

Voice Typing to Enhance English Intermediate Level Learners' Speaking Proficiency: A Case Study

Alaa Alnajashi¹

¹ King Abdullaziz University, Saudi Arabia

Correspondence: Alaa Alnajashi, King Abdullaziz University, Saudi Arabia.

Received: February 25, 2024 Accepted: April 23, 2024 Online Published: May 20, 2024

doi:10.5539/ijel.v14n3p67 URL: <https://doi.org/10.5539/ijel.v14n3p67>

Abstract

This mixed-methods case study explores the effectiveness of Google Docs' voice typing feature in enhancing the English oral skills of nine university students at the intermediate level. These university students were selected to utilize voice typing for their end-of-term oral presentations, engaging in two practice sessions each week for six weeks period. Training was provided to ensure proficiency in the use of voice typing. To gauge progress, a qualitative analysis of the data was performed, focusing on any advancements in speaking capabilities, with the students' final speaking project scores serving as the primary metric for evaluation. Evaluation of the final speaking project was conducted by two teachers, and to validate the assessment process, inter-rater reliability tests were implemented. In addition to the quantitative assessment, qualitative data were gathered through interviews, wherein students conveyed their initial challenges with the voice typing feature, particularly concerning speech recognition inaccuracies. Despite these initial hurdles, participants reported a smoother experience as they became more accustomed to the tool. The study underscores the educational benefits of voice typing and advocates for continued investigation into this promising field, particularly given the rise of Artificial Intelligence technologies adept at accommodating diverse accents and proficiency levels.

Keywords: Voice typing, speaking, technology

1. Introduction

Mastery of speaking is a critical component in the field of foreign language education and is essential for acquiring proficiency in a new language (Ohta, 2005). According to Mitra (2003) speech to text software, can improve English pronunciation for students without any required intervention from the teacher. However, these systems are not specifically designed for non-native speakers, as they were originally created to assist in the text-to-speech transformation process. Voice typing can be used by English learners to practice until they can hypothetically perfect their pronunciation. The accuracy of this hypothesis remains uncertain, as there is limited research in this area. This study aims to test this hypothesis and answer the following research questions:

- 1) How effective is voice typing in improving students' speaking proficiency levels?
- 2) What are students' experiences with using voice typing to improve their speaking proficiency?

2. Literature Review

Speech recognition, or automatic speech recognition (ASR), converts spoken language into text and is distinct from voice recognition, which identifies individual users' voices. Additionally, the technology for voice typing, also known as speech-to-text (STT), has made significant strides in accuracy and functionality, with applications across various fields, including assistance for those with disabilities, language learning, and vehicle control (Negoita et al., 2021).

Speech recognition systems have been effectively used for learning spoken English, with Hidden Markov Models (HMM) achieving greater than 90% accuracy across age groups (Jiang, 2022). Additionally, speech recognition software has been shown to effectively assess oral proficiency in English learners, enhancing students' ability to evaluate and improve their speaking skills. However, its effectiveness is contingent upon speaker-dependent technology, necessitating individualized training to address the diverse needs of learners (Coniam, 2013). Further, intelligent voice systems have been recognized for their role in invigorating students' enthusiasm, interest, and self-confidence, which in turn fosters greater interaction and participation within classroom settings (Wang, 2022).

Media technologies such as Vocaroo, VoiceThread, and text-to-speech (TTS) platforms have demonstrated potential in bolstering oral proficiency and autonomy. These advancements are particularly effective when coupled with instructor feedback, which remains a crucial element in the development of speaking abilities (Kim, 2018). VoiceThread, in particular, offers learners valuable opportunities for independent planning, rehearsal, and controlled production, which are vital components of oral proficiency development (Dugartsyrenova & Sardegna, 2016; Nguyen & Yukawa, 2022).

Moreover, the integration of voice typing within VoiceThread has been suggested to further enhance speaking proficiency and pronunciation accuracy (Lee, 2019). Voice typing, as a form of interaction with technology, may potentially contribute to enhancing learners' speaking proficiency, but further research is needed for a conclusive understanding. In fact, the use of text-to-speech (TTS) technology in learning English has been explored in various studies (Shadiev et al., 2017 ; Bione, 2016). However, speech to text didn't receive similar attention and the research on the effects of voice typing on speaking proficiency is somewhat limited and more research is needed in this area.

3. Methodology

This study utilized a mixed-methods approach to understand how the use of the free voice typing feature in Google documents can enhance students' speaking proficiency levels. Nine students were recruited for this study. These students were foundation-year students at an anonymous university with proficiency levels at B1 according to the Common European Framework. Phase one includes students' actual use of technology over six weeks, as will be discussed below, and phase two includes one-to-one interviews with the nine students to explore their experiences.

3.1 Phase One

Initially, students were instructed to prepare a written script for their final speaking project prior to their final presentation. Then, they were required to use a shared Google document for voice-typing their presentation and were encouraged to practice voice typing twice a week over a six-week period (see screenshot below—Figure 1). The researcher monitored the progress throughout the six weeks, and the collected data were analyzed qualitatively using NVivo software. This analysis aimed to observe the progress and identify areas where improvement plateaued. Their final speaking marks served as a benchmark to gauge their success, as they were rated by two independent teachers. Then, an inter-rater reliability test was conducted to validate the accuracy of the benchmarks.

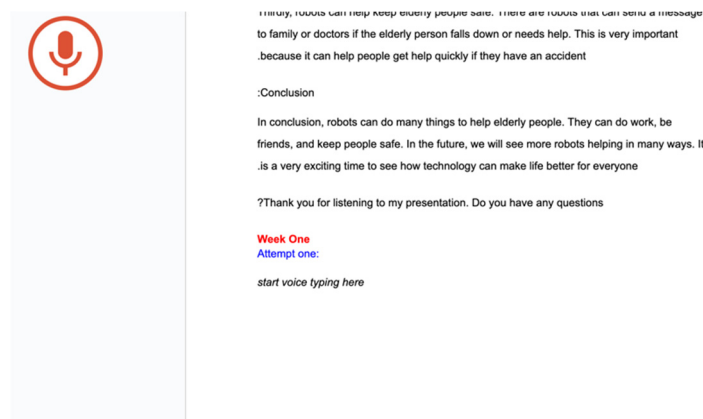


Figure 1. Voice Typing in Google Doc

3.2 Phase Two

In this phase, all participating students were invited to partake in one-on-one interviews to discuss their perspectives on voice typing and their likelihood of utilizing it in the future. The data from the interview were automatically audio-transcribed using Windows 10's audio-to-text feature. Then, after reading the transcripts and checking their accuracy, the files were uploaded into NVivo and prepared for analysis. I employed inductive coding for my thematic analysis, where I was trying to fit the codes under emerging themes (Saldaña, 2013).

4. Results

Initially, upon completion of the students' practice phase, the Google documents for the nine students were

uploaded into NVivo and prepared for coding. Each of the nine participants' data was tabulated according to the number of practices, words that the computer didn't recognize, and samples of students' verbatim responses in the interviews. The analysis aimed to assess the progress of the students and also to evaluate their experiences, as will be shown in the coming section.

4.1 Observing Students' Progress

To observe student progress, I investigated the data for each of the nine participants, starting from reading the original text to reaching the final practice before the final presentations. Below is a sample of how the data was analyzed for each participant (see Tables 1 and 2):

Table 1. Sample of the analysis of the first student data

Week	1 st week	2 nd week	3 rd week	4 th week	5 th week	6 th week
Number of practices	X2	X2	X2	X2	X2	X2
Words that the computer didn't pick	<u>5 words</u> Computer Science Robot People Hobby	<u>3 words</u> Computer Robot People	<u>3 words</u> Computer Robot People	<u>3 words</u> Computer Robot People	<u>2 words</u> Robot People	<u>1 word</u> People
Observation from the researcher	Most of the text seems intelligible, except for these errors, and there is progress in terms of the intelligible words recognized by the computer.					
Students' experience	The student mentioned that she really "enjoyed" the experience and felt more "confident" in the final presentation as the computer recognized most of her speech.					

The bar graph (Figure 2) shows the number of words the computer didn't recognize over the practice period:

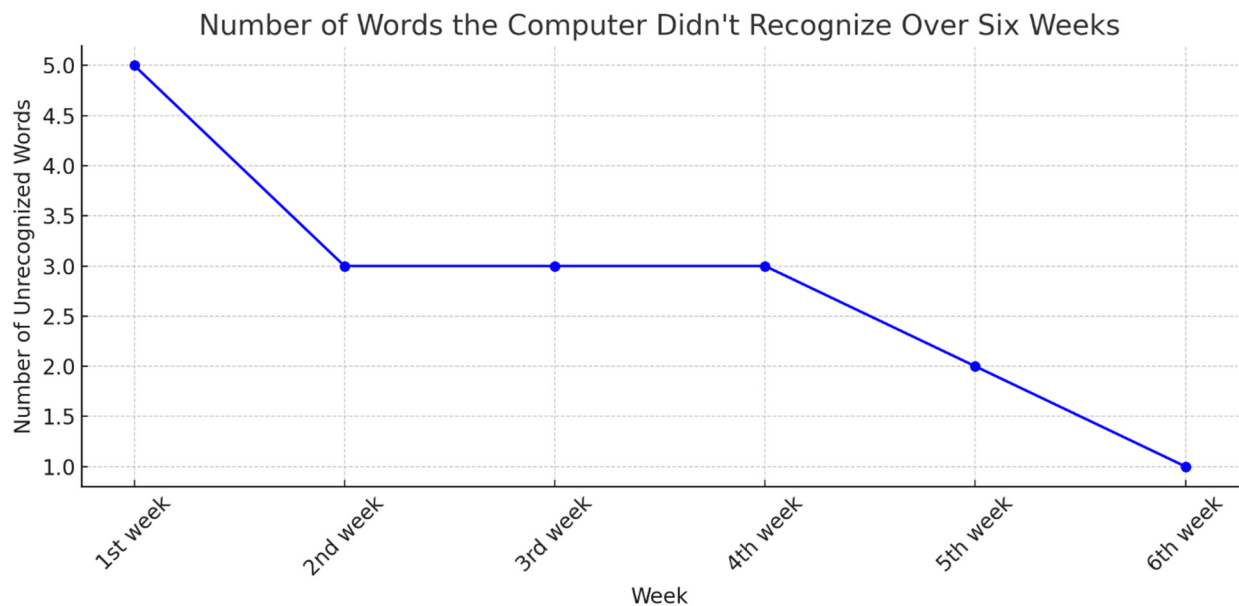


Figure 2. Sample of student progress

Table 1. Sample of the analysis of the second student data

Week	1 st week	2 nd week	3 rd week	4 th week	5 th week	6 th week
Number of practices	X2	X2	X2	X2	X2	X2
Words that the computer didn't pick	10 words Computer Chip Right Regular Obvious Science Psychology Sophisticated Feel Sponsor	11 words Computer Chip Right Regular Obvious Science Psychology Sophisticated Feel Sponsor Owner	11 words Computer Chip Right Regular Obvious Science Psychology Sophisticated Feel Sponsor Owner	7 words Computer Chip Right Regular Science Psychology Sponsor	5 words Computer Chip Right Regular Science Psychology Sponsor	3 words Computer chip Sponsor
Observation from the researcher	At the onset, the student seemed to reach a plateau, and then there was an improvement					
Students' experience	"It was kind of frustrating as I kept trying and trying, and I really wanted to give up."					

The bar graph (Figure 3) shows the number of words the computer didn't recognize over the practice period:

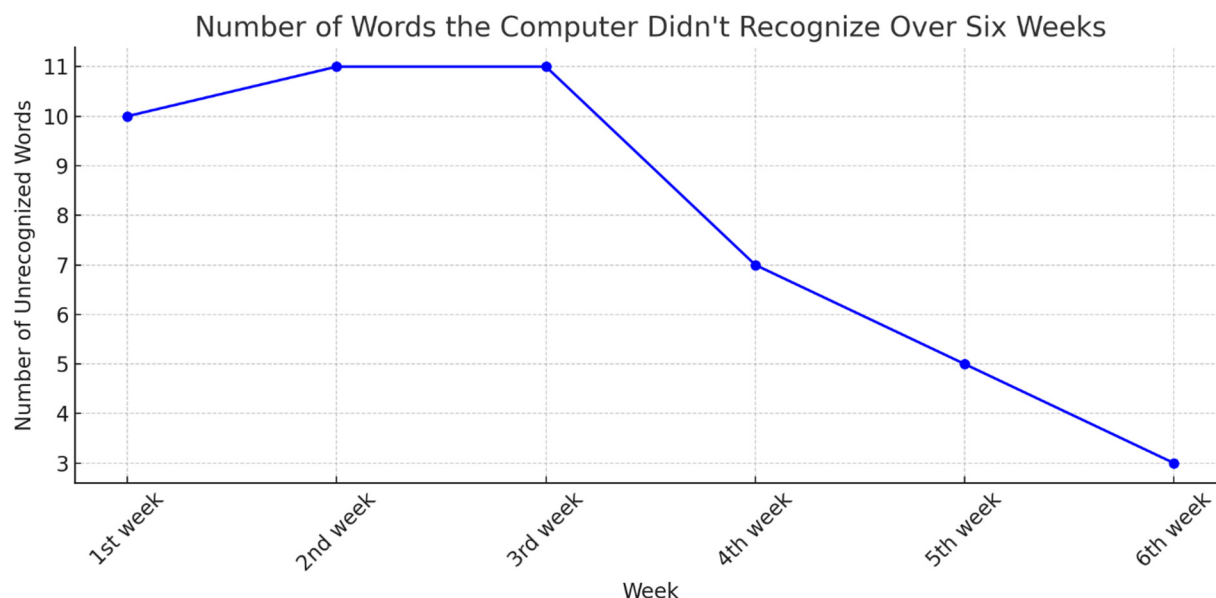


Figure 3. Participant 2 progress

As for the overall results, Table 3 below presents the data on the total number of practices for each participant, along with the final scores assigned by the two teachers. The figure below (Figure 4) offers a word cloud analysis of the list of words that the computer failed to recognize.

Table 2. Students' overall results

Student Serial No.	Number of words in students' presentations	Number of practices	Words that the computer couldn't pick	Final score in the presentation out of 15 by Teacher 1	Final score in the presentation out of 15 by Teacher 2
1	189	12	Openness	14	15
2	200	12	Robot	15	15
3	205	12	Computer	15	15
4	200	12	Non	15	15
5	199	10	People, PlayStation	13	14
6	200	12	Topic	15	15
7	200	12	Artificial	15	15
8	200	12	People	15	15
9	200	12	Computer Science, People	15	15

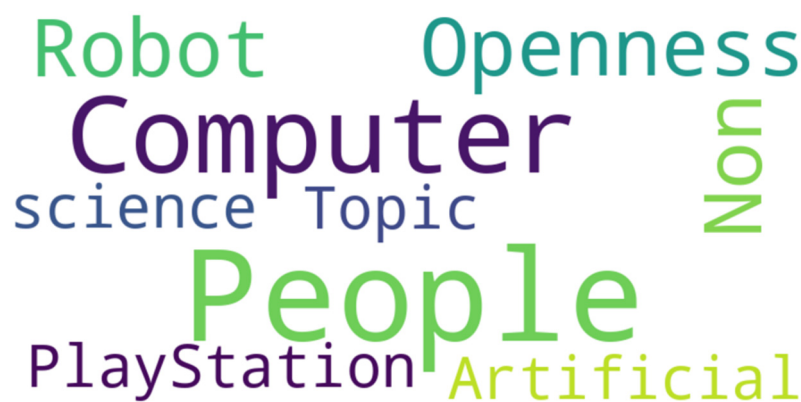


Figure 4. Word cloud of the words that the computer didn't recognize

Additionally, interrater reliability between the teachers to gauge the agreement between teachers' evaluation was calculated. To calculate the inter-rater reliability (IRR) between teachers, I run the test using Python.

The inter-rater reliability test between the two teachers, using Cohen's kappa, yields a score of 0.25. This score indicates a fair level of agreement beyond chance between the two teachers' ratings (Haslwanter, 2016). Cohen's kappa scores can be interpreted as follows:

- Less than 0: Less than chance agreement
- 0.01–0.20: Slight agreement
- 0.21–0.40: Fair agreement
- 0.41–0.60: Moderate agreement
- 0.61–0.80: Substantial agreement
- 0.81–0.99: Almost perfect agreement

Therefore, in this context, the teachers are in fair agreement regarding the final scores for the presentations. Additionally, there was a noticeable relation between the number of attempts and students final score by the two raters.

4.2 Students' Reflections in Their Experiences with Voice Typing

Upon completion of this phase, audio recordings were transcribed using Windows 10 (voice-to-text service). Subsequently, the files/texts were uploaded into NVivo, and data were coded to identify themes among the data. Five main themes emerged:

1) Ease of use:

Students expressed that using voice typing is very intuitive and doesn't require any training and that they can use it from their phones or computer .

2) Cost-free use

Many students expressed that what they really liked about this feature is that it is cost free and gives them a lot of training for free.

3) Technical difficulties

Some students said that the microphone freezes a lot and they were not sure how to fix it.

4) Lack of motivation

Few students felt that the experience of using voice typing is frustrating and they don't really feel motivated to learn or use English language in their real life.

5) Increased self-confidence:

Three students mentioned that the use of voice typing helped them improve their self-confidence, as they felt that if the computer could understand them, then they really spoke good English.

5. Discussion

The study aims to assess the usability of a simple, free feature—voice typing—to enhance students' speaking proficiency by exploring how effective this feature is in improving students' speaking proficiency levels through constant practice. As can be seen from the results section in Table 3, students did indeed improve their pronunciation, and the number of words recognized by the computer increased every week. This strongly suggests that students are enhancing their pronunciation and overall speaking proficiency with continuous practice. This also supports the literature that posits continuous practice improves students' speaking proficiency (Buckingham & Alpaslan, 2017; Qi et al., 2020). Additionally, Qi et al. (2020) stated that combining listening to materials with practicing them in synchronous communication can have a significant impact on students' overall speaking proficiency.

The study also suggested that there were certain words as shown in the word cloud in the result section (Figure 4) that the computer couldn't understand the students' pronunciations at all. Looking at these words and trying to find a common pattern in relation to pronunciation shows that most words contain the letter "p" or "r" two sounds that usually mispronounced by Arab learners. The pronunciation of the letter "p" by Arab learners of English can be challenging due to the absence of this phoneme in the Arabic phonetic inventory. Arabic speakers might substitute it with the phoneme "b" because Arabic does not have a direct equivalent to the English "p" sound. This substitution is a common phonological challenge faced by Arab learners when learning English or other languages that use the "p" sound. This difficulty is part of a broader set of pronunciation challenges and is influenced by the phonological and phonetic differences between Arabic and English (Hammami et al., 2020).

Additionally, research on Arabic speech sound errors in children, including the pronunciation of the letter "r", demonstrates that Arab learners may struggle with this sound when it appears at the beginning, middle, or end of words. Techniques like using Mel Frequency Cepstral Coefficients (MFCC) features and training probabilistic classifiers have been explored to classify spoken words and address pronunciation errors effectively, achieving classification accuracy rates of 71.75%, 77.20%, and 74.06% on average for words containing the letter "r" in different positions (Hammami et al., 2020). This is evident in our data, and it seems that this issue is related to the challenges faced by these learners rather than to the computer's inability to comprehend their speech. Additionally, our data showed a surge in students' performance as the number of practice sessions increased, as depicted in the Figure (2, 3) above. This supports the findings of Qi et al. (2020) and Zen et al. (2022), who advocated for additional practice to enhance learners' speaking proficiency.

Looking at the interrater reliability tests between the two teachers in evaluating overall speaking proficiency, the results showed that the two teachers agreed on their evaluation of the students' speaking proficiency. Future studies could, of course, conduct pre-speaking proficiency tests and post-tests to gain a holistic understanding of how voice typing feature improves the learners' speaking proficiency.

In relation to students' experiences with using voice typing, the students mentioned that the feature is very intuitive and easy to use, and according to Chapelle and Sauro (2019), it is extremely important to consider the ease of use of a certain technology when recommending it. Hence, I can suggest that teachers might recommend this simple feature to learners, where they can use it and benefit from it. Additionally, many students mentioned that the fact

this feature is free of cost is another main reason why they potentially might continue to use it. However, some students said that the microphone freezes a lot, and they were not sure how to fix it. In fact, technical difficulties are a given thing with the use of technology. In fact, technical difficulties can lead to students' frustration, as suggested by Runstova et al. (2020), who explored technical difficulties in general and how they lead to students' frustration. Additionally, a few students felt that the experience of using voice typing is frustrating, and they don't really feel motivated to learn or use the English language in their real life. This is related to the general lack of motivation, which is essential in any learning experience, especially in learning English (Alizadeh, 2016). However, some students mentioned that voice typing helped them in improving their speaking proficiency level. This is in line with Leeming and Harris (2022), who mentioned that speaking practice and achieving satisfying results after the practice leads to a higher level of self-confidence.

Overall, the study highlighted the potential use of a free feature in Google Docs, namely voice typing, to enhance learners' speaking proficiency. This study commenced in 2022, and during that time, there were very limited text-to-speech free software options available for voice typing. However, artificial intelligence was introduced to public users, and by 2023, the use of artificial intelligence tools became mainstream, with many affordable speaking applications like Loora and ELSA. All of these applications provide instant feedback on speaking, and future research should focus on this promising area.

6. Conclusion

This mixed method study investigated the effectiveness of voice typing in enhancing English speaking proficiency among students, finding that consistent practice improved pronunciation and word recognition by computers, supporting previous research that continuous practice boosts speaking proficiency as well as students' self confidence. The study revealed difficulties in pronouncing the "p" and "r" sounds among the sample of this study, which aligns with known pronunciation challenges stemming from differences between Arabic and English phonetics. Despite technical issues and varying levels of motivation, the voice typing feature was found to be intuitive and beneficial, especially as a free tool.

This study presents several pedagogical implications: teachers can guide students to this free, easy-to-use feature to enhance their speaking proficiency. The study highlights the importance of continuous practice in improving learners' speaking proficiency and emphasizes the crucial role of teachers in raising awareness among learners about the significance of practicing the language to achieve an acceptable level of speaking proficiency. Learners' motivation plays a pivotal role in the success or failure of utilizing certain technological applications. Speaking practice, along with the use of voice typing, can increase students' overall self-confidence.

This mixed-methods case study is insightful but has several limitations. The timeline extended beyond initial estimates because the study's design necessitated waiting until after a six-week period and the semester's end for a speaking exam to gather comprehensive data. Subsequent interviews were delayed to the next academic year due to the summer vacation. Future research might benefit from a strategy that involves collecting data over a series of shorter intervals. Furthermore, the study's participant pool was homogenous, consisting of students at the same proficiency level and engaging with similar subject matter. Future studies could explore the efficacy of voice typing across a broader spectrum of proficiency levels and a wider array of speaking topics. Additionally, as highlighted in the discussion, the relevance of this feature may be diminishing in light of the rapid advancements in artificial intelligence and voice recognition technologies, which open up new and more promising avenues in the field. The research also notes the growing role of artificial intelligence and speech recognition technologies in language learning, suggesting that future studies should focus on these areas.

Acknowledgments

I acknowledge the participants who participated in this study.

Authors' contributions

Not applicable

Funding

Not applicable

Competing interests

Not applicable

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Canadian Center of Science and Education.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

- Alizadeh, M. (2016). The Impact of Motivation on English Language Learning. *International Journal of Research*, 1, 11–15.
- Bonces, J. (2020). *Implementing Text-to-Speech Technology as a Means of Enhancing L2 Reading Fluency*. <https://doi.org/10.37467/gka-edurev.v1.2401>
- Buckingham, L., & Alpaslan, R. (2017). Promoting speaking proficiency and willingness to communicate in Turkish young learners of English through asynchronous computer-mediated practice. *System*, 65, 25–37. <https://doi.org/10.1016/j.system.2016.12.016>
- Chapelle, C. A., & Sauro, S. (2019). *The handbook of technology and second language teaching and learning*. Wiley-Blackwell, Oxford.
- Coniam, D. (2013). The Use of Speech Recognition Software as an English Language Oral Assessment Instrument: An Exploratory Study. *The CALICO Journal*, 15, 7–23. <https://doi.org/10.1558/cj.v15i4.7-23>
- Dugartsyrenova, V., & Sardegna, V. (2016). Developing oral proficiency with Voice Thread: Learners' strategic uses and views. *ReCALL*, 29, 59–79. <https://doi.org/10.1017/S0958344016000161>
- Hammami, N., Lawal, I., Bedda, M., & Farah, N. (2020). Recognition of Arabic speech sound error in children. *International Journal of Speech Technology*, 23, 705–711. <https://doi.org/10.1007/s10772-020-09746-3>
- Haslwanter, T. (2016). *An Introduction to Statistics with Python: With Applications in the Life Sciences*. Springer. <https://doi.org/10.1007/978-3-319-28316-6>
- Jiang, M. (2022). Evaluation of spoken English self-study system considering speech knowledge recognition algorithm. *Applied Mathematics and Nonlinear Sciences*. <https://doi.org/10.2478/amns.2021.2.00235>
- Kim, S. (2018). Exploring Media Literacy: Enhancing English Oral Proficiency and Autonomy Using Media Technology. *Studies in English Education*, 23(2). <https://doi.org/10.22275/SEE.23.2.03>
- Lee, L. (2019). *An Exploration of Using Audio Recorded Speaking Tasks With VoiceThread for L2 Oral Proficiency Development*. Assessing the Effectiveness of Virtual Technologies in Foreign and Second Language Instruction. <https://doi.org/10.4018/978-1-5225-7286-2.ch006>
- Leeming, P., & Harris, J. (2022). *Measuring speaking proficiency growth in the language classroom: An investigation of practical approaches for teachers*. Language Teaching Research. <https://doi.org/10.1177/13621688221130856>
- Mitra, S., Tooley, J., Inamdar, P., & Dixon, P. (2003). Improving English Pronunciation: An Automated Instructional Approach. *Information Technologies and International Development*, 1, 75–84. <https://doi.org/10.1162/itid.2003.1.1.75>
- Negoita, A., Suci, G., Segarceanu, S., & Trufin, D. (2021). *Speech recognition systems*. eLearning and Software for Education. <https://doi.org/10.12753/2066-026X-21-095>
- Nguyen, T., & Yukawa, T. (2022). Online Course Design Using VoiceThread with TPACK Model to Enhance English Speaking Skills for Non-native Learners. *Int. J. Eng. Pedagog.*, 12. <https://doi.org/10.3991/ijep.v12i1.24407>
- Ohta, K. (2005). Language the challenges encountered by EFL postgraduate students from the teacher's perspectives: Interviews with seven experienced ESL/EFL teachers. *Journal of Language and Learning*, 3,

133–155.

- Qi, W., & Liu, Y. (2020). Fostering Learners' Language Proficiency in Oral Class: A Lesson Plan for Integrated Instruction of English Listening and Speaking Skills. *Studies in English Language Teaching*. <https://doi.org/10.22158/selt.v9n1p1>
- Rubtsova, A., Odinkaya, M., Krylova, E., & Smolskaia, N. (2020). Problems of Mastering and Using Digital Learning Technology in the Context of a Pandemic. *Knowledge in the Information Society*, 184, 324–337. https://doi.org/10.1007/978-3-030-65857-1_28
- Saldaña, J. (2013). *The coding manual for qualitative researchers*. Los Angeles, LA: Sage.
- Shadiev, R., Huang, Y., & Hwang, J. (2017). Investigating the effectiveness of speech-to-text recognition applications on learning performance, attention, and meditation. *Educational Technology Research and Development*, 65, 1239–1261. <https://doi.org/10.1007/s11423-017-9516-3>
- Wang, Y. (2022). *Quantitative Analysis of Intelligent Speech System Applied to English Speech Assessment*. Proceedings of the 2022 3rd International Conference on Modern Education and Information Management. Atlantis Press. https://doi.org/10.2991/978-94-6463-044-2_118
- Zhen, L., & Hashim, H. (2022). *The Usage of MALL in Learners' Readiness to Speak English*. Sustainability. <https://doi.org/10.3390/su142316227>

Copyrights

Copyright for this article is retained by the author, with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).