EFL Students' Perception of Using AI Text-to-Speech Apps in Learning Pronunciation

Eman Abdel-Reheem Amin[1]

Abstract

The present study examined students' perception of using Artificial intelligence-powered text-to-speech apps (AI TTS Apps) to learn aspects of English language pronunciation. This investigation was conducted considering the extended Unified Theory of Acceptance and Use of Technology (UTAUT2). To ensure participants had prior experience using AI TTS for pronunciation, they engaged in five sessions focused on learning both segmental and suprasegmental aspects of pronunciation through a text-to-speech app. This research employed a mixed-method approach. A questionnaire was used to examine participants' perceptions of AI TTS Apps, while interviews were conducted to gain deeper insights into their experiences with AI TTS Apps. According to the results, participants positively perceived the use of AI TTS Apps in learning pronunciation. It is recommended to utilize AI Text-to-Speech Apps both within and outside pronunciation classes, due to their positive pedagogical effects.

Keywords: *Aspects of pronunciation, AI text-to-speech applications, Majmaah University, The Extended Unified Theory of Acceptance and Use of Technology (UTAUT2).*

Introduction

Advances in technology have facilitated learning procedures, providing more opportunities for self and independent learning. The development of software and apps, utilizing cutting-edge speech technology, is increasingly aiding foreign language learning. AI Text-to-Speech (TTS) Apps, employing artificial intelligence algorithms and voice synthesis technology, convert written text into lifelike spoken words. These apps have emerged as a viable choice in language classes (Bione, 2017). Previous studies have examined their effect on students' reading skills (e.g., Bone & Bouck, 2017; Wood et al., 2018; Young et al., 2019; Amin, 2022). In the context of pronunciation and phonological awareness, researchers have reported significant results (Cardoso, 2018; Ekşi & Yeşilçınar, 2016; Grimshaw et al., 2018; Meihami & Husseini, 2014). TTS Apps can also be utilized in listening classes (Oktalia & Drajati, 2018). Studies by Chang (2019), Cunningham (2013), Huang, Liao, (2015), and Young and Stover (2014) have examined their effect on writing and editing skills. Besides, they can be used as self-learning tools which in turn enhance learners' motivation and achievement (Svensson et al. 2019). Students have also shown a positive attitude toward using TTS (Bione et al., 2016)

As a new technology, it is necessary to examine learners' acceptance of these tools. Various models and theories have been developed to investigate users' and learners'

[1] Department of English, College of Education, Majmaah University, Majmaah, Saudi Arabia 11952, e.abdelrahim@mu.edu.sa

acceptance of, and perceptions toward, new technology, such as the Extended Unified Theory of Acceptance and Use of Technology (UTAUT2) by Venkatesh et al. (2012).

The rationale of the study:

In the learning context, researchers used to examine learners' acceptance and perception of new technologies through certain models such as TAM and UTAUT. However, they often neglected the impact of the three additional constructs of the UTAUT2 model (hedonic motivation, habit, and price value), possibly because UTAUT2 primarily targets consumers of technology. Recognizing learners as consumers of technology, especially considering that almost all college learners in Saudi Arabia have smartphones, this study adopted the UTAUT2 model as a framework for investigation. Consequently, the current study attempts to bridge the gap in the literature by investigating learners' perceptions in the light of the variables of the UTAUT2. Additionally, as suggested by Venkatesh et al. (2016), it examines learners' acceptance of tasks within the target technology contexts. Therefore, this study explores students' perception and acceptance of using AI TTS Apps as assistive technology tools that support learning aspects of pronunciation, such as segments and prosody elements, or suprasegmentals such as stress and intonation, and the relationship between sounds (i.e., assimilation, elision, and linking to provide connected speech.

Review of literature

Learning and teaching pronunciation

Pronunciation was not given much emphasis in second-language classrooms before the second half of the nineteenth century, due to the adoption of the grammar-translation method which prioritized written over spoken language. Later with the appearance of the behaviorist approach and the total physical response approach, there was a shift towards communication and spoken language; nevertheless, teaching and learning pronunciation depended on imitation and mimicry. At the beginning of the twentieth century, the intuitive-imitative and analytic-linguistic approaches shaped the teaching of pronunciation. Later with the appearance of the audio-lingual method, situational language teaching, and communicative language teaching, new teaching and learning practices changed pronunciation activities inside classes. With Krashen's monitor theory of second language acquisition (SLA), the focus became on comprehensible input to enable learners to communicate and acquire a language (Jarrosz, 2019; Murphy & Baker, 2019).

Advances in technology had the greatest impact on teaching and learning language in general, and pronunciation. Pedagogically, computer-assisted language learning (CALL) facilitates learners' autonomous learning of languages (Schmitt & Celce-Murcia, 2020). Later on, new concepts appeared in the field of pronunciation pedagogy for instance, Computer-assisted pronunciation training (CAPT) and Mobile-assisted pronunciation training (MAPT) which incorporate technologies e.g., automatic speech recognition (Rogerson-Revell, 2021; Walesiak, 2021). They introduce suggested strategies that help learners cope on their own in learning pronunciation such as listening to the pronunciation of words, sentences, or short texts and recording earners' production, employing correction strategies e.g., soliciting repetition, paraphrasing, and checking feedback (Burns & Seidlhofer, 2020).

CAPT and MAPT enable learners to be exposed to comprehensible input of language that can aid their acquisition of a language; they also allow learners to listen, repeat, and imitate spoken language which is considered a method to improve their pronunciation (Safavi, 2021). The research by Walesiak (2021) concluded that Students are enthusiastic about using mobile applications to learn pronunciation. Fouz-González (2020) found that

training learners on using a pronunciation app improved their perception and production of English pronunciation.

Using technology to enhance pronunciation has been a subject of interest in pronunciation pedagogy. Teachers have sought relevant and innovative technologies to assist learners in achieving intelligibility, comprehensibility, and interpretability of pronunciation (Jarosz, 2019). Furthermore, researchers have investigated learners' perceptions of technologies and applications for teaching and learning pronunciation, such as AI TTS. For example, Bione's research (2017) evaluated the use of TTS in terms of comprehensibility, naturalness, accuracy, and intelligibility. In their subsequent studies, Bione et al. (2016, 2017) concluded that EFL learners have a positive attitude toward using TTS synthesizers in learning pronunciation. Proença et al. (2019) recommended the application of AI TTS Apps to teach pronunciation rules to L2 learners.

AI-powered Text-to-speech applications (AI TTS Apps)

Natural Language Generation (NLG) is a subfield of artificial intelligence and computational linguistics concerned with designing software to produce appropriate output in written or spoken form (Reiter & Dale, 2000). In recent years, NLG, coupled with Speech-processing technology and Speech synthesis, has garnered significant attention. Speech synthesis involves the concept of enabling computers to generate human-like speech, encompassing coding, decoding, and speech production. This has led to the emergence of related technologies such as text-to-speech, speech recognition systems, and machine translation.

The term "text-to-speech" refers to the process of converting written texts into speech, essentially making it an AI-powered voice generator. Conversely, speech recognition systems aim to convert spoken words into written texts, while machine translation deals with translating between languages through either oral or written texts and presenting the results in either spoken or written form. Initially, these systems produced robotic sounds and mechanical voices; however, with advancements in technology, they have evolved to generate more human-like voices (Taylor, 2009).

TTS software stands as one of the NLG applications based on AI-powered voice generators. The function of TTS comprises two integral modules: Natural Language Processing (NLP) and Digital Signal Processing (DSP). NLP aims to read, decode, and interpret the input, generating a phonetic transcription of the text with appropriate intonation and rhythm. On the other hand, DSP converts the received information into a natural-sounding voice (Dutoit, 1997)

According to Dutoit (1997), TTS systems exhibit segmental and supersegmental qualities that vary across different systems. The segmental quality relies on producing natural sounds, provided that the NLP delivers human-like information. Meanwhile, the supersegmental qualities are linked to the richness of prosodic features delivered by the DSP module. These qualities play a crucial role in determining the intelligibility and naturalness of the produced voice by the TTS system. Reiter and Dale (2000) later suggested that TTS systems can evolve into concept-to-speech systems. They argued that intonation is not solely related to syntactic features but is also intricately connected to the communicative goal of the text.

AI Text-to-Speech (TTS) Apps hold significant value as pedagogical and learning tools (Cardoso et al., 2015). They serve as assistive technologies, particularly beneficial for individuals with visual impairments. In recent times, the application of these apps in addressing learning disabilities and difficulties has demonstrated remarkable results, with a focus on learners facing reading difficulties (Young et al., 2019). Furthermore, TTS Apps prove beneficial in enhancing pronunciation skills for foreign language learners, as evidenced by studies conducted by Bione et al. (2016), Grimshaw et al. (2018), Proença et al. (2019), and Bione & Cardoso (2020).

The Extended Unified Theory of Acceptance and Use of Technology (UTAUT2)

Researchers have explored various theoretical models, including the Theory of Reasoned Action (TRA) and the Theory of Planned Behavior (TPB), to investigate technology acceptance. One prominent model that builds upon these theories is the Technology Acceptance Model (TAM). Furthermore, Venkatesh et al. (2003) combined perspectives from eight different models to create the Unified Theory of Acceptance and Use of Technology (UTAUT). UTAUT comprises four key constructs: "Performance Expectancy (PE), Effort Expectancy (EE), Social Influence (SI), and Facilitating Conditions (FC). Additionally, the model includes four moderators—age, gender, experience, and voluntariness" (Venkatesh et al., 2003, p. 447). According to UTAUT, these constructs collectively influence Behavioral Intention (BI) to use technology, thereby providing insight into the understanding of technology acceptance.

In 2012, Venkatesh et al. broadened their UTAUT by integrating three additional constructs: hedonic motivation (HM), price value (PV), and habit (H). They incorporated "hedonic motivation and price value as new predictors of behavioral intention and habit as a new predictor of both intention and technology use". Venkatesh et al. (2016, p. 335). Thus, UTAUT2 examines users' acceptance of modern technology in the light of "Performance Expectancy, Effort Expectancy, Social Influence, Facilitating Conditions, Hedonic Motivation, Habit, and Price value".

Several researchers have attempted to expand the UTAUT2 by incorporating new constructs. For instance, Ain et al. (2015) suggested considering learning value in a learning context instead of price value, particularly when learners are not required to pay for their learning. The present study aims to examine learners' acceptance of using AI TTS Apps to improve their English pronunciation through the constructs of the UTAUT2 Model. Therefore, the study's questions are as follows:

Research questions

What are students' perceptions of using AI TTS Apps in learning pronunciation?

What expectations do students have regarding the use of AI TTS Apps in learning pronunciation skills?

What factors may influence students' use of AI TTS Apps in learning pronunciation?

Methodology

Research design

This study adopted the mixed method approach where both quantitative and qualitative data are analyzed to check the acceptance or rejection of the hypotheses.

Participants

Ninety students enrolled in the English language department at Zulfi College of Education, Majmaah University, Saudi Arabia, were the participants of the study during the academic year 2022. They were selected from pronunciation, and speaking, and listening courses, volunteers from other classes participated. They were trained in how to use AI TTS Apps to be familiar with them. They studied five sessions on aspects of pronunciation such as segments (i.e., vowels, consonants, and individual sounds) and super segmental such as stress and intonation, and the relationship between sounds (e.g., assimilation, elision, and linking to provide connected speech. Each session included some sentences and examples that illustrated the pronunciation aspect being studied. These sentences were imported into the TTS AI Apps. After the training sessions, students conducted a questionnaire and interview to check their perception of using AI TTS Apps in learning pronunciation.

Tools

A questionnaire

A questionnaire based on the UTAUT2 model was developed. It consisted of 34 items over eight dimensions as follows: performance expectancy (n= 5), effort expectancy (n=4), social influence (n=4), facilitating conditions (n=6), Behavioral intention (n=5), hedonic motivation (n=3), habit (n=3), and price value (n=3).

Validity and Reliability of the questionnaire

A panel of TEFL specialists validated the questionnaire, and Cronbach's alpha values, detailed in Table 1, were examined to ensure its reliability. The obtained values surpassed the recommended threshold of 0.7 for all variables, signifying a high level of reliability. This observation supports the overall dependability of the questionnaire.

Table 1 The reliability of the questionnaires

Dimensions	Cronbach's alpha
PE	.81
EE	.73
SI	.83
FC	.80
BI	.90
HM	.82
H	.83
PV	.85
Total	.95

** Correlation is significant at the 0.01 level (2-tailed).

Additionally, internal consistency within the questionnaire dimensions was assessed, as outlined in Table 2. The results indicate that all correlation coefficients between individual dimensions and the overall questionnaire score reached statistical significance at the 0.01 level.

Table 2 The Correlation Coefficient between each dimension and the total score of the questionnaires

Dimensions	Correlation with the total score of the questionnaire
PE	(**).82
EE	(**).77
SI	(**).75
FC	(**).84
BI	(**).83
HM	(**).87
H	(**).80
PV	(**).70

** Correlation is significant at the 0.01 level (2-tailed).

An interview

A structured digital interview was conducted to get a deeper understanding of students' perception of using AI TTS Apps to learn aspects of pronunciation. The interview consisted of two open-ended questions. The questions aimed to check participants' expectations of using AI TTS Apps and the factors that influence the use of such apps. Students received the questions and then recorded their responses. The interview questions were submitted to a group of specialists in TEFL and applied linguistics to check their validity.

Results

Questionnaire results

The descriptive statistics and the results of the questionnaires are presented below.

Table 3 Descriptive Statistics of the questionnaire.

	N	No. of items	Minimum	Maximum	Mean	Std. Deviation
Total	90	34	2.94	5.00	4.01	.52
PE	90	5	2.80	5.00	4.24	.59
EE	90	4	2.75	5.00	4.11	.54
SI	90	4	1.50	5.00	3.66	.79
FC	90	6	2.67	5.00	3.93	.58
BI	90	5	2.33	5.00	4.13	.65
HM	90	3	2.67	5.00	4.12	.66
H	90	3	2.00	5.00	4.02	.76
PV	90	3	2.00	5.00	3.76	.73

The descriptive statistics of the questionnaire, as shown in Table 3, indicate that a total of 90 students participated (n=90). The second column illustrates the number of items in the questionnaire and the distribution for each dimension. The mean values demonstrate that students, on average, expressed an overall agreement on the dimensions covered in the questionnaire.

Table 4 The frequency of students' responses, means, standard deviations, rank, and weighted mean for Performance Expectancy (PE)

PE		Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	SD	Rank	result
PE1	F	46	38	6	0	0	4.44	.620.	2	Strongly agree
	%	51.1	42.2	6.7	0	0				
PE2	F	48	36	5	1	0	4.45	656.2	1	Strongly agree
	%	53.3	40.0	5.6	1.1	0				

		Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	SD	Rank	result
PE3	F	30	45	1	1	3	4.08	.8949	4	Agree
	%	33.3	50.0	12.2	1.1	3.3				
PE4	F	37	38	12	2	1	4.20	.8373	3	Agree
	%	41.1	42.2	13.3	2.2	1.1				
PE5	F	30	40	15	5	0	4.05	.8527	5	Agree
	%	33.3	44.4	16.7	5.6	0				
Weighted mean		4.244								Agree

The results presented in Table 4 reveal that students had a positive perception of the benefits of using AI TTS Apps in learning pronunciation. They strongly agreed that these apps were useful in their pronunciation studies, with mean scores of 4.5 and 4.4, respectively. Furthermore, students showed agreement on expectations of achieving better grades, increased participation, and progress in pronunciation classes (items no. PE3, PE4 & PE6). The weighted mean (4.24) indicates a positive performance expectancy of using AI TTS Apps.

Table 5 The frequency of students' responses, means, standard deviations, rank, and weighted mean for Effort Expectancy (EE)

EE		Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	SD	Rank	result
EE1	F	34	47	9	0	0	4.20	.8373	1	Agree
	%	37.8	52.2	10.0	0	0				
EE2	F	37	46	7			4.05	.8527	3	Agree
	%	41.1	51.1	7.8	0	0				
EE3	F	33	37	17	3	0	4.11	.8271	2	Agree
	%	36.7	41.1	18.9	3.3	0				
EE4	F	14	45	25	6	0	3.74	8011.	4	Agree
	%	15.6	50.0	27.8	6.7	0				
Weighted mean		4.025								Agree

In terms of effort expectancy, the results in Table 5 indicate that using AI TTS Apps in learning pronunciation was perceived as easy by students, with a weighted mean of 4.0. Participants felt skillful at using the apps (m=4.2) and agreed upon their overall ease of use (m=4.0). Furthermore, they found it easy to scan, upload, and type texts into the TTS Apps (m=4.1), although creating audiobooks and files in TTS AI Apps was rated slightly lower (m=3.7).

Table 6 The frequency of students' responses, means, standard deviations, rank, and weighted mean for Social Influence (SI)

SI		Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	SD	Rank	result
SI1	F	13	31	28	15	3	3.40	1.03	4	Agree
	%	14.4	34.4	31.1	16.7	3.3				
SI2	F	12	40	20	15	3	3.47	1.03	3	Agree
	%	13.3	44.4	22.2	16.7	3.3				
SI3	F	26	35	21	6	2	3.85	.9893	2	Agree
	%	28.9	38.9	23.3	6.7	2.2				
SI4	F	22	43	21	4	0	3.92	.8104	1	Agree
	%	24.4	47.8	23.3	4.4	0				
Weighted mean		3.66								Agree

As shown in Table 6, participants agreed on the influence of social factors on their use of AI TTS Apps, with a weighted mean of 3.66. The most impactful factor was colleagues' beliefs in the usefulness of the apps, followed by the teacher's encouragement to use them, with means of 3.9 and 3.8, respectively. The impact of other people (items no. SI1 & SI2) had the least consequence, with a mean of 3.4.

Table 7 The frequency of students' responses, means, standard deviations, rank, and weighted mean for Facilitating Conditions (FC)

FC		Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	SD	Rank	result
FC1	F	23	53	14	0	0	4.10	.6368	1	Agree
	%	25.6	58.9	15.6	0	0				
FC2	F	19	52	14	3	2	3.92	.8377	3	Agree
	%	21.1	57.8	15.6	3.3	2.2				
FC3	F	22	43	22	2	1	3.92	.8241	4	Agree
	%	24.4	47.8	24.2	2.2	1.1				
FC4	F	24	53	10	3	0	4.08	.7133	2	Agree
	%	26.7	58.9	11.1	3.3	0				
FC5	F	23	39	19	7	2	3.82	.9781	5	Agree
	%	25.6	43.3	21.1	7.8	2.2				

		Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	SD	Rank	result
FC6	F	16	43	24	7	0	3.75	.8388	6	Agree
	%	17.8	47.8	26.7	7.8	0				
Weighted mean		3.93								Agree

In Table 7, the results indicate agreement on the availability of facilitation conditions to use TTS Apps, with a weighted mean of 3.9. Participants expressed consensus that the instructions for using AI TTS Apps are clear (m=4.10) and the user guide is sufficiently clear (m=4.0). Additionally, students agreed on the appropriateness of voice options, including tone, gender, speed, and accent (items no. FC2 & FC3, m=3.9). Other factors, such as the clarity of the user guide, tutorial videos, and Qs & As, showed consistent agreement among students (items no. FC4, FC5, & FC6, m=4.0, 3.8, & 3.7)."

Table 8 The frequency of students' responses, means, standard deviations, rank, and weighted mean for Behavioral Intention (BI)

BI		Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	SD	Rank	result
BI1	F	31	39	17	2	1	4.07	.85101	6	Agree
	%	34.4	43.3	18.9	2.2	1.1				
BI2	F	31	43	12	4	0	4.12	.80487	5	Agree
	%	34.4	47.8	13.3	4.4	0				
BI3	F	29	46	14	1	0	4.14	.71203	2	Agree
	%	32.2	51.1	15.6	1.1	0				
BI4	F	33	45	10	2	0	4.21	.72661	1	Agree
	%	36.7	50.0	11.1	2.2	0				
BI5	F	32	44	9	5	0	4.14	.81504	3	Agree
	%	35.6	48.9	10.0	5.6	0				
BI6	F	32	42	13	2	1	4.13	.82380	4	Agree
	%	35.6	46.7	14.4	2.2	1.1				
Weighted mean		4.13								Agree

Table 8 results imply that students had behavioral intention to use AI TTS Apps in learning pronunciation (Weighted mean =4.13). The highest agreement goes for item no. BI4 as students would like to use TTS APPs to practice pronunciation (m=4.2). They intended to use them outside classes (item no. B2, m=4.1). They agreed that they would encourage colleagues to use the Apps (item no. BI5, M=4.1). They also agreed upon using the APPs to improve and learn aspects of pronunciation items no. BI6 &BI1, M=4.1&4.0 respectively.

Table 9 The frequency of students' responses, means, standard deviations, rank, and weighted mean for Hedonic Motivation (HM)

HM		Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	SD	Rank	result
HM1	F	28	43	7	2	0	4.07	.7677	3	Agree
	%	31.1	47.8	18.9	2.2	0				
HM2	F	33	40	14	3	0	4.14	.8011	1	Agree
	%	36.7	44.4	15.6	3.3	0				
HM3	F	31	42	16	1	0	4.14	.7429	2	Agree
	%	34.4	46.7	17.8	1.1	0				
Weighted mean		4.11								Agree

Table 9 results suggest students' agreement on the hedonic motivational effect of TTS App use, with a weighted mean of 4.1. They indicated enjoyment, interest, and pleasure in using the Apps (items no. HM2, HM3, & HM3, M=4.1, 4.1, 4.0 respectively).

Table 10 The frequency of students' responses, means, standard deviations, rank, and weighted mean for Habit (H)

H		Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	SD	Rank	result
	%	34.4	46.7	17.8	1.1	0				
H1	F	31	39	12	8	0	4.03	.9173	2	Agree
	%	34.4	43.3	13.3	8.9	0				
H2	F	28	47	13	2	0	4.12	.7317	1	Agree
	%	31.1	52.2	14.4	2.2	0				
H3	F	27	38	18	5	2	3.92	.9625	3	Agree
	%	30.0	42.2	20.0	5.6	2.2				
Weighted mean		4.02								Agree

Table 10 displays students' habits of using AI TTS Apps, showing a weighted mean of 4.0. For items H1 &H2, participants agreed on using the APPs to learn pronunciation and

listen to English texts (m= 4.1, 4.0). Moreover, they agreed that using AI TTS Apps became a habit for learning pronunciation (m=3.92).

Table 11 The frequency of students' responses, means, standard deviations, rank, and weighted mean for price value (PV)

PV		Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	SD	Rank	result
PV1	F	18	36	32	4	0	3.75	.8253	2	Agree
	%	20.0	40.0	35.6	4.4	0				
PV2	F	18	34	31	7	0	3.70	.8798	3	Agree
	%	20.0	37.8	34.4	7.8	0				
PV3	F	19	42	25	4	0	3.84	.8060	1	Agree
		21.1	46.7	27.8	4.4	0				
Weighted mean		3.76								Agree

Results in Table 11 show students' agreement on the price value of AI TTS Apps, with a weighted mean of 3.7. They found it valuable and affordable (m= 3.8 & 3.7).

Interview results

Ninety students responded to the digital interview questions. Table 12 shows students' answers to the question, 'What expectations do you have regarding the use of AI TTS Apps in learning pronunciation skills?'

Table 12 Participants' expectations of using TTS AI Apps in learning pronunciation skills

	Category	Frequency	Percentage
Responses			
	Improve pronunciation an intonation	31	34%
	Enhance connected speech sounds	15	17%
	Correct mistakes in pronunciation	10	11%
	Gain good marks.	13	14%
	Facilitate learning of pronunciation	21	23%
	Total no of responses	90	100%

The results indicate that 34% of the participants expected improvement in their pronunciation, while 23% believed that it would facilitate their learning of pronunciation. Approximately 17% of the students noted that the apps enhanced connected speech and linking sounds. Fourteen percent mentioned the possibility of achieving good grades and marks using AI TTS Apps. For 11% of the students, it is considered a good method to correct mistakes in pronunciation.

Table 13 summarizes answers to the second question of the interview: 'What factors may influence students' use of AI TTS Apps in learning pronunciation?'

Table 13 Factors affect participants' use of TTS AI Apps

	Category	Frequency	Percentage
Responses	Ease of use	29	32%
	Cost and price	13	14%
	Effectiveness	28	31%
	Time constraints	20	22%
	Total no of responses	90	100%

Table 13 categorizes the factors that affect participants' use of AI TTS Apps. The primary factor for students was ease of use, accounting for 32%. The second influential factor was the effectiveness of the apps, representing 31%. The third reason identified was time constraints, at 22%. Only 14% of the students suggested that cost and price might hinder their use of AI TTS Apps.

Discussion

The study aimed to investigate learners' perception of using AI TTS Apps in learning pronunciation. To answer the study questions, a questionnaire and an interview were applied. Table 3's findings display that they had a positive perception towards the use of AI TTS Apps. Their perception of pronunciation learning via AI TTS Apps was investigated within each construct of the extended unified theory of acceptance and use of technology (UTAUT2) model. The results of this study are consistent with Bione et. (2016), who concluded that students had an overall positive attitude and perception of the pedagogical use of TTS in EFL classes.

To address the first question of the study, which examines students' perception of using AI TTS Apps in learning pronunciation in the light of the UTAUT2, the results of each construct are discussed below.

Performance expectancy

The results of Table 4 indicate students' performance expectancy (PE) of using TTS. Students believed that using AI TTS Apps was beneficial and helped them in learning pronunciation. It facilitated their learning, helping them achieve better grades and actively participate in pronunciation classes. This result aligns with Bione et al. (2017), who reported the usefulness of using AI TTS Apps in learning pronunciation in terms of intelligibility. The study also shares similar findings with Oktalia (2018) and Oktalia, and Drajati (2018), who indicated that AI TTS Apps were helpful in teaching pronunciation.

Other studies, such as De Araújo Gomes (2018) and Fouz-González (2020), concluded that TTS helped students develop phonological awareness and positively influenced their pronunciation production. These Apps can also enhance the correct pronunciation of fossilized words by affecting both the production and perception of segmental features of pronunciation.

Effort expectancy

As for effort expectancy, the results from Table 5 reveal that learners reported a positive agreement on the ease of using AI TTS Apps in learning pronunciation. This result is consistent with Fouz-González (2020), who reported the ease of use of CAPT Apps. Similarly, ease of use has been identified as a significant factor influencing mobile learning, as affirmed by Chao (2019).

Social influences

Results of the effect of social influences on learners' use of AI TTS Apps are indicated in Table 6. Students agreed that their colleagues' beliefs in the Apps' usefulness and the encouragement of their teachers positively affected their use of TTS AI Apps. This result is congruent with the results of Venkatesh et al. (2003), who emphasized the importance of social influences on learners' acceptance of modern technologies. Similarly, Ab Jalil et al. (2022) found that social influence variables directly affect the behavioral intentions of using new technologies.

Facilitating conditions

The results from Table 7 highlight the availability of facilitating conditions of AI TTS Apps, such as clear instructions for use. Additionally, these Apps provide a clear voice, enabling learners to easily grasp pronunciation. Learners can also adjust the speed of the voice and choose from various accents. These facilitating features align with the findings of Cardoso (2018) and Grimshaw, Bione, and Cardoso (2018). Similarly, the clarity of voice provided by mobile pronunciation applications has been noted to facilitate students' use and enhance their pronunciation perception and production (Fouz-González, 2020). Moreover, Oktalia (2018) reported that TTS may produce a less natural voice at reduced speeds, but students still benefit from the correct pronunciation it produces.

On the contrary, Bione and Cardoso (2020) pinpointed that the synthetic voice provided by TTS Apps did not affect EFL students' potential benefits in pronunciation classes. Thus, AI TTS Apps offer pedagogical potentialities that could be effectively utilized in L2 pronunciation classes.

Behavioral intention

Behavioral intention is one of the dimensions measured by the questionnaire, and Table 8 indicates that students expressed the intention to use AI TTS Apps in their pronunciation learning even outside the classrooms. As a form of Computer-Assisted Pronunciation Training (CAPT), TTS provides self-use and independent learning opportunities for learners, as indicated in the review by Rogerson-Revell (2021). The results align with the findings of Bione et al. (2016), suggesting that students are likely to use TTS outside classes, especially when they lack the natural environment for exposure to a foreign language. The present study agrees with previous research that TTS Apps allow for autonomous learning of pronunciation (e.g., Bione et al., 2017; Ekşi & Yeşilçınar, 2016). Other studies also support the idea of using apps for the retention of pronunciation learning, such as the study by Walesiak (2021).

Hedonic motivation

The results from Table 9 demonstrate students' agreement on the hedonic motivational effect of using AI TTS Apps. They expressed that the experience was interesting, enjoyable, and pleasurable. This positive hedonic motivation suggests that students may be willing to use AI TTS Apps in their pronunciation learning. The hedonic motivation construct has been identified in previous studies as a significant factor in the acceptance of new technologies, including the use of smartphones and their applications in learning (e.g., Ahmed & Kabir, 2018; Chao, 2019; Moorthy et al., 2019).

Habit

The results from Table 10 indicate that students agreed that using AI TTS Apps could become one of their habits for learning pronunciation. The habit construct is a significant factor that may influence higher education students' use of smartphone apps in their learning. This finding aligns with the results reported by Moorthy et al. (2019) and Nikolopoulou et al. (2020). The latter study concluded that habit was the strongest factor influencing students' use of TTS Apps.

Price Value

The cost of the AI TTS App is free of charge, but students can upgrade to access additional features. As indicated in Table 11, students agreed that they could afford the price of these apps. However, this result contrasts with the findings of Fouz-González (2020), who pointed out that, while pronunciation apps are affordable, most students were not willing to pay for them.

The interview results provide answers to the second question of the study: 'What expectations do you have regarding the use of AI TTS Apps in learning pronunciation skills?' As indicated in Table 12, students demonstrated that AI TTS Apps improve pronunciation and intonation, enhance connected speech sounds, correct mistakes in pronunciation, and facilitate the learning of pronunciation. This result aligns with Grimshaw et al. (2018), who concluded that TTS is accepted in pronunciation classes for its comprehensibility, naturalness, and intelligibility.

Students' responses reveal that they expect to use TTS Apps independently to improve and learn pronunciation, a finding consistent with Ekşi & Yeşilçınar (2016). However, it's worth noting that the results of the interview indicate that students perceive TTS as a learning tool that enhances both segmental and suprasegmental features of pronunciation. This differs from the study by Bione et al. (2016), which reported that only segmentals were perceived as good, while suprasegmentals could be problematic, as TTS lacks native rhythm, pauses, and phrase stress, etc. Their result may be affected by an outdated version of TTS Apps.

Table 13 summarizes answers to the third question of the study: What factors may influence your use of AI TTS Apps in learning pronunciation?' Students reported that ease of use, cost and price, effectiveness, and time constraints are common factors that affect their use of TTS Apps. These findings align with previous studies by Oktalia (2018), Oktalia, and Drajati (2018), and Fouz-González (2020).

In summary, learners agreed that using AI TTS Apps would be useful in learning segments (i.e., vowels, consonants, and individual sounds), as well as in learning the suprasegmental aspects (i.e., stress, intonation, and rhythm) and connected speech rules (assimilation, linking, elision). Additionally, they found it easy to use TTS Apps and expressed the intention to use them for learning and improving various aspects of pronunciation. Furthermore, TTS Apps were perceived as enjoyable, potentially becoming one of their learning habits, and the price of TTS Apps was considered reasonable and worth their value."

Conclusions, recommendations, and suggestions

In conclusion, the study found that students have a positive attitude and perception of using AI TTS Apps for learning and improving English language pronunciation. The facilitating conditions, ease of use, and effort expectancy associated with AI TTS Apps enable learners to access the target language anytime and anywhere, thereby facilitating their foreign language learning. These Apps can be effectively incorporated into pronunciation and listening classes, offering pedagogical benefits that extend beyond the confines of traditional classroom settings. Moreover, AI TTS Apps provide self-learning opportunities for both second language (L2) and foreign language (FL) learners, particularly those with limited exposure to authentic learning environments.

Future experimental research could investigate the effectiveness of using AI TTS Apps in improving students' pronunciation, intonation, and overall speaking skills. Moreover, there is an opportunity to enhance and develop other variables such as oral reading, fluency, and listening comprehension within empirical studies. Research endeavors may

also focus on exploring students' motivation and attitude toward using AI TTS Apps in learning or acquiring foreign and second languages.

Further studies could explore whether using AI TTS Apps can effectively address some of the pronunciation challenges faced by adult learners, including issues related to fossilization. Additionally, researchers may explore whether providing opportunities to listen to the voices of native speakers via AI TTS Apps could enhance students' episodic memory and aid in learning vocabulary.

Given that this study was limited to female learners, future research could replicate the study to investigate potential gender differences between males and females in their perception of using AI TTS Apps in learning and acquiring pronunciation.

Acknowledgments

References

Ab Jali, H., Rajakumar, M.,& Zaremohzzabieh,Z. (2022). Teachers' Acceptance of Technologies for 4IR Adoption: Implementation of the UTAUT M (Placeholder1)odel. International Journal of Learning, Teaching and Educational Research, 21(1),18-32. https://doi.org/10.26803/ijlter.21.1.2

Ahmed, M. S., & Kabir, A. (2018). The acceptance of smartphone as a mobile learning tool: Students of business studies in Bangladesh. Malaysian Online Journal of Educational Technology, 6(2), 38–47. DOI: 10.17220/mojet.2018.02.003

Ain, N., Kaur , K., & Waheed, M. (2015) The influence of learning value on learning management system use: An extension of UTAUT2. Information Development 1–16. https://doi.org/10.1177/0266666915597546

Amin, E. A. (2022). Using Repeated-- Reading and Listening—While-- Reading via Text—To Speech APPs. in Developing Fluency and comprehension. World Journal of English Language, 12(1), 211-220. https://doi.org/10.5430/wjel.v1122nn11pp2211

Bione, T. (2017). Synthetic voices in the foreign language context (Doctoral dissertation, Concordia University). https://core.ac.uk/reader/211520078

Bione, T., Grimshaw, J., & Cardoso, W. (2016). An evaluation of text-to-speech synthesizers in the foreign language classroom: learners' perceptions. In S. Papadima-Sophocleous,, L. Bradley & S. Thouësny (Eds), CALL communities and culture – short papers from EUROCALL 2016 (pp. 50-54). Research-publishing.net. https://doi.org/10.14705/rpnet.2016.eurocall2016.537

Bione, T., Grimshaw, J., & Cardoso, W. (2017). An evaluation of TTS as a pedagogical tool for pronunciation instruction: the 'foreign' language context. In K. Borthwick, L. Bradley & S. Thouësny (Eds), CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017 (pp. 56-61). Research-publishing.net. https://doi.org/10.14705/rpnet.2017.eurocall2017.689

Bione, T., & Cardoso, W. (2020). Synthetic voices in the foreign language context. Language Learning & Technology, 24(1), 169–186. https://doi.org/10125/44715

Bone, E. K., & Bouck, E. C. (2017). Accessible Text-to-Speech Options for Students Who Struggle with Reading. Preventing School Failure, 61 (1), 48-55. https://doi.org/10.1080/1045988X.2016.1188366

Burns, A. & Seidlhofer , B. (2020). Speaking and pronunciation. In N. Schmitt and M. P.H. Rodgers (Eds.), An Introduction to Applied Linguistics(pp.240-258) (3rd ed.). Routledge

Cardoso, W. (2018). Learning L2 pronunciation with a text-to-speech synthesizer. In P. Taalas, J. Jalkanen, L. Bradley & S. Thouësny (Eds), Future-proof CALL: language learning as

exploration and encounters – short papers from EUROCALL 2018 (pp. 16-21). Research-publishing.net.https://doi.org/10.14705/rpnet.2018.26.806

Cardoso, W., Smith, G., & Garcia Fuentes, C. (2015). Evaluating text-to-speech synthesizers. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds), Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy (pp.108-113). Dublin: Research-publishing.net. http://dx.doi.org/10.14705/ rpnet.2015.000318

Chao, C-M (2019). Factors Determining the Behavioral Intention to Use Mobile Learning: An Application and Extension of the UTAUT Model. Front. Psychol. 10:1652. doi: 10.3389/fpsyg.2019.01652

Cunningham, R. (2013). The Effects of Word Prediction and Text-to-Speech on the Writing Process of Translating. (Doctoral dissertation 322).https://irl.umsl.edu/dissertation/322

De Araújo Gomes, A. A., Cardoso, W., & De Lucena, R. M. (2018). Can TTS help L2 learners develop their phonological awareness? In P. Taalas, J. Jalkanen, L. Bradley & S. Thouësny (Eds), Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018 (pp. 29-34). Research-publishing.net. https://doi.org/10.14705/rpnet.2018.26.808

Dutoit, T. (1997). An Introduction to Text-to-speech Synthesis. Dordrecht: Kluwer Academic.

Ekşi G. Y., & Yeşilçınar, S. (2016). An Investigation of the Effectiveness of Online Text-to-Speech Tools in Improving EFL Teacher Trainees' Pronunciation. English Language Teaching, 9 (2), 205-214. URL: http://dx.doi.org/10.5539/elt.v9n2p205

Fouz-González, J. (2020). Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. Language Learning & Technology, 24(1), 62–85. https://doi.org/10125/44709

Grimshaw, J., Bione, T., & Cardoso, W. (2018). Who's Got Talent? Comparing TTS Systems for Comprehensibility, Naturalness, and Intelligibility. In p. Taalas, J. Jalkanen, L. Bradley, & S. Thouësny (EDs.). Future-proof CALL: language learning as exploration and encounters – short papers from EUROCALL 2018. (pp. 83-88). https://doi.org/10.14705/rpnet.2018.26.817.

Jarosz, A. (2019). English Pronunciation in L2 Instruction: The Case of Secondary School Learners. Springer

Meihami, H., & Husseini, F. (2014). Bringing TTS software into the classroom: The effect of using text to speech software in teaching reading features. Teaching English with Technology, 14(1), 23-34. http://www.tewtjournal.org

Moorthy, K., Tsen, T. Y., Loh, C. T., & Vikniswari, V. K. (2019). Habit and hedonic motivation are the strongest influences in mobile learning behaviours among higher education students in Malaysia. Australasian Journal of Educational Technology, 35(4), 174-191. https://doi.org/10.14742/ajet.4432

Murphy, J.M., & Baker, A.A. (2019). History of ESL Pronunciation Teaching. In M. Reed, J. Levis (Eds.), The Handbook of English Pronunciation (pp.36-66). Wiley Blackwell

Nikolopoulou, K., Gialamas, V., & Lavidas, K. (2020). Acceptance of mobile phone by university students for their studies: an investigation applying UTAUT2 model. Education and Information Technologies. https://doi.org/10.1007/s10639-020-10157-9

Oktalia, D. (2018). The Implementation of E-Speak in EFL Classroom What Do Teachers Think? 2 nd English Language and Literature International Conference (ELLiC). Semarang ,5 May 2018. https://jurnal.unimus.ac.id/index.php/ELLIC/article/download/3475/3316

Oktalia, D., & Drajati, A. (2018). English teachers' perceptions of text to speech software and Google site in an EFL Classroom: What English teachers really think and know, International Journal of Education and Development using Information and Communication Technology (IJEDICT), 14 (3),183-192 https://www.learntechlib.org/p/188288/.

Proença, J., Raboshchuk, G., Costa, Â., Lopez-Otero, P., Anguera, X. (2019) Teaching American English pronunciation using a TTS service. Proc. SLaTE 2019: 8th ISCA Workshop on Speech and Language Technology in Education, 59-63, DOI: 10.21437/SLaTE.2019-11.

Reiter, E., & Dale, R. (2000). Building Natural Language Generation Systems. Cambridge University press

Rogerson-Revell, P. M. (2021). Computer-Assisted Pronunciation Training (CAPT): Current Issues and Future Directions. RELC Journal: A Journal of Language Teaching and Research, 52 (1) ,189-205. https://doi.org/10.1177/0033688220977406

Safavi, S. (2021). Accent conversation in Computer –Assisted Pronunciation Training (CAPT). In K. Kelch, Peter Byun, S. Safavi, S. Cervantes (Eds.). CALL Theory Applications for Online TESOL Education (pp127-160). IGI Global. DOI: 10.4018/978-1-7998-6609-1

Schmitt, N., & Celce-Murcia, M. (2020). An overview of applied linguistics. In N. Schmitt and M. P.H. Rodgers (Eds.), An Introduction to Applied Linguistics (pp1-16) (3rd ed.). Routledge

Svensson, I., Nordström, T., Lindeblad, E., Gustafson, S., Björn, M., Sand, C., & Nilsson, S. (2019). Effects of assistive technology for students with reading and writing disabilities. Disability and Rehabilitation: Assistive Technology, 1-13 https://doi.org/10.1080/17483107.2019.1646821

Taylor, P. (2009). Text-to-Speech Synthesis. Cambridge University Press

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425-478DOI: 10.2307/30036540

Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Quarterly, 36(1) 157-178. https://doi.org/10.2307/41410412

Venkatesh, V., Thong, J. Y., & Xu, X. (2016). Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead. Journal of the Association for Information Systems., 17 (5), 328 – 376. DOI: 10.17705/1jais.0042

Walesiak, B. (2021). Mobile apps for pronunciation training: Exploring learner engagement and retention. In A. Kirkova-Naskova, , A. Henderson, & J. Fouz-González (Eds.) English Pronunciation Instruction: Research-based insights (pp. 357-385) (Vol. 19). John Benjamins Publishing Company.https://doi.org/10.1075/aals.19.15wal

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, v51 n1 p73-84. https://doi.org/10.1177/0022219416688170

Young, M. C., Courtad, C. A., Douglas, K. H., & Chung, Y. C. (2019). The Effects of Text-to-Speech on Reading Outcomes for Secondary Students With Learning Disabilities. Journal of Special Education Technology, 34(2), 80-91. https://doi.org/10.1177/0162643418786047