, they explained that their skills tend to improve and be better than before. Not only pronunciation skills but also vocabulary has also greatly improved. Because of this feature many are introduced, and trained to improve new vocabulary. Meanwhile, there was only one participant who felt there had been no change in his pronunciation skills. Accent because he hasn't been using the Elsa Speak application for a long time.

1.2 DISCUSSION

The rapid advancements in Artificial Intelligence (AI) have opened up new possibilities for enhancing various aspects of human life, including communication skills. One promising area is the integration of speech recognition technology into AI systems, which can significantly impact language learning and interpersonal interactions (Hohenstein et al., 2023).

Speech recognition technology has made significant strides in recent years, enabling AI-powered systems to accurately transcribe and interpret spoken language (Chen et al., 2022). This capability has profound implications for language learning, as it allows students to practice their speaking and listening skills with immediate feedback and correction. By providing real-time analysis and feedback on pronunciation, intonation, and grammar, speech recognition can help learners identify and rectify their weaknesses, ultimately improving their overall communication proficiency.

In the context of English language education, the integration of speech recognition features in applications like ELSA Speak has the potential to greatly impact the development of learners' communication skills. Existing research has highlighted the benefits of technology-enhanced language learning for English language learners (ells) (shaji & nagaraj, 2020). Computer-assisted language learning not only provides increased speaking opportunities, but also helps to address the challenge of student reticence that is commonly observed in English as a second language (ESL) classrooms.

The incorporation of speech recognition capabilities in ELSA Speak further enhances the interactive and engaging nature of the learning experience, which can in turn improve the effectiveness of English language skill development.

The ELSA Speak app is a popular language learning tool that utilizes speech recognition technology to provide real-time feedback on pronunciation. However, the application of speech recognition in this context poses several challenges that must be addressed to ensure effective pronunciation practice.

One of the primary challenges is the variability of non-native speech. As language learners, users of the ELSA Speak app may exhibit diverse pronunciation patterns, accents, and articulatory characteristics that can be challenging for the speech recognition system to accurately process. This challenge is further exacerbated when dealing with student's that has difficult to articulate something clearly. So that is make challenge to optimize this future function.

CONCLUSION

This research aims to find out students' opinions regarding the AI-free speech recognition feature in the ELSA Speak application. Participants consisted of 6th-semester students majoring in English education at STAIN Majane. Based on the results and discussion, although it has disadvantages such as not having to have internet access, subscription features, or features that sometimes don't work. Researchers concluded that the AI-based speech recognition feature in the ELSA Speak application received a positive opinion from participants or students because it had been proven to have helped improve their pronunciation skills. Apart from that, it can add to the vocabulary of the Arabic language that students are studying.

Plus, the application display is very pleasant and easy to use. So that students are not pressured into playing the application. Let alone the speech recognition feature

contains training in pronunciation skills which helps in learning foreign languages. Researchers found that students' perceptions of the AI-based recognition feature in the ELSA Speak application for pronunciation skills proved positive.

ACKNOWLEDGMENT

Thank God, the researcher would like to express his gratitude to Allah SWT for all His blessings, grace, and gifts that have given the researcher knowledge, experience, strength, patience, and opportunities so that he is able to complete this research. However, the researcher realizes that without the help and support of various parties, the preparation of this research cannot run well. So on this occasion, the researcher would like to express his deepest gratitude to:

1. The researcher's parents who always provide prayers and support
2. Prof. Dr. Wasilah Sahabuddin, ST., MT. Rector of STAIN Majene.
3. Achmad Taqlidul Chair Fachruddin, S.Pd., M.Pd. the Head of English Program of Education.
4. Rina Marliana, S.S., M.Hum. our lectures and also as the first author and supervisor of this research. Thank you for your support, advice and patience in guiding the completing of this research,
5. All of our lovely friends Rahmi and Sariana who always support us. Thanks a lot
6. Subject of the research, English program education. The researchers are very grateful to them, because they have prepared time to help make this research a success
7. Then to the researchers. Thank you very much for your struggle, affection, gratitude, and patience and for being such a great person until now. We succeeded, and it was amazing. The researcher realized that this thesis could be better. The mistake was the researcher's mistake. Therefore, constructive criticism and suggestions would be highly appreciated. May all our efforts always be blessed by Allah SWT. Amen, O Lord of the Universe.

REFERENCES

Ames, H., Glenton, C., & Lewin, S. (2019). Purposive sampling in a qualitative evidence synthesis: A worked example from a synthesis on parental perceptions of vaccination communication. BMC Medical Research Methodology, 19(1), 26.

Braun, V., & Clarke, V. (2021). Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. Counseling and Psychotherapy Research, 21(1), 37-47.

Chen, X., & Li, Y. (2023). Personalized learning pathways in AI-driven language applications: Effectiveness and user satisfaction. Artificial Intelligence in Education, 4(2), 145-162.

Chen, J.; Lai, P.; Chan, A.; Man, V.; Chan, C.-H. AI-Assisted Enhancement of Student Presentation Skills: Challenges and Opportunities. *Sustainability* **2023**, *15*, 196. https://doi.org/10.3390/su15010196

Dai, T. H., Pham, T. M., Le, T. A., & Le, H. M. (2020). ELSA Speak: An AI-based app for pronunciation training.

Handoko, F., Ardiasmo, G., & Solehudin, M. (2021). Students' perceptions of an AI-based pronunciation training app: A case study of ELSA Speak. CALL-EJ, 22(1), 51-67

Hohenstein, J., Kizilcec, R.F., DiFranzo, D. *et al.* Artificial intelligence in communication impacts language and social relationships. *Sci Rep* **13**, 5487 (2023). https://doi.org/10.1038/s41598-023-30938-9

Korstjens, I., & Moser, A. (2021). Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. European Journal of General Practice, 24(1), 120-124.

Levis, J., & Muller Levis, G. (2018). Teaching High-Value Pronunciation Features: Priorities, Strategies, and Syllabi. CATESOL Journal, 30(1), 145-162.

Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile app: English speakers' user experience. The EUROCALL Review, 23(1), 38-56.

Quintero, L. M., Saldaña, A., & Quintero, L. E. (2019). Enhancing pronunciation through an intelligent mobile application: An experience with Venezuelan EFL learners.

Roth, W. M., & von Unger, H. (2023). Current perspectives on research ethics in qualitative research. Forum Qualitative Sozialforschung/Forum: Qualitative Social Research, 19(3).

Serra, F. F., Araújo, A. B., & Guerreiro, A. S. (2023). Purposive sampling as a tool for insider research in professional sport. Qualitative Research in Sport, Exercise, and Health, 15(2), 297-312.

Shaji, S., & Nagaraj, P. (2020). Integration of Technology in English Language Classrooms: A Research Review. *Shanlax International Journal of English*, *9*(S1-i1-Dec), 26–29. https://doi.org/10.34293/english.v9is1-dec2020.3608

Sharma, S., Chengalvarayan, R., & Bellur, A. (2020). Recent advances in speech recognition using deep learning methods. International Journal of Speech Technology, 23(4), 679-694.

Smith, J. A., & Nobles, M. R. (2020). Revisiting the use of qualitative research in criminology and criminal justice. Journal of Criminal Justice Education, 31(4), 503-524.

Xu, J., Jennings, M., & Young, R. (2017). A comparative study of AI-based pronunciation training for ESL learners. The EuroCALL Review, 25(2), 3-16.