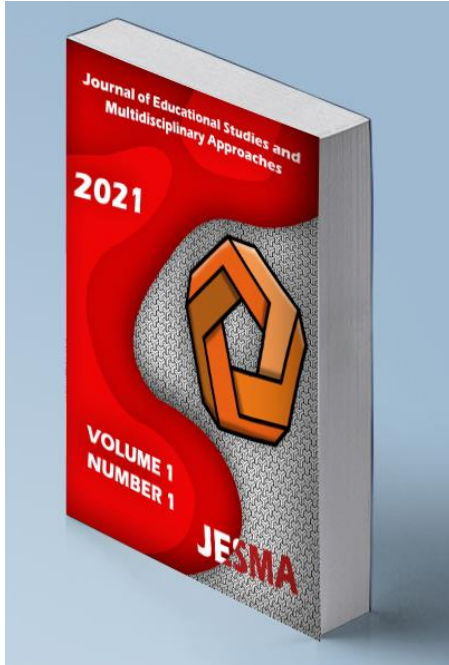




Journal of Educational Studies and Multidisciplinary Approaches (JESMA)

www.jesma.net

The Effectiveness of Online Text-To-Speech Tools in Improving EFL Students' Pronunciation¹

Fatma Önder²

To cite this article:

Önder, F. (2025). The Effectiveness of Online Text-To-Speech Tools in Improving EFL Students' Pronunciation. *Journal of Educational Studies and Multidisciplinary Approaches (JESMA)*, 5 (1), 29-47. <https://doi.org/10.51383/jesma.2025.116>

The Journal of Educational Studies and Multidisciplinary Approaches (JESMA) is an international scientific, high-quality open access, peer-reviewed scholarly journal that provides a comprehensive range of unique online-only journal submission services to academics, researchers, advanced doctoral students, and other professionals in their field. This journal publishes original research papers, theory-based empirical papers, review papers, case studies, conference reports, book reviews, essays, and relevant reports twice a year (March and September)

¹ This article was presented at 15th international METU ELT Convention, which was hosted at Middle East Technical University, Ankara, Türkiye, on May 10-11, 2024

² Lecturer, Konya Food and Agriculture University, Department of Foreign Languages, Konya, Türkiye, fatmaparlakci85@gmail.com

Fatma ÖNDER <https://orcid.org/0009-0006-83408344> 

ARTICLE INFORMATION

Original Research

DOI: 10.51383/jesma.2025.116

Received 20 December 2024

Revision 27 February 2025

Accepted 16 March 2025

ABSTRACT

This study attempts to investigate the effectiveness of an online text-to-speech application as a pedagogical tool in improving Turkish EFL students' pronunciation. Learners' pronunciation is expected to be accurate and intelligible to ensure clear and successful communication, allowing for meaningful interaction with native speakers and enhancing overall language proficiency. This study aimed to explore a free web-based text-to-speech tool implemented in one class during a 4-week pronunciation training. The study was conducted with 14 university students in prep classes at a private university in Türkiye. The current investigation employed an experimental design with a pre- and post-test. Quantitative data were collected through a read-aloud test for both the pre- and post-test. The investigation also aimed to shed light on university students' perceptions regarding text-to-speech training using the online application TTSReader. The results indicate that the learners have shown improvement in their pronunciation at the word level, and students have also expressed positive attitudes relating to the online text-to-speech application.

Keywords: Artificial intelligence, text-to-speech, speech-to-text, speech recognition, pronunciation



This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original authors and source are credited.

Introduction

A crucial component of language learning is communication. Accurate pronunciation and grammar are essential for effective communication, particularly in settings where English is used as a foreign language (EFL). As a fundamental aspect of language acquisition, pronunciation significantly influences learners' ability to comprehend and express themselves effectively. Therefore, improving pronunciation is crucial for gaining confidence and fluency in a second language; emphasizing the need to examine technology's role in facilitating this process. As technology continues to transform educational methods, its ability to improve pronunciation learning through AI-powered technologies deserves more research.

In educational contexts, the internet and other digital technologies have become increasingly beneficial. Incorporating communication and information technologies makes learning more effective than conventional educational methods. Similarly, artificial intelligence (AI)-based technologies offer multiple benefits to learners, especially in improving pronunciation. This highlights the expanding role of technology in language learning, especially through AI-powered tools such as Speech-to-Text Recognition and Text-to-Speech.

Vargas (2023) conducted a phenomenological study on teachers' perceptions of AI as an instructional tool in secondary English Language Arts (ELA) classrooms. The findings of this research revealed an overwhelming number of positive perceptions, mainly indicating the need for effective training for educators, leaders, and students to successfully integrate AI into education. These findings highlight the increasing need to provide educators with the necessary skills as this is crucial for successfully integrating AI-driven technologies such as Speech-to-Text and Text-to-Speech into language learning.

With the support of AI-driven technology, such as Speech-to-Text Recognition (STR), or Text-to-Speech (TTS), students can study and practice beyond the classroom at their own pace. Several studies have demonstrated that these technologies are invaluable resources for language instruction (Liu & Shadiev, 2023; McCrocklin, 2019; Zou et al., 2023). By facilitating self-paced learning, these tools provide opportunities for continuous practice, which emphasizes the need for further research on their impact on pronunciation development.

Shadiev et al. (2014) reviewed previous research on STR technology and its potential to enhance learning. They mentioned that STR technology was primarily used to support individuals with special needs, such as learning disabilities or physical impairments, as well as international students. Over time, a wider range of target consumers has been included in STR technology research. According to Liu & Shadiev (2023), several studies have been conducted on STR to assist with a variety of language skills. These include listening, speaking, writing, grammar, vocabulary, word recognition, communication, and pronunciation. Nevertheless, the use of technology, including TTS, in EFL courses specifically for improving pronunciation remains underexplored in previous studies.

A variety of techniques aimed at teaching L2 pronunciation have gained widespread acceptance and approval. Among these are shadowing and self-recording for error identification. While these approaches can help L2 students improve their pronunciation, most of them don't provide an opportunity for feedback on how they sound when they speak; and students are often unable to identify their pronunciation errors (McCrocklin, 2019). Consequently, to help them improve their L2 pronunciation, it may be necessary to provide precise and quick feedback (Saito & Lyster, 2012). Using technologies, including TTS and STR, students can receive instant feedback on their pronunciation using technologies, including TTS and STR, allowing them to make corrections and improvements quickly. Studies have indicated that these technologies are beneficial in improving learners' phonetic accuracy while minimizing pronunciation-related anxiety, as they create an encouraging environment for continuous practice and development. Therefore, AI-powered tools appear to be a valuable solution to the pronunciation challenges faced by language learners.

Problem statement

Pronunciation is an essential task for Turkish EFL (English as a Foreign Language) students due to a variety of related difficulties. Some frequent pronunciation errors made by Turkish students learning English are due to the characteristics of their mother tongue. For instance, according to Demirezen (2009), a very common pronunciation error occurs when Turkish English language learners pronounce the /ng/ sound combination, as /Ink/ instead of the correct pronunciation of /Inj/. It is important to note that poor pronunciation obscures and interferes with automatic control of pronunciation in the target language, as well as intelligibility, accuracy, and fluency (Demirezen, 2009). The absence of phonetic parallels in the native language, in particular, hinders the smooth acquisition of English word pronunciations. Furthermore, the short time allocated for pronunciation practice in educational contexts worsens the problem. As a result, Turkish EFL learners often struggle to effectively and accurately reproduce English words that lack direct counterparts in their mother tongue, hindering their overall proficiency in spoken English.

Chen (2011) states that most Taiwanese students struggle with communicating effectively and efficiently with individuals from all over the world. According to Kung (2017), many learners are unable to speak and express themselves successfully in English, as English education in Taiwan places a lower priority on oral training. A similar situation exists in the Turkish educational setting with EFL students. When students begin their education at university, the pronunciation errors they have acquired become fossilized, making these errors incredibly difficult to change (Selinker, 1972). However, Acton (1984) asserts that fossilized pronunciations can be reversed.

Purpose of the study

Given the challenges posed by fossilized pronunciation errors, particularly in the context of EFL learners, it is essential to explore possible solutions, such as the audio-articulation pronunciation instruction approach and emerging AI-powered tools. Since accurate pronunciation is essential for proper intonation and since good pronunciation is a necessary component of communicative ability, the audio-articulation pronunciation instruction approach may be very helpful in addressing this issue (Demirezen, 2009). However, technological advancements and the creation of AI-powered tools such as automated speech recognition or text-to-speech may also be useful in solving this pronunciation problem. A limited number of studies, however, have investigated the effectiveness of AI-driven TTS tools in enhancing learners' pronunciation. Therefore, the current study aims to investigate the efficacy of the free online text-to-speech program TTSReader, particularly, focusing on learners' pronunciation improvement at the word level. Consequently, the following set of research questions served as a guide for the current study:

1. Is the text-to-speech tool an effective way to improve EFL students' pronunciation?
3. Does the TTSReader-Text-to-Speech online play a role in improving English pronunciation accuracy?
4. What are students' attitudes towards the web-based TTSReader in their language learning?
5. Is there a noticeable improvement between the pre-test and post-test scores of the EFL students?

Literature review

Text-to-Speech

Text-to-speech (TTS) technology is a tool that transforms written text into spoken words. With read-aloud software, like text-to-speech, users may read along while listening to written text. It was around the 1980s that TTS technology began to take shape. As technology has advanced, educators and researchers have increasingly adopted text-to-speech, alongside related tools, not only to support children with reading disabilities but also to enhance language learning for a broader range of students (Wood et al., 2018). It can transform several forms of text input, including written documents, e-books, and other written information, into speech. Instead of having people read the information, TTS software

analyzes the text and generates a spoken version. Speech synthesis, to put it simply, is “the process of making the computer talk” (Handley, 2009). In simple terms, they are built using two modules. The first module converts the text into phonemes, while the second module translates those phonemes into voice (Handley, 2009).

This capability of TTS technology makes it particularly valuable in language learning contexts. Peters and Bell (2007) state that hearing a book or passage read aloud may be helpful for novice readers, both adults and children. Individuals studying a second language can also benefit, particularly if the software is advanced enough to provide accurate vocalization of the spoken language, potentially even in several dialects. Handley (2009) suggests that TTS technology can be utilized as a model for pronunciation in exercises emphasizing both segmental (practice of individual and combined phonemes) and suprasegmental (prosody and intonation) levels.

Google Text to Speech

There are numerous text-to-speech tools and services available, both commercial and open-source. Although we will not be reviewing every TTS tool in this study, we will highlight a few of the most well-known ones. One popular application is Google Text-to-Speech, which is part of the Google Cloud Platform and offers high-quality TTS voices. A broad spectrum of voices in many languages and dialects, is available through Google TTS. A wide range of voices is available for users to select from, meeting their needs and preferences.

Natural Reader

Natural Reader (www.naturalreaders.com) is also a popular TTS program that turns written material into spoken words. It is used to assist those who might have trouble comprehending or reading text because of visual impairments, learning difficulties, or they are searching for an alternative method to take in written information. Natural Reader comes with a variety of features and voices. Users can adjust the speech pace, voice type, and other parameters using Natural Reader to create a more customized listening experience.

TTSReader

TTSReader is a web-based application that provides TTS features. It enables users to convert written material into speech from a variety of sources, including webpages, documents, eBooks, and PDFs. Customization options available with TTSReader include the ability to change the voice, highlight text being read aloud, and adjust the speech pace. This functionality can help users follow along with the spoken text and improve their understanding. Students may improve their pronunciation and understanding by listening to the text read aloud. TTSReader offers both free and premium versions. Generally, the free version can read aloud and convert text to speech; however, some advanced features or additional voices may only be available in the premium version, which a subscription. This tool is user-friendly for pedagogical purposes, which is why we employed TTSReader in the current study during the training sessions.

Pronunciation

Considering TTSReader's potential to improve both pronunciation and comprehension, it is important to recognize the significance of pronunciation in language learning and its contribution to effective communication. Pronunciation refers to how a word or language is uttered, including the sounds, emphasis, rhythm, and intonation used when speaking words or phrases. According to Burgess & Spencer (2000), pronunciation in language learning involves the use of the target language's phonological components in speaking. Phonology encompasses both segmental and suprasegmental aspects of the target language's sound system, which includes theoretical knowledge (Burgess & Spencer, 2000). Exposing learners to the target language's sound system is essential in language

education as it sets the groundwork for correct pronunciation, fosters effective communication skills, and enhances listening comprehension.

It is crucial to consider how exposure to language input plays a key role in helping learners acquire all aspects of the language, including pronunciation. Language is primarily learned by individuals through the input they are exposed to (Krashen, 1982). To help students acquire all aspects of language—pronunciation, vocabulary, grammar, and discourse—educators should work to increase their exposure to the language of instruction and motivate them to broaden their areas of linguistic competency (Celce-Murcia et al., 2010).

While increasing exposure to language input is essential, it is also important to recognize that the goal of pronunciation instruction is not to make learners sound like native speakers but to help them communicate effectively. Instead, encouraging students to overcome barriers so that their pronunciation doesn't interfere with their communication skills is a more reasonable and achievable objective (Celce-Murcia et al., 2010). According to Celce-Murcia et al. (2010), when instructing pronunciation, the focus has mostly been on teaching students how to pronounce words correctly at the word level, whether they are working with single words or words in highly controlled and unnatural sentence structures. In response to the question, "How can pronunciation be taught as part of Communicative language teaching?", they offered a list of various pronunciation techniques. This list, according to them, is considered fairly comprehensive: listen and imitate, phonetic training, minimal-pair drills, contextualized minimal pairs, visual aids, tongue twisters, developmental approximation drills, practice of vowel shifts and stress shifts, recitation, reading aloud, and recordings of learners' production (Celce-Murcia et al., 2010). In addition to these established pronunciation techniques, it has been accepted that promoting active learner participation is important for attaining successful pronunciation outcomes.

As stated by Morley (1991), teaching pronunciation is more beneficial and enjoyable when students actively participate in their own learning rather than passively repeating exercises. Morley (1991) also adds that learner self-engagement should be developed at an early stage in the learning process. Educators and instructional materials might help students get involved (Morley, 1991). Today, when it comes to tools to enhance language teaching and learning, AI-driven technology is essential in creating innovative and successful solutions for educators and learners.

While active learner participation is critical in pronunciation instruction, it is also important to examine other external factors that influence L2 pronunciation, such as the quality of input and the learner's linguistic background. Several factors affect L2 pronunciation teaching. The quality and quantity of L2 input, particularly the amount of native target language input, might improve learners' pronunciation. Additionally, the learner's L1 background and phonetic variations between the L1 and L2 may create obstacles in perception and production (Piske, 2012; Piske et al., 2001). As researchers continue to explore factors influencing L2 pronunciation, technological tools like speech recognition and text-to-speech software have emerged as potential resources for improving pronunciation instruction.

Previous research on TTS Technology

In a recent review article, speech recognition technologies have received little attention in previous studies, while researchers who have used speech recognition tools in their studies have primarily focused on facilitating pronunciation skills (Liu & Shadiev, 2023). In a meta-analysis study, Wood et al. (2018) found that text-to-speech technology enhances reading comprehension among individuals with reading difficulties. They pointed out that there is an increasing number of studies into the effectiveness of text-to-speech and comparable read-aloud tools in improving reading comprehension. However, additional empirical research is required, as it remains unclear whether AI-based voice recognition and TTS software can help language learners improve their pronunciation. In addition to studies focusing on pronunciation and reading comprehension, other research has explored the use of AI tools in improving a variety of language skills.

Some researchers have focused on improving specific language skills, while others have addressed different aspects of language learning. For instance, Cavus and Ibrahim (2017) studied how to improve both listening and pronunciation, whereas Tsai (2019) used speech recognition technology to improve students' oral performance only. According to Cavus and Ibrahim (2017), students who used the

developed speech recognition application in the experimental group showed statistically significant improvements in their learning outcomes. The study, which was experimental in nature, focused on the potential of using an interactive speech recognition engine that works entirely on a mobile phone. These findings demonstrate that speech recognition technology may help language learners improve their speaking skills.

Building on this, additional studies have further supported the effectiveness of speech recognition technology in improving speaking proficiency. For example, Ahn & Lee (2016) discovered that students had positive attitudes toward using speech recognition technology for speaking exercises, suggesting that such tools not only improve proficiency but also engage students in the learning process. This implies that voice recognition software can help learners improve both their speaking skills and their motivation to practice.

Further expanding on the advantages of AI-driven tools for language learning, some studies have examined the impact of TTS technology on pronunciation improvement. In their pre- and post-test experimental study, Eksi and Yesilcinar (2016) explored the pedagogical potential of AI-based text-to-speech websites for self-study, examining how English language teacher trainees used these tools for pronunciation development. According to the pre- and post-test findings, text-to-speech websites can be a useful tool for self-study aimed at improving pronunciation. Additionally, the results of the reflection questionnaire used in their study showed that web-based text-to-speech tools were efficient and easy to use (Eksi & Yesilcinar, 2016).

In addition to studies focusing on pronunciation in English, research has also extended to other languages, such as Chinese, exploring the use of AI-driven technology in foreign language teaching. In a multiple-case study, Yeh (2014) investigated effective strategies, based on teachers' perspectives, for instructing and learning Chinese as a foreign language using various tools, such as TTS, SRT, and machine translation technologies. The research found that instructors' trust in technology, administrative support, ease of use, and accessibility all played a significant role in their adoption of these tools for teaching. On the other hand, students' social behaviour, motivation, learning attitude, and guidance from parents all had an impact on their capacity to successfully use technology tools when learning Chinese.

Similarly, research on pronunciation training for Arabic-speaking learners highlighted the potential of CAPT (Computer Assisted Pronunciation Teaching) to offer individualized learning experiences, free from peer pressure. Ehbara (2021) states in her research on the impact of CAPT training on Arabic-speaking young learners that the use of CAPT can assist students in working individually, at their own speed, and with fewer worries about peer criticism. Such applications can be employed as additional practice in combination with regular teaching settings.

Computer Assisted Pronunciation Teaching (CAPT)

CAPT is gaining attention from language educators due to its potential to enhance pronunciation skills through technological advancements. Hismanoglu (2006) highlighted that language instructors and experts are becoming more interested in CAPT because of the progress in computer technology. CAPT offers learners a private, relaxed setting where they can practice at their own pace and receive immediate feedback (Hismanoglu, 2006). In this context, the present study utilized a CAPT system to teach English pronunciation to Turkish-speaking students, offering them native English speaker input generated by TTS technology. Thus, the integration of CAPT systems, such as those using TTS technology, presents an innovative approach to improving pronunciation, particularly for language learners like Turkish-speaking students.

Methodology

This study aims to investigate whether a course incorporating TTS tools can help improve EFL students' pronunciation. A pre-test and post-test approach was used to examine the effectiveness of the web-based TTS application TTSReader. Data were collected using two quantitative instruments: a read-aloud test administered both as a pre-test and a post-test, and a questionnaire measuring students' beliefs,

attitudes, and perceptions of the effectiveness of the TTS tool. The 4-week pronunciation training with the TTS tool was also embedded in the course. Descriptive statistics were employed to analyze the data. For the read-aloud test, the number of correct answers was counted, and an overall score was calculated as a percentage.

Participants

The research study included 14 university students as participants in preparatory classes at a private university in Türkiye. A convenience sampling strategy was employed for this quantitative study as participants were willing and available to participate (Creswell & Creswell, 2018). The students' ages ranged from 18 to 24, with an equal distribution (7 female and 7 male students). Initially, there were 20 participants; however, six participants dropped out due to absences during training sessions and failure to complete the post-test. Consequently, the study was completed with 14 participants. The students had an A1 level of English and faced various pronunciation challenges. Students in preparatory classes take three different language courses. The main course focuses on grammar, while the other courses emphasize reading and writing skills. Lastly, the listening and speaking course develops students' abilities in listening, speaking, and pronunciation.

Instruments

A read-aloud test was utilized in the study as both the pre-test and post-test. The read-aloud test was generated using the AI-driven technology ChatGPT, which was instructed to create sentences containing 45 targeted words. Each sentence included one target word, focusing on specific pronunciation challenges such as consonant clusters and silent letters—voiced th, voiceless th, engma -ŋ, -ng, -ph, silent k, silent h, silent t, silent l, and silent b. In total, 45 targeted words were embedded in the read-aloud test.

The read-aloud test was recorded by the researcher, using a voice recorder app. The researcher then evaluated the recordings by providing correct/incorrect feedback on the targeted words. When the particular phoneme was pronounced correctly in the targeted word, it was considered correct. When the targeted phoneme was mispronounced, the word was evaluated by the researcher as incorrect. The test was evaluated by the researcher herself, as she was also the teacher of the participants.

Example: Targeted word: There /ðeər/

For instance, if the voiced th in the word *there* is not pronounced correctly, the pronunciation is considered incorrect.

To complement the pre- and post-test assessments, participants' perceptions of the TTS technology were also gathered to gain a more comprehensive understanding of their experiences with the tool. Participants' opinions and perceptions of the TTS technology were collected using a twelve-item questionnaire with two response options: *agree* and *disagree*. The questionnaire used in this study was adapted from Noviyanti (2020), who developed a questionnaire based on relevant literature. The reliability coefficient of the instrument was reported as 0.75, indicating moderate reliability. Since the students' proficiency levels were too low to fully comprehend the English version, the Turkish translation of the questionnaire was used.

Training

The study included a four-week pronunciation training period. Due to the short duration of the course, it was not feasible to address all segmental differences that posed challenges for the learners. Instead, the training focused on pronunciation problems at the word level. Specifically, the study emphasized the acquisition of consonant clusters and silent letters, as these features are either uncommon or challenging for native Turkish speakers to pronounce. We targeted these phonemes because Turkish students often struggle with pronouncing consonant clusters. Additionally, silent letter rules are not always evident to Turkish learners, as every letter in a Turkish word is pronounced.

The pronunciation training covered the phonemes displayed in *Table 1*.

Table 1

Phonemes, Letters, and Transcription Examples

Phoneme description	Letters	Transcription	Example
The voiced dental fricative	th	/ð/	this
The voiceless dental fricative	th	/θ/	three
Velar nasal phoneme (engma)	ng	/ŋ/	sing
Consonant digraph	ph	/f/	photo
Silent ⟨k⟩	k	—	knife
Silent ⟨h⟩	h	—	honest
Silent ⟨l⟩	l	—	could
Silent ⟨t⟩	t	—	listen
Silent ⟨b⟩	b	—	debt

Each pronunciation session lasted 40 minutes. ChatGPT generated new passages for every phoneme, incorporating the targeted phonemes into context-related passages derived from the students' curriculum. The university computer lab was used for pronunciation practice. The lab contained 20 PCs, each equipped with headphones with built-in microphones and an internet connection. This setup allowed each student to practice independently on a computer.

Procedure

A four-week training program was planned, with each session lasting 40 minutes. The training was divided into four stages. First, the teacher explained the pronunciation topic, showed the IPA phonetic symbols, demonstrated sound articulation with particular examples, and modeled the correct pronunciation.

Second, the teacher projected a passage generated by ChatGPT (containing words with the targeted phonemes) onto the board. Students listened to the paragraph, which was read aloud by an AI voice generator called TTSReader: Text-to-Speech (TTS), while following the text on the screen. The students had to pay attention to the specific phonemes or consonant clusters while listening. Dual coding was applied in this part, as students listened to the projected paragraph, they visually tracked the specific phoneme being taught in that session. TTSReader allows users to read along by highlighting the current sentence and instantly scrolling to keep it in view. This allows learners to follow the passage with their eyes while simultaneously listening. This approach can help learners read faster, comprehend more effectively, and focus on the pronunciation of the targeted words. Students listened to the paragraph twice, ensuring they heard the correct pronunciation of the words in context.

The third stage was the most active and intensive part of the training: shadowing practice. Students read the passage aloud, checking their pronunciation and assessing whether they pronounced the targeted words correctly. Learners could attempt as many times as desired to achieve the perfect pronunciation of the words. Peer evaluation was also incorporated at this stage. After each session, students recorded themselves while shadowing the passage, paying particular attention to the targeted words. They then sent their recordings via WhatsApp to an assigned partner, who listened to the recording and provided some feedback on their peer's pronunciation. When the pronunciation was correct, the students had to give positive feedback, and when they noticed mispronunciation of the words in the passage, they could give constructive feedback to help improve pronunciation. Finally, there was a group discussion on the difficulties and pronunciation errors they made. Students also received a printed document of the paragraphs, with the targeted words in bold, to help them identify which words and phonemes to focus on. An overview of the four-week training sessions is provided in *Table 2*.

Table 2
Training Sessions: Pronunciation training with TTSReader-Text-to- Speech (TTS)

Week	Pronunciation Training	Phonemes	AI-Driven Programs Used
1	The voiced dental fricative <th> The voiceless dental fricative <th>	/ð/, /θ/	ChatGPT (creating passages) TTSReader-Text-to-Speech (TTS) Online
2	Silent <k> Silent <h> Velar nasal phoneme <ng>	—, —, /ŋ/	
3	Consonant digraph <ph>	/f/	
4	Silent <l> Silent <t> Silent 	—, —, —	

Ethical Considerations

In this study, all rules stated within the scope of "Higher Education Institutions Scientific Research and Publication Ethics Directive" were followed. None of the actions stated under the title "Actions Against Scientific Research and Publication Ethics," which is the second part of the directive, were taken.

Ethical review board name: Konya Food and Agriculture University

Date of ethics review decision: 14/12/2023

Ethics assessment document issue number: 2023/06

Findings

We present the results using descriptive statistics in percentages due to the study's narrow focus and small number of participants. For the read-aloud test, the correct answers were counted, and an overall score was calculated. Table 3 shows an increase in students' pronunciation accuracy across all nine categories. Average percentages of correct answers were collected for each category, as shown in Table 3.

Table 3
Average accuracy percentages for all 9 categories

	Pre-test	Post-test
	%	%
Voiceless- th	27.14	67.14
Voiced -th	48.56	68.56
Engma -ng	42.84	87.16

Ph	59.98	81.42
Silent -k	60.00	95.74
Silent-h	38.58	64.30
Silent-l	15.70	68.56
Silent-t	60.00	78.56
Silent-b	8.58	48.56

In the voiceless th /θ/ category, the average correct answer percentage in the pre-test was 27.14%, **while** in the post-test, it rose to 67.14%. In the voiced th /ð/ category, the average percentage of correct pronunciation in the pre-test was 48.56%. The average percentage of correct answers in the post-test increased to 68.56%. In the **engma-ng** category, the average percentage of correct answers in the pre-test was 42.84%. In the post-test, the rate rose to 87.16%. The average percentage of correct answers in the Ph category, in the pre-test, was 59.98%. In the post-test, the rate rose to 81.42%. In the silent **-k** category, the average percentage of correct answers in the pre-test was 60%, **while** in the post-test, the rate increased to 95.74%. In the silent **-h** category, the average percentage of correct answers in the pre-test was 38.58%, **while** in the post-test, the rate increased to 64.30%. In the silent **-l** category, the average percentage of correct answers in the pre-test was 15.70%, **and** in the post-test, the rate increased to 68.56%. In the silent **-t** category, the average percentage of correct answers in the pre-test was 60%, **while** the rate increased to 78.56% in the post-test. In the silent **-b** category, the average percentage of correct pronunciation in the pre-test was 8.58%, while in the post-test the rate increased to 48.56%. The read-aloud pre- and post-test findings indicate that learners' pronunciation **accuracy** improved.

To gather participants' user perceptions and attitudes about the TTS technology used during training sessions, a questionnaire with a dual Likert scale is administered. The results of the participants' perceptions and attitudes about the TTS tool, TTSReader, are presented in *Table 4*.

Table 4

Distribution of answers to the 12-item evaluation questionnaire

Article no.		Agree		Disagree	
		n	%	n	%
1	I believe that the application has pedagogical value	13	92.9	1	7.1
2	I believe that this application can improve my pronunciation	14	100.0	0	0
3	I believe that this application gives opportunities to improve my pronunciation accuracy	14	100.0	0	0
4	I think that students need to use this application to increase their pronunciation accuracy	14	100.0	0	0
5	I believe that I cannot improve my pronunciation with this application.	0	0	14	100.0
6	I believe that this application cannot support independent learning.	2	14.3	12	85.7

7	I believe that this application can be time-consuming.	6	42.9	8	57.1
8	I find it difficult to use this application	0	0	14	100.0
9	I believe that this application can decrease my anxiety toward pronunciation learning	14	100.0	0	0
10	I believe that this application is not too effective for pronunciation learning	0	0	14	100.0
11	I believe that this application promotes pronunciation accuracy	14	100.0	0	0
12	I believe that this application promotes independent learning	14	100.0	0	0

Items 2, 3, 4, 9, 11, and 12 were answered by 100% of respondents with "I agree." 100 percent of the respondents answered "I do not agree" with articles 5, 8, and 10. Forty-two point nine percent (42.9%) of respondents said "I agree" with article 7. Based on the results of the first question, the majority of students believe that using the TTSReader application will help them pronounce words more accurately. The responses to questions 2, 3, and 4 indicate that every participant believes the application offers students the chance to improve their pronunciation skills. The responses to question 5 show that all participants disagree with the claim that the program cannot enhance students' pronunciation. All of the participants stated that this application can decrease anxiety toward pronunciation learning and that they believe it promotes independent learning. Question 8 demonstrates that learners generally do not have problems operating the TTSReader; however, almost half of the students stated that this application can be time-consuming. All the participants disagreed with question 10, suggesting that they believe in the effectiveness of the application for pronunciation learning, though the contents of question 10 are not specified here.

Discussion

Technology and English language instruction are closely linked, as digital tools and platforms facilitate interactive and immersive learning experiences that improve language proficiency, enhance communication skills, and provide access to a variety of learning materials. The primary goal of this study was to explore improving EFL students' pronunciation via TTS technology, specifically TTSReader, as a pedagogical tool. The purpose of using TTS technology was to determine whether it might help EFL learners accurately pronounce words in English with particular phonemes and consonant clusters.

A 4-week pronunciation training program using TTSReader was conducted with 14 EFL students at a private university in Türkiye. The pronunciation training focused on consonant clusters (voiced and voiceless -th, -ph, -ng) and silent letters (-k, -l, -t, -b, -h). Following the pronunciation training sessions, a dual Likert questionnaire was used to investigate participants' attitudes, perceptions, and beliefs regarding the TTSReader technology employed in the sessions. The TTSReader tool employed in this research is a free online text-to-speech platform. According to the responses of the participants, this tool was perceived as user-friendly and effective for pedagogical purposes.

The findings of this experimental study confirm earlier findings of several studies about the incorporation of technology to enhance language teaching success. The results of research conducted by Zou et al. (2023), for instance, confirm that AI speech evaluation programs were beneficial for

enhancing speaking skills, grammatical accuracy and range, pronunciation and oral rhythm, reading aloud skills, spontaneous speaking, and idea-organization.

In a similar vein, the findings of Noviyanti's (2020) study showed that students' pronunciation improved based on the speaking test results. The average pre-test score increased by about 30 points, rising from 56 in the pre-test to 90 in the post-test.

While AI speech evaluation programs have proven beneficial in many areas of language learning, other forms of technology, such as Automatic Speech Recognition (ASR) software, have been shown to improve pronunciation skills by providing learners, more autonomy, and a low-anxiety practice environment. A different study indicates that using ASR software is a simple and useful way to give students more control while learning pronunciation. The same research also states that with the help of ASR, students were able to conduct significant language experimentation in a safe setting, free from the potential anxiety that comes with speaking with others (McCrocklin, 2016).

Pronunciation-related anxiety can often hinder effective communication; therefore, the use of AI-driven technologies, such as ASR or TTS tools, offers a practical solution to overcoming such barriers in language learning settings. In this study, all participants agreed that the TTS application helped reduce their anxiety toward pronunciation learning, supporting the findings that such technologies can promote a more relaxed and effective learning environment.

Similarly, Neri et al. (2008) found that CAPT helps children pronounce words accurately. In their study, one group received teacher-led pronunciation training, while the other used a computer-based system (Parling) with an ASR component. Furthermore, research conducted by Eksi and Yesilcinar (2016) on the effectiveness of online TTS tools found that these tools are both efficient and easy to use. Additionally, teacher candidates, who participated in the study, said that they would utilize such tools in the future, which is noteworthy. Additionally, Gilakjani and Narjes Banou Sabouri (2017) found in their research that pronunciation software is an affordable and useful technique for teaching pronunciation. Interviews with Iranian instructors showed that using pronunciation software improved their teaching skills. However, while these studies highlight the benefits of technology in pronunciation learning, other research has examined the varied impacts of technology, shedding light on both its potential advantages and limitations.

In his study, Chiang (2019) explored how well students performed when dictating English vocabulary using dictation conducted by the teacher and TTS technology. The findings revealed that students did not prefer TTS technology in the language classroom. However, in a recent single-case research study on how learners with disabilities comprehended and enhanced their listening comprehension using Text-To-Speech software and a human reader, Brunow and Cullen (2021) concluded that reading proficiency and individual experiences affected how effective TTS was.

This contrasts with other studies that highlight the broader applicability of TTS technology across languages. Soler-Urzuá (2011) found that employing TTS effectively can help Spanish learners acquire the English /ɪ/ in an educational context and concluded that educators might benefit from its use. She mentions several benefits of using TTS in educational settings, such as taking on repetitive tasks and listening to texts repeatedly as learners wish. Instructors may feel uncomfortable reading a text repeatedly during a class to help learners understand distinct linguistic elements. Additionally, TTS can save teachers time searching for listening material by allowing them to enter written content and reuse it for aural exercises. This demonstrates several advantages of TTS technology, which not only promotes pronunciation improvement but also leads to significant advances in pronunciation accuracy, as evidenced by the pre- and post-test results.

The pre- and post-test findings of this study reveal improvement in learners' pronunciation accuracy across all categories, with percentages increasing significantly for both voiced and voiceless sounds, as well as silent letters. Notably, the silent-k category experienced the highest gain, increasing from 60% in the pre-test to 95.74% in the post-test, indicating significant improvement in overall pronunciation accuracy.

The participants' perceptions of the TTSReader application were overwhelmingly positive. All of the participants agreed that the TTS tool may help them improve their pronunciation, promote independent learning, and reduce anxiety toward pronunciation practice. While all students reported no trouble using the application, nearly half believed it could be time-consuming. Overall, the findings indicate that TTSReader is highly effective in pronunciation learning, with participants expressing trust in its potential.

Limitations and Recommendation

Limitations

It must be acknowledged that the current study has several limitations. Firstly, the lack of a control group and the 4-week duration of the training might affect the validity of the results. Furthermore, the group's size was not large enough to adequately generalize the results. Future research might use a greater number of participants and set up the experimental and control groups at the same time in the study. One limitation of the study is related to the questionnaire used to collect participants' opinions. The original questionnaire employed two-choice response options (agree/disagree), which may not have allowed for a nuanced understanding of participants' attitudes, particularly for those who were indifferent or neutral towards certain items. While a 5-point Likert scale could have provided a more objective assessment of the responses, the choice of a binary scale was made for simplicity from the participants' perspective. Despite this limitation, the findings provide valuable insights into participants' general perceptions of the TTS tool used in the study. In light of these limitations, the study might offer helpful insights and suggestions to educators across various environments who want to utilize AI-driven TTS tools for instructional reasons, to improve learners' pronunciation. Additionally, it provides suggestions for AI-driven TTS program makers worldwide to help them discover which AI programs could be more advantageous for language learners.

Conclusion

In conclusion, the study suggests that the web-based text-to-speech program improved Turkish EFL students' pronunciation skills, as evidenced by increased read-aloud performance and favorable student views. The research findings demonstrated that the web-based TTSReader technology assisted students while learning English as a foreign language. Their pronunciation improved at the word level across nine categories, including consonant clusters and silent letters. The questionnaire revealed that users found the TTSReader user-friendly and gave positive reviews of it. All participants stated that this application can decrease anxiety toward pronunciation learning and that they believe it promotes independent learning.

Based on these results, it can be concluded that TTSReader tool can be used as a pedagogical tool for teaching pronunciation in EFL classes, especially for learning correct pronunciation at the word level. Additionally, TTS technology allows students to refine their pronunciation at their own pace and comfort. Learners can repeat words or entire sentences as many times as needed to improve proper pronunciation. Individualized learning with TTS can be beneficial for students with diverse learning styles and preferences. Furthermore, TTS offers learners the opportunity to work on their pronunciation in a wide range of contexts, including the classroom, at home, and on the road. This ability to adapt fosters constant learning and practice.

References

- Acton, W. (1984). Changing Fossilized Pronunciation. *TESOL Quarterly*, 18(1), 71-85. JSTOR. <https://doi.org/10.2307/3586336>
- Ahn, T., & Lee, M. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology*, 47, 778–786. <https://doi.org/10.1111/bjet.12354>
- Brunow, D. A., & Cullen, T. A. (2021). Effect of Text-to-Speech and Human Reader on Listening Comprehension for Students with Learning Disabilities. *Computers in the Schools*, 38(3), 214-231. <https://doi.org/10.1080/07380569.2021.1953362>
- Burgess, J., & Spencer, S. (2000). Phonology and Pronunciation in Integrated Language Teaching and Teacher Education. *System*, 28, 191-215. [https://doi.org/10.1016/S0346-251X\(00\)00007-5](https://doi.org/10.1016/S0346-251X(00)00007-5)
- Cavus, N., & Ibrahim, D. (2017). Learning English using children's stories in mobile devices. *British Journal of Educational Technology*, 48(2), 625-641. <https://doi.org/10.1111/bjet.12427>
- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching Pronunciation Hardback with Audio CDs (2): A Course Book and Reference Guide*. Cambridge University Press.
- Chen, H. H.-J. (2011). Developing and evaluating an oral skills training website supported by automatic speech recognition technology. *ReCALL*, 23(1), 59-78. Cambridge Core. <https://doi.org/10.1017/S0958344010000285>
- Chiang, H.-H. (2019). A Comparison Between Teacher-Led and Online Text-to-Speech Dictation for Students' Vocabulary Performance. *English Language Teaching*, 12(3), 77. <https://doi.org/10.5539/elt.v12n3p77>
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- Demirezen, M. (2009). A model to rehabilitate a fossilized pronunciation error of Turkish English language teachers: The nasal devoicing of /ŋ/ as /ŋk/. *Procedia - Social and Behavioral Sciences*, 1(1), 2722-2727. <https://doi.org/10.1016/j.sbspro.2009.01.482>
- Ehbara, H. (2021). Impact of Computer-Assisted Pronunciation Training on Libyan Child Learners of English. *PhD Thesis*.
- Eksi, G. Y., & Yesilcinar, S. (2016). An Investigation of the Effectiveness of Online Text-to-Speech Tools in Improving EFL Teacher Trainees' Pronunciation. *English Language Teaching*, 9(2), 205. <https://doi.org/10.5539/elt.v9n2p205>
- Gilakjani, A. P. & Narjes Banou Sabouri. (2017). A Detailed Analysis over Iranian EFL Teachers' Beliefs towards Using Pronunciation Software in Teaching English Pronunciation. *International Journal of Applied Linguistics & English Literature*, 6(7), 227-235. Publicly Available Content Database. <https://doi.org/10.7575/aiac.ijalel.v.6n.7p.227>
- Handley, Z. (2009). Is text-to-speech synthesis ready for use in computer-assisted language learning? *Speech Communication*, 51(10), 906–919. <https://doi.org/10.1016/j.specom.2008.12.004>
- Hismanoglu, M. (2006). Current Perspectives on Pronunciation Learning and Teaching. *Journal of Language and Linguistic Studies*, 2(1), 101-110.
- Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*.

- Kung, F.-W. (2017). *Teaching and Learning English as a Foreign Language in Taiwan: A Socio-cultural Analysis*.
- Liu, J., & Shadiev, R. (2023). Review of research on applications of speech recognition technology to assist language learning. *ReCALL*, 35(1), 74-88. Cambridge Core. <https://doi.org/10.1017/S095834402200012X>
- McCrocklin, S. (2019). ASR-based dictation practice for second language pronunciation improvement. *J. Second Lang. Pronunciation*, 5(1), 98–118. <https://doi.org/10.1075/jslp.16034.mcc>
- McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of Automatic Speech Recognition. *System*, 57, 25-42. <https://doi.org/10.1016/j.system.2015.12.013>
- Morley, J. (1991). The Pronunciation Component in Teaching English to Speakers of Other Languages. *TESOL Quarterly*, 25(3), 481-520. <https://doi.org/10.2307/3586981>
- Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, 21(5), 393-408. <https://doi.org/10.1080/09588220802447651>
- Noviyanti, S. D. (2020). 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), 162. <https://doi.org/10.24235/eltecho.v5i2.7246>
- Peters T & Bell L. (2007). Choosing and using text-to-speech software. *Computers in Libraries*, 27(2), 26-29. CINAHL Ultimate.
- Piske, T. (2012). Factors affecting the perception and production of L2 prosody: Research results and their implications for the teaching of foreign languages. In P. A. Hunt & D. J. K. (Eds.), *Educational linguistics* (Vol. 15, pp. 41–59). https://doi.org/10.1007/978-94-007-3883-6_4
- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29(2), 191-215. <https://doi.org/10.1006/jpho.2001.0134>
- Saito, K., & Lyster, R. (2012). Effects of Form-Focused Instruction and Corrective Feedback on L2 Pronunciation Development of /ɪ/ by Japanese Learners of English. *Language Learning*, 62(2), 595-633. <https://doi.org/10.1111/j.1467-9922.2011.00639.x>
- Selinker, L. (1972). INTERLANGUAGE. *IRAL - International Review of Applied Linguistics in Language Teaching*, 10(1-4). <https://doi.org/10.1515/iral.1972.10.1-4.209>
- Shadiev, R., Hwang, W.-Y., Chen, N.-S., & Huang, Y.-M. (2014). Review of Speech-to-Text Recognition Technology for Enhancing Learning. *Journal of Educational Technology & Society*, 17(4), 65-84.
- Soler-Urzuá, F. (2011). *The Acquisition of English /ɪ/ by Spanish Speakers via Text-to-Speech Synthesizers: A Quasi-Experimental Study*.
- Tsai, P. (2019). Beyond self-directed computer-assisted pronunciation learning: A qualitative investigation of a collaborative approach. *Comput. Assist. Lang. Learn.*, 86, 713–744. <https://doi.org/10.1080/09588221.2019.1614069>
- Vargas, N. (2023). Secondary ELA teacher perceptions of the use of artificial intelligence as an instructional tool [Ed.D. dissertation, Lamar University]. *ProQuest Dissertations and Theses Global*. <https://www.proquest.com/dissertations-theses/secondary-ela-teacher-perceptions-use-artificial/docview/2900412932/se-2?accountid=159111>

- Wood, S. G., Moxley, J. H., Tighe, E. L., & Wagner, R. K. (2018). Does Use of Text-to-Speech and Related Read-Aloud Tools Improve Reading Comprehension for Students With Reading Disabilities? A Meta-Analysis. *Journal of Learning Disabilities*, 51(1), 73-84. <https://doi.org/10.1177/0022219416688170>
- Yeh, R. (2014). Effective Strategies for Using Text-to-Speech, Speech-to-Text, and Machine-Translation Technology for Teaching Chinese: A Multiple-Case Study [Ph.D., Northcentral University]. In *ProQuest Dissertations and Theses* (1734041300). ProQuest Dissertations & Theses Global. <https://www.proquest.com/dissertations-theses/effective-strategies-using-text-speech-machine/docview/1734041300/se-2?accountid=159111>
- Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023). An Investigation Into Artificial Intelligence Speech Evaluation Programs With Automatic Feedback for Developing EFL Learners' Speaking Skills. *SAGE Open*, 13(3), 21582440231193818. <https://doi.org/10.1177/21582440231193818>

Biographical notes:

Fatma Önder is a lecturer within the Department of Foreign Languages at Konya Food and Agriculture University and is currently pursuing a PhD in English Language Teaching at Necmettin Erbakan University.