People's Democratic Republic of Algeria
Mohamed Khider University of Biskra
Faculty of Letters and Languages
Department of English Language and Literature

MASTER THESIS

Faculty of Letters and Foreign Languages

English Language
Sciences of the language

Submitted and Defended by:
KOUIDRI Halima Saadia

Assessing the Effectiveness of Text-to-Speech and Automatic Speech Recognition in Improving EFL Learner's Pronunciation of Regular Past-ed

the Case of A1 Beginner's Level at the Center of Intensive Training at the University of Biskra

---

Dissertation Submitted to the Department of Foreign Languages as Partial Fulfillment of the Requirements for the Degree of Master in Sciences of Language

Board of Examiners

Dr. Saihi Hanane	Examiner
Dr. Ilham Tigane	Supervisor
Miss Meriam Ghennai	President

Academic Year: 2022-2023

DECLARATION OF INTEGRITY

Supervisor: Dr. Ilham Tigane

Candidate: KOUIDRI HALIMA SAADIA

Specialty: Science of Language

Date: 07/06/2023

I, **KOUIDRI Halima Saadia**, solemnly declare that the dissertation entitled **"Assessing The Effectiveness of Text-to-Speech and Automatic Speech Recognition in Improving EFL Learners' Pronunciation of Past-ed".**The Case of A1 Beginner's Learners Level at The Center of Intensive Language Training at the University of Biskra .submitted to the Department of English Language and Literature at Biskra University is entirely my work, free from plagiarism, and has not been submitted to any other educational institution. I have appropriately acknowledged and cited all sources used, and I have conducted myself with academic integrity throughout the process. I understand the severe consequences of academic misconduct and affirm the authenticity of my dissertation.

Signature

Halima Kou

Dedication

To my dear mother Dalila, your love and support have guided me through life's ups and downs. I dedicate my achievements to you, for your unwavering belief in me has propelled me forward. You are my embodiment of love, compassion, and resilience. Thank you for being my pillar of strength.

To my father and my teacher Nacer, your dedication and strength have laid the foundation for my dreams. I am honoured to be your child and I dedicate my accomplishments to you. Your belief in my abilities has given me the courage to overcome any obstacle. Thank you for being my inspiration. You have ignited a flame of knowledge within me. Your guidance and encouragement have shaped my intellect. I dedicate my achievements to you for your impact on my educational journey.

To my dear sisters Amina, Khadija, Soumia, and Hiba, you bring joy and laughter to my life. I dedicate my love and support to each of you. May our bond continue to be a source of strength and cherished moments.

To my brothers Abdo, Mohamed, Oussama, and Ahmed, you are my closest allies. I dedicate my achievements to each of you for your unwavering support and shared experiences. Thank you for being there for me.

To my best friend Asma, your friendship has been my rock. I dedicate my accomplishments to you for your constant support and encouragement.

Lastly, to my departed grandparents, though you are no longer with us, your love and

wisdom continue to guide me. I dedicate my achievements to you, carrying your spirit as I journey through life. I hope to make you proud.

To my family, teachers, friends, and departed loved ones,

Your love, support, and inspiration have shaped my journey. I offer my heartfelt dedication to each of you. Thank you for believing in me and being part of my success.

Acknowledgements

I am thrilled to acknowledge the successful completion of this research work and would like to express my deep sense of gratitude to those who have contributed to its realization.

First and foremost, I am immensely grateful to my research supervisor, guide, and mentor, Dr Ilham Tigane, for granting me this invaluable opportunity to conduct this research. Her unwavering support, guidance, and expertise have been instrumental at every stage of this study. Her timely advice, suggestions, and provision of necessary resources have significantly facilitated the completion of my work. I truly appreciate the immense knowledge and advice she shared with me throughout this journey.

I would also like to extend my heartfelt appreciation to the members of the jury, Dr Saihi Hanane and Miss Meriam Ghennai, for their valuable time, insightful feedback, and constructive criticism, which have contributed to the refinement of this research.

Special thanks are also due to Mrs Diamonta and Miss Hadjer, who played significant roles in the study, as well as to all the participants from the CEIL center who generously dedicated their time and efforts to be part of this research. Their participation has been crucial in gathering meaningful data and insights. without which this study would not have been possible.

Lastly, I would like to express my gratitude to all those who have supported me during this research journey, including my family, friends, and colleagues. Their encouragement, motivation, and belief in my abilities have been a constant source of inspiration.

ABSTRACT

In recent years, there has been a growing interest in computer-assisted pronunciation training (CAPT) tools, especially with the integration of speech synthesis and automatic speech recognition (ASR) technologies in language learning applications. This quasi-experimental study aims to assess the effectiveness of text-to-speech (TTS) and ASR in improving the pronunciation of English past -ed among A1 EFL Learners level. The study used a pretest-posttest control-experimental group design to assess learners' pronunciation skills. A questionnaire was also conducted to evaluate participants' perceptions and attitudes towards the use of both tools. The results of the t-test analysis demonstrated significant improvements in the post-test scores of the experimental group compared to the control group, highlighting the positive impact of TTS and ASR technologies on learners' phonological awareness, perception, and production skills. These findings suggest that TTS and ASR have the potential to facilitate different stages of pronunciation development and can aid learners in acquiring desired pronunciation features. Additionally, the questionnaire responses revealed positive perceptions and high satisfaction with TTS and ASR technologies, affirming their effectiveness as valuable resources for pronunciation practice and self-assessment. This study underscores the pedagogical implications of integrating TTS and ASR technologies in pronunciation instruction, enhancing learners' pronunciation skills and expanding the learning environment beyond traditional classroom boundaries.

Key Words:

Computer-assisted pronunciation training (CAPT), Speech synthesis, Automatic speech recognition, Text to speech synthesis, Awareness, Perception, and Production.

LIST OF ABBREVIATIONS AND ACRONYMS

- ASR: Automatic Speech Recognition.
- CALL: Computer-Assisted Language Learning.
- CAPT: Computer-Assisted Pronunciation Training.
- EFL: English as a Foreign Language.
- MALL: Mobile-Assisted Language Learning.
- TAPT: Technology-Assisted Pronunciation Training.
- TTS: Text-to-Speech.

LIST OF FIGURES

LIST OF TABLES

List of Contents

General Introduction

Chapter I: Literature Review

Chapter II: Methodology

Chapter III: Results and discussions

This page intentionally left blank.

GENERAL INTRODUCTION

1. Background of the study

Pronunciation is a critical aspect of language learning, and English as a Foreign Language (EFL) is no exception. Accurate and clear pronunciation can enhance a learner's comprehension, fluency, and overall communicative competence, whereas poor pronunciation can lead to misunderstandings, confusion, and lack of confidence in language use.

EFL learners encounter several challenges when it comes to mastering English pronunciation. One of these challenges is the lack of exposure to the language, differences in the sound systems of their first language and English, limited opportunities for speaking practice and the lack of feedback on their pronunciation. Furthermore, EFL learners in Algeria face additional challenges due to the limited availability of English-speaking environments and native speakers of English. These challenges result in reduced opportunities to practice and improve their pronunciation skills.

To address these challenges, technology-based interventions such as Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) have been proposed as effective tools to enhance EFL learners' pronunciation skills. TTS technology converts written text into speech, which learners can use to improve their perception of English sounds. In contrast, ASR technology enables learners to receive immediate feedback on their pronunciation accuracy While previous research has shown the effectiveness of using either TTS or ASR, few studies have investigated the potential benefits of using both technologies together. Therefore, this research aims to address this gap in the literature and assess the effectiveness of using TTS and ASR in combination to improve EFL learners' pronunciation of the final "-ed"

2. Statement of the problem

Pronunciation is an essential aspect of language learning, especially in the English as a foreign language (EFL) context. The importance of pronunciation in EFL is to enable effective communication, both in personal and professional settings. However, EFL learners face several challenges when it comes to mastering English pronunciation. These challenges include the lack of exposure to native speakers, differences in the sound systems of their first language and English, and limited opportunities for speaking practice. Among the specific difficulties faced by EFL learners is the accurate pronunciation of the final "-ed" in English words, which has three different pronunciations - /t/, /d/, or /ɪd/ - depending on the preceding sound.

With the development of information technology, a plethora of speech tools were developed to assist students in improving their pronunciation. Text-to-Speech Synthesizers (TTS) and Automatic Speech Recognition (ASR) are two speech technologies that can promote input and output practice. Accordingly, little research has been conducted to investigate the use of both tools combined to help learners acquire target L2 features. According to Levis and Suvorov (2013), the potential of combining ASR and text-to-speech software has not been fully investigated, but it holds promise for improving non-native speech in various contexts. Similarly, Liakin, et al.(2017) suggest that there has been limited exploration of the pedagogical applications of TTS and ASR in second language (L2) education. Nevertheless, the existing studies demonstrate positive results when these technologies are employed as additional resources in classroom instruction In a specific investigation conducted by Khademi (2021), significant improvements were observed in participants' awareness and perception of English past -ed allomorphy through the utilization of Google Translate's speech capabilities, TTS, and ASR. However, when it came to production, improvements were only evident for the /id/ sound. It is important to note that this study has certain

limitations, including the absence of a control group and the relatively short duration of the study, which lasted for only two hours. Additionally, the attitudes of EFL learners towards TTS and ASR as tools for improving their pronunciation have not been fully explored.

Therefore, to address these gaps this study aims to assess the effectiveness of using TTS and ASR in combination to improve EFL learners' pronunciation of final regular past tense marker-ed allomorphy in terms of awareness, perception, and production -three stages in pronunciation development (Celce-Murcia et al., 2010), using a pretest-posttest control and experimental group study as well as to examine learners' attitudes towards the use of these technologies.

In this regard, the present study aims first, at reviewing some literature about EFL pronunciation, and the use of TTS and ASR to improve pronunciation, and then we empirically test the effectiveness of combining the above mentions tools in one application called “Neural Reader” In addition, this research examines the participants’ attitudes toward the use of TTS and ASR tools.

3. Research questions

1- What is the impact of integrating text-to-speech (TTS) and automatic speech recognition (ASR) technologies on the awareness, perception, and production of the final "-ed" sound among EFL learners?

2- What are the learners’ attitudes towards the use of ASR and TTS for improving their pronunciation?

4. Research Hypothesis

Based on the above research questions, we hypothesize that if learners used TTS and ASR as pronunciation learning tools:

- Ho: There is no significant impact on students' scores after utilizing TTS and ASR between the control group and the experimental group.
- Hi: Learners who utilize TTS and ASR as pronunciation learning tools will demonstrate significantly higher scores compared to the control group who does not use these tools.

5. Research aims

The study aims to investigate the effectiveness of TTS and ASR in improving EFL learners' awareness, perception, and production of past "ed" pronunciation. Whereas objectives of the present study are:

1. To determine the attitudes and perceptions of EFL learners towards the use of TTS and ASR for pronunciation training.
2. To explore the potential benefits and limitations of using TTS and ASR for pronunciation training in an EFL context.

6. Methodology

The present research employed a pretest-posttest quasi-experimental design with a control group. This design allows us to compare the effectiveness of the intervention (text-to-speech and automatic speech recognition) on the experimental group's pronunciation of past-ed with the control group who will not receive any intervention. The pretest and posttest data were collected through a jotform platform, which allows participants to listen to audio and record their audio responses. Additionally, a semi-structured questionnaire was administered to explore the attitudes and perceptions of the experimental group towards the use of text-to-speech and automatic speech recognition in improving their pronunciation.

6.1 Sample

The participants of the present study are 28 students of A1 level at the center of intensive training at Mohamed Kheider University. EFL learners were randomly assigned to either an experimental or a control group of 14 participants in each.

6.2 Data Collection Methods

The data collection methods for this study involved a pretest, a posttest, and a questionnaire. The pretest and posttest were used to compare the effectiveness of the TTS and ASR intervention in improving pronunciation of regular past "ed." The pretest established the baseline pronunciation skills, while the posttest measured any improvements after the intervention. The control group served as a comparison group, not receiving any intervention. Both the pretest and posttest were administered through a jotform platform. The questionnaire gathered information on learners' attitudes and perceptions towards using TTS and ASR technology for improving pronunciation skills. The questionnaire included close-ended questions, such as the Likert scale and multiple-choice questions, to collect quantitative data on learners' experiences and perceptions of TTS and ASR technology for pronunciation learning tools.

6.3 Research Procedures

During the study, the researcher gained access from the center, both teachers and students from both the experimental and control groups. The research took place over a period of eight days, from April 25th to May 3rd, 2023. To collect the necessary data, several procedures were implemented. Firstly, a pretest was conducted with both groups to establish a baseline for comparison. Then, the experimental group received presentations on Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) technologies, along with instructions on how to use the application effectively. In the classroom, the researcher introduced the application and guided the

students in its usage. During the treatment phase, participants independently utilized the application, with the researcher providing support through an online group. A follow-up session was conducted to ensure familiarity with the application, and a posttest was administered, following the same procedure as the pretest. Additionally, participants completed a questionnaire to gather further information.

6.4 Data Analysis

The data collected in this quasi-experimental research study were analyzed using descriptive and inferential statistical techniques. The descriptive analysis provides an overview of participants' initial and final performance and assesses consistency or variability of pronunciation improvement within grouThe inferentialtial analysis included an independent t-test to compare pre-test and post-test scores between the control and experimental groups, aiming to identify significant differences in pronunciation improvement. The questionnaire data were analyzed descriptively to summarize participants' responses and gain insights into their perceptions of the effectiveness of TTS and ASR technologies. SPSS 23 software was used for all data analyses, and a significance level of 0.05 was set for inferential statistics.

7. Significance of the study

This study shed light on the use of TTS and ASR software in teaching English pronunciation. The use of this software can enhance pronunciation practice owing to the comprehensive input provided by TTS and the opportunities to practice pronunciation and receive feedback via ASR

The principle of this research is that phonetic training improves pronunciation skills. Thus, theoretically, this research is expected to extend and prove that phonetic training using TTS and

ASR is an effective way to improve pronunciation skills. Practically, this research is expected to be useful for English teachers as a way of using speech synthesis in teaching English, especially in pronunciation. For students, this study will help them to realize the importance of making benefits from technology to help them improve their pronunciation. Then, this study is expected to help them to learn pronunciation more efficiently and correctly.

Therefore, the significance of the study is to provide valuable insights and recommendations to English language teachers and learners on the effectiveness of using Text-to-Speech and Automatic Speech Recognition technologies in improving EFL learners' pronunciation of regular past “ed”. The study can contribute to the existing literature on the use of technology in language learning and teaching, particularly in the area of pronunciation instruction. Additionally, the study may also help improve the quality of EFL education and enhance learners' communicative competence.

8. Organization of the Study

The dissertation opens with General Introduction Chapter which provides a preliminary hint about the whole work, its background, its significance and objectives, and its organization. Moreover, it introduces the research questions that are to be addressed, and it presents a brief description of the steps that are followed to achieve the objectives of the study

Chapter One provides the readers with a literature review which presentes the theoretical work of major studies in the field of teaching pronunciation. It also highlights some practical investigations relevant to the subject area. This chapter presents studies about the use of technological tools namely text-to-speech and automatic speech recognition and their affordance to teaching and learning L2 pronunciation.

Chapter Two serves as the methodology chapter in this study, focusing on the fieldwork

and analysis of the collected data. The chapter then delves into the applied research methodology, describing the specific procedures, techniques, and instruments utilized in data collection. It outlines the pretest, posttest, and participants' questionnaire as the primary data sources.

Chapter Three is dedicated to presenting and analyzing the results of this study. It provides an overview of the collected data, presents the key findings, and discusses their pedagogical implications.

CHAPTER ONE

Literature Review

Introduction

Pronunciation is a crucial aspect of learning a new language, and English as a Foreign Language (EFL) is no exception. Accurate pronunciation is essential for effective communication and can enhance a learner's overall communicative competence. However, mastering accurate pronunciation in a second language is often hindered by multiple factors including unfamiliar phonetic patterns and sounds, limited exposure to the target language, as well as insufficient opportunities for practice and to receive feedback. In response to these challenges, Technology-Assisted Pronunciation Training (TAPT) has emerged as a promising approach. TAPT capitalizes on advanced technologies like Text-to-Speech (TTS) and Automatic Speech Recognition (ASR). Text-to-speech and Automatic speech recognition have been proposed to overcome the shortcoming of traditional settings by providing learners with comprehensible input through TTS and ample opportunities to practice pronunciation in addition to personalized feedback via ASR. TTS software converts written text into spoken language, allowing learners to listen to and imitate native-like pronunciation. ASR technology, on the other hand, provides learners with real-time feedback on their pronunciation accuracy

The first section of the present literature review provides an overview of English pronunciation, highlights the challenges faced by EFL learners in mastering pronunciation, and examines pedagogical approaches to English pronunciation. The second section is dedicated to the use of Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) technologies in pronunciation training and reviews the literature on the use of TTS and ASR in EFL pronunciation.

Section One

1.1 Definition of Pronunciation

Clear and accurate pronunciation plays a vital role in effective communication as it directly influences the clarity and understanding of spoken language. Pronunciation encompasses the way words are produced and the sounds they generate. When a speaker mispronounces words, it can create barriers to communication and lead to misunderstandings. Therefore, recognizing the significance of proper pronunciation is crucial for maintaining successful and meaningful communication.

Hornby (1995, p.928) provides a comprehensive definition of pronunciation as "the manner in which a word is said, the manner in which a word is pronounced, or the manner in which a person speaks a language." This definition emphasizes the importance of how words are spoken in determining their correct pronunciation. Similarly, According to Robinett (1981), pronunciation can be defined as “the process of generating speech sounds, encompassing articulation, vowel formation, accent, inflexion, and intonation. It often pertains to the accuracy or acceptability of the produced speech sounds” (p. 64). These elements are often considered in relation to the correctness or acceptability of the speech sounds.

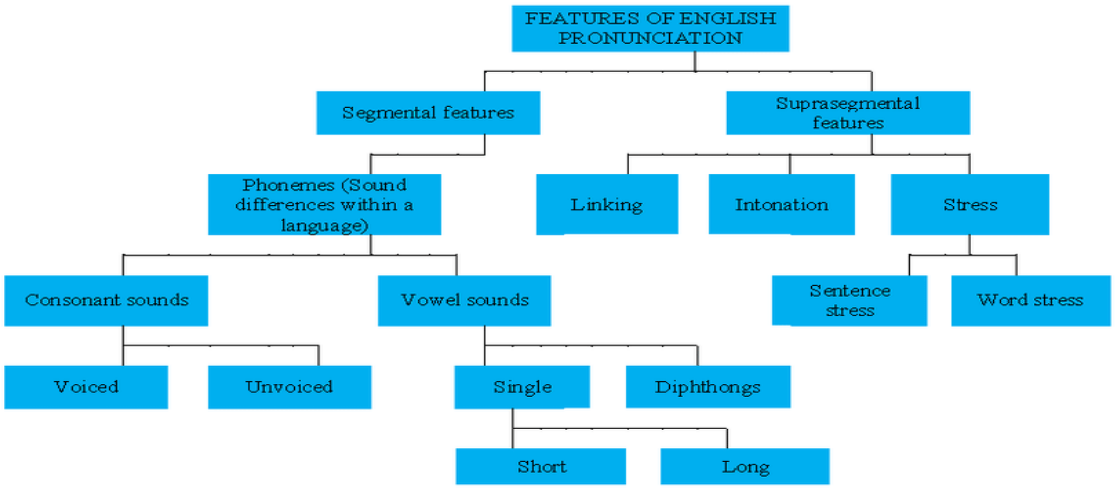
According to these definitions, pronunciation encompasses not only the physical production of speech sounds but also the overall manner in which a person speaks a language. It includes factors like accent and intonation, which impact the clarity and comprehensibility of speech. In essence, clear and accurate pronunciation is crucial for effective communication, facilitating the smooth transfer of meaning between speakers and listeners. In essence, pronunciation can be defined as the way in which a person speaks or makes a word sound, with the aim of being easily understandable to others.

1.2. Features of English pronunciation

Segmental and suprasegmental features are two categories used to classify the components of English pronunciation. Scholars define English segmental features as discrete sounds, such as vowels and consonants, which convey the meanings of words. Suprasegmental elements, such as intonation, stress, and pitch, provide additional information about the speaker's intended meaning and can help distinguish between different types of sentences and convey emotions (Most & Peled, 2007; Yenkimaleki & Van Heuven, 2021)

In pronunciation teaching, there is ongoing debate concerning whether segmental or suprasegmental features are more significant (Wang, 2020). Some researchers have emphasized the significance of suprasegmental elements in improving speaking skills, claiming that if L2 pronunciation teachers prioritize suprasegmental aspects of English, learners' comprehensibility will improve (Yenkimaleki & Van Heuven, 2021). nevertheless, most studies have focused on the perception of segmental aspects, with little emphasis on the perception of suprasegmental features (Most & Peled, 2007).

Figure 1: *Various Features of English Pronunciation* (Adapted from Gilakjani 2012, p.120).



As depicted in the figure above, pronunciation encompasses features at both the segmental (micro) and suprasegmental (macro) levels. At the segmental level, which focuses on individual sounds, phonemes are divided into consonants and vowels. Consonants can be further classified as either voiced or unvoiced, while vowels can be categorized as single vowels or diphthongs, which are combinations of two vowel sounds.

Moving on to the suprasegmental level, which involves features that extend beyond individual sounds, there are two types of features: intonation and stress. Intonation refers to the pitch pattern and melody of speech, while stress pertains to the emphasis placed on certain syllables or words within an utterance, encompassing both word stress and sentence stress. A more detailed explanation of the segmental features is stated as follows:

1.2.1 Segmental Features of Pronunciation

English segmental features include vowels, diphthongs, and consonants. The categories are as follows:

1.2.2 Phonemes

Phonemes are defined as "the smallest unit in a language's sound system" (Crystal,2008,p.361). They are the smallest unit of speech sounds that can differentiate between words. Not all languages have the same number of phonemes. for instance, there are 44 phonemes in English (24 consonants and 20 vowels).

1.2.2.1 Vowels

The vowel is a sound produced with a free passage. A free passage here means that vowel sounds are produced without obstruction. English vowels are divided into two kinds, which are long vowels and short vowels. Long vowels consist of the following /iː/, /ɜː/, /aː/, /uː/, /ɔː/, while short vowels consist of /ɪ/, /e/, /æ/, /ə/, /ʌ /, /ʊ/,/ɔ/.

Some processes are responsible for the production of vowels, those are the shape of the lips, the opening between the jaws, the position of the soft palate, and the shape of the tongue (Williamson, 2018). Vowels are described in terms of height, backness/frontness, and roundedness.

1.2.2.2 Consonants

Consonants are speech sounds which are produced with some degree of constriction or obstruction of airflow in the vocal tract. There are 24 consonants in English, and they play a crucial role in shaping the sounds of the language.

According to Proctor (2021), Consonants can be further categorized based on their voicing. Voiced consonants are those that produce a vibration in the vocal cords when pronounced by the speaker. This vibration creates a distinct quality in the sound. Examples of voiced consonants in English include /b/, /d/, /dʒ/, /g/, /j/, /l/, /m/, /n/, /r/, /v/, /ð/, /y/, /z/, /ʒ/, and /ŋ/. On the other hand, unvoiced consonants are those that do not produce any vibration in the vocal cords. They are pronounced without any vocal cord movement, resulting in a different quality of sound. Examples of unvoiced consonants in English include /f/, /p/, /t/, /ʧ/, /k/, /θ/, /s/, and /ʃ/. One way to determine if a consonant is voiced or unvoiced is to place a finger on the throat and check for vibrations while pronouncing the sound. If there is a noticeable vibration, the consonant is voiced; if there is no movement in the neck and only a brief blast of air, the consonant is unvoiced.

Understanding the distinction between voiced and unvoiced consonants is important in phonetics and phonology, as it can have implications for the meaning and pronunciation of words in different languages. For example, in English, the voicing distinction is phonemic, meaning that changing the voicing of a consonant can result in a change in meaning. For instance, the pair of words "bat" and "pat" differ only in the voicing of the initial consonant, which changes the meaning

from one word to another. Consonants are an essential part of the segmental features of pronunciation in English, and understanding their voicing distinction is crucial in studying phonetics and phonology.

1.3. The Rules of Regular Verbs in Past Tense

There is a consensus among authors (Celce-Murcia et al., 2010, Fraizer and Mills, 2015, Schoenberg, 2016) regarding the rules for pronouncing the -ed ending of regular verbs in the simple past tense. According to these authors, the rules are as follows:

Rule number one states that if the verb base ends in a voiced sound (excluding /d/), the -ed ending should sound like /d/. The /d/ sound should blend with the previous consonant and not be pronounced as an additional syllable. A voiced sound is characterized by the vibration of the vocal cords and includes consonants such as /b/, /v/, /g/, /z/, /j/, /ð/, /l/, /m/, /n/, /r/, and all the vowels.

The second rule explains that if the verb base ends in a voiceless sound (excluding /t/), the -ed ending should sound like /t/. The /t/ sound should blend with the previous consonant and not be pronounced as an additional syllable. A voiceless sound is characterized by the absence of vibration of the vocal cords and includes consonants such as /p/, /f/, /k/, /s/, /ʃ/, /tʃ/, /ʧ/, /θ/.

Finally, rule three specifies that if the verb base ends in a /t/ or /d/ sound, the -ed ending should sound like /ɪd/ or /əd/ and should be pronounced as an additional syllable. Table 1 provides examples related to these rules.

According to DeCapua (2017), the most common pronunciation mistake made by EFL learners when using the past tense is mispronouncing the "-ed" ending of regular verbs. This is because the "-ed" ending can be pronounced in three different ways depending on the final sound of the verb.

1.4. EFL learners' challenges with "ed-endings" Pronunciation

Pronunciation is an essential aspect of language learning, and it can be particularly challenging for non-native English-speaking students. Among the most common pronunciation errors made by EFL learners is mispronouncing the "-ed" ending of regular verbs in the past tense. This is because the "-ed" ending can be pronounced in three different ways, leading to confusion and incorrect pronunciation.

Lluïsa (2015) claims that students struggle with the correct pronunciation of "–ed endings" and tend to pronounce them as written. Lluïsa added that EFL students often fail to recognize the distinction between the /d/, /t/, and /ɪd/ sounds, which can lead to incorrect pronunciation. This is a common error in language learning and requires proper guidance from teachers to help students master this aspect of pronunciation.

Recent research by Wahyuni et al. (2021) discusses the difficulties faced by non-native English-speaking students in pronouncing words with -ed endings (both adjectives and regular past tense). Data collection and analysis were done by the researchers using a qualitative approach that involved interviews and observation. The study's findings revealed that the majority of students had trouble pronouncing words with -ed endings, in addition to problems in differentiating between the final -id, -d, or -t sounds. The researchers identified several factors contributing to this problem, including a lack of knowledge of English pronunciation rules, a lack of practice, and a lack of motivation to improve pronunciation skills.

In response to these challenges, Benitez-Correa et al. (2020) suggest the use of technology and software such as Audacity to improve EFL learners' pronunciation. Audacity is a software tool commonly used for analyzing and studying speech sounds. It allows researchers to visualize and study waveforms and spectrograms to explore phonetic properties and

analyze phonological patterns. Further, the results suggest that using the software was an effective tool not only to improve learners' pronunciation but also to practice the English language in a free and relaxed environment.

Overall, the studies conducted by DeCapua (2017), Lluïsa (2015), and Wahyuni et al. (2021) collectively highlight the challenges encountered by EFL learners in correctly pronouncing words with "-ed" endings. These challenges can be addressed through various strategies, including providing clear explanations of pronunciation rules, offering ample practice opportunities, and integrating technology tools like Audacity. By implementing these approaches, teachers can play a crucial role in supporting their students in overcoming pronunciation difficulties and ultimately enhancing their overall proficiency in English. The findings from these studies offer valuable insights for educators seeking effective methods to improve pronunciation skills among EFL learners.

1.5. Approaches to teaching pronunciation

In the field of modern language teaching, three approaches to pronunciation instruction are generally proposed: The Intuitive-Imitative approach, the analytic-Linguistic Approach and the integrative approach. The above-mentioned approaches are associated with different methods of language teaching.

1.5.1 The Intuitive - imitative Approach

An Intuitive-Imitative Approach depends on the learner's ability to listen to and imitate the rhythms and sounds of the target language without the requirement for explicit instruction. As Hismanoglu and Hismanal (2010) describe the Intuitive Imitative Approach as "a student's ability to listen to and imitate the rhythms and sounds of the target language will give rise to the development of an acceptable threshold of pronunciation without the intervention of any explicit

information" (p. 984). Furthermore, the intuitive-imitative approach refers to the student's capacity to mimic the sounds and rhythm of the target language. The learner employs some specific technologies like audiotapes, videos, and computer-based programs to emulate the uttered sound of the target language.

1.5.2 The Analytic-Linguistic Approach

An Analytic-Linguistic Approach, on the other hand, integrates listening, imitation, and production with information and tools like a phonetic alphabet, articulatory descriptions, vocal apparatus charts, contrastive information, and other aids. It explicitly informs and focuses the learner's attention on the target language's sounds and rhythms. In this approach, the learners are provided with explicit information on pronunciation using the phonetic alphabet, articulatory descriptions, and vocal charts. The explicit information can be presented in different interactive speech software and websites (Lee, 2008). An Analytic-Linguistic Approach was developed to complement the intuitive-imitative approach instead of replacing it (Celce-Murcia et al.,1996) since this approach emphasizes omitation and listening; nevertheless, it introduced the use of various aids such as the phonetic alphabet.

1.5.3The integrative approach

The integrative approach, as described by Lee (2008, p. 1), recognizes pronunciation as an integral component of communication rather than a standalone drill or practice sub-skill. In this approach, Lee emphasizes that learners should engage in meaningful task-based activities to practice pronunciation. This approach entails a dual-focus oral communication program that addresses both micro-level linguistic competence, including phonetic and phonological aspects, through segmental practices and suprasegmentals, and macro-level communicability, which involves the development of sociolinguistics, discourse, and strategic competence. The ultimate

goal of this approach is to achieve communicative goals while fostering enjoyment of the language (Hismanoglu & Hismanoglu, 2010).

The above-mentioned approaches to teaching pronunciation incorporate both traditional methods and modern techniques. In the intuitive-imitative approach, learners are encouraged to listen to and imitate the rhythms and sounds of the target language without explicit instruction. This approach often utilizes technologies such as audiotapes, videos, computer-based programs, and websites to facilitate the learning process. Learners can immerse themselves in the language and imitate native speakers to develop their pronunciation skills.

On the other hand, the analytic-linguistic approach provides learners with explicit information about pronunciation, such as phonetic alphabet, articulatory descriptions, and vocal charts. This approach aims to help learners understand the underlying principles of pronunciation and how different sounds are produced. Today, interactive speech software and websites are available to present this explicit information in a variety of engaging ways.

The integrative approach, views pronunciation as an essential component of communication, rather than a separate drill or practice sub-skill. It acknowledges the importance of integrating pronunciation with other language skills and emphasizes meaningful communication through task-based activities. This approach recognizes that pronunciation is not just about producing accurate sounds, but also about using those sounds effectively in real-life communication situations. Thus, the integrative approach considers pronunciation as an integral part of overall language proficiency and promotes its integration into the language learning process.

1.5.4 Community Language Learning

In the realm of pronunciation instruction, Celce-Murcia et al. (1996) explain this approach

by stating that the pronunciation syllabus was mostly student-initiated and designed. Several tools and techniques are essential in the treatment of pronunciation in Community Language Learning such as the audiotape recorder and the human-computer technique. including the audiotape recorder which not only captures what is said in the student-generated utterances but also allows students to distance themselves from what was said and focus on how it was pronounced. Thus, allowing them to compare their pronunciation with that of the teacher. The human-computer technique does not provide explicit pronunciation correction and enables the learner to begin pronunciation practice by selecting the item(s) to practice and determining the quantity of repetition required. Students can thus approximate the desired pronunciation to the extent that they choose.

1.5.5 Communicative Language Teaching (CLT)

The Communicative Approach in the 1980s was dominant in language teaching. It holds that since the primary purpose of language is communication, using language to communicate should be central in all classroom language instruction.V arious techniques used to teach pronunciation in this approach include listening and imitating, phonetic training, minimal pair drills, contextualized minimal pairs, visual aids, tongue twisters, developmental approximation drills, practice of vowel shifts and stress shifts related by affixation, reading aloud/recitation, and recordings of learners' production. In a CLT classroom, teachers are encouraged to employ a variety of audio-visual tools. Because the resources are mostly authentic, there are numerous ways to incorporate photographs and videos into the classroom to make it more fascinating and authentic. Typically, class activities are focused on activities connected to real-world communication (Freeman, 2000).

1.6. The Significance of Pronunciation in Language Learning: Focusing on Intelligible Communication

Proper pronunciation is an indispensable aspect of successful communication, reflecting a crucial direction in current foreign language teaching. In line with the goal of language learning, which is communication in both written and oral forms, achieving intelligible pronunciation, rather than a native-like accent, is emphasized (Celce-Murcia et al., 2010). The importance of intelligible pronunciation in successful oral communication has been supported by researchers such as Derwing and Munro (1997) in addition to Levis and Suvorov (2013). Both learners and teachers acknowledge its significance in language learning (Grim & Sturm, 2016). Thus, without proper pronunciation training, learners may experience misunderstandings and communication breakdowns due to a lack of mutual intelligibility.

Therefore, Celce-Murcia et al. (2010. p,45) propose that students should engage in several learning activities with at least five phases that progress from sound analysis and awareness raising to listening discrimination and production. These phases include (1) description and analysis of the target feature; (2) listening discrimination; (3) controlled practice; (4) guided practice; and (5) communicative practice. The authors suggest that these phases should be implemented over a number of lessons. However, the main challenge in implementing Celce-Murcia et al.'s (2010) recommendations for achieving appropriate levels of intelligibility is the issue of time.

Pronunciation is often considered one of the challenging skills in second language teaching and learning. Achieving intelligible pronunciation requires extensive practice in listening (input) and speaking (output), preferably with corrective feedback (Saito & Lyster, 2012). Nevertheless, the limited classroom time for EFL teachers and students has always been a concern, potentially hindering phonological acquisition in terms of awareness, perception, and production

(Collins & Muoz, 2016).

1.6.1 Phonological Awareness

Insufficient instruction and practice in pronunciation can hinder English as a Foreign Language (EFL) learners' ability to notice important aspects of the target language (L2) and impede their language development (Svalberg, 2012). Language awareness, which involves consciously perceiving and understanding certain aspects of a second language, plays a crucial role in language acquisition. Specifically, language awareness encompasses explicit knowledge about language and a conscious sensitivity in language use. From a phonological perspective, awareness can be defined as "one's level of sensitivity to the sound structure of spoken language" (Anthony & Francis, 2005, p. 255).

According to Carlet and Kivistö-de Souza (2018, p. 104), L2 phonological awareness “can be developed through any activity that brings a specific aspect into the language learners’ consciousness” The researchers also present some instances of awareness-raising practices, such as explicit comparisons of L1-target language phonologies, input enhancement, and feedback approaches. In addition, it is argued that raising learners' awareness of target language phonology not only positively reflects on their L2 pronunciation but also allows them to take control of their pronunciation learning by developing self-monitoring abilities.

Linebaugh and Roche (2015) emphasize the importance of incorporating both input and output practice in pronunciation training to develop awareness and enhance perception skills. When learners engage in input practice, they focus on listening to and comprehending the target language sounds produced by native speakers. This exposure to authentic pronunciation models helps learners become familiar with the sounds, rhythm, and intonation patterns of the language. On the other hand, output practice involves actively producing the target language sounds

themselves. By actively articulating the sounds and receiving feedback on their pronunciation, learners gain a deeper understanding of the phonetic features and nuances of the language. This combined approach of input and output practice helps learners develop a heightened awareness of the target language's phonological features, enabling them to perceive and discriminate sounds more accurately. Improved perception skills allow learners to recognize and interpret spoken language more effectively, contributing to overall language comprehension and communication proficiency.

1.6.2 Perception: Aural Discrimination

Perception refers to a speaker's ability to distinguish between sounds they heard (Soler-Urza, 2011). Research has shown that learners' perception is affected by the quantity and quality of input they receive (Flege, 1991). Thus, Linebaugh and Roche (2015) recommend that learners engage in focused aural exposure to L2 sounds, or input, to improve their ability to discriminate them. The input concept is directly related to perception since learners cannot develop their perceptual skills unless they have access to the target language.

In order to improve the perception of sounds or phonemes, Liakin et al. (2015) suggest employing strategies such as slow speaking, repetition, and other input improvement techniques. However, as previously discussed, insufficient practice and/or instructions as a result of insufficient time committed to pronunciation in the classroom In addition, the scarce exposure to the language outside the classroom leaves learners with insufficient time for input (and hence perceiving) practice. Assuming that perception precedes oral production (Celce-Murcia et al., 2010; Flege, 1995), the insufficient input practice may hinder learners' aural discrimination skills and subsequently impacts their ability to accurately produce the target language sounds. the lack of practice and exposure can lead to difficulties in aural discrimination (perception) and

subsequently impact oral production skills.

1.6.3 Production: Output and Corrective Feedback

While input is important in the development of L2 pronunciation, production (or output) is equally significant in language development. According to Swain (1999), output practice can help learners acquire autonomy, noticing abilities, and hypothesis testing, in which learners use language production to verify their assumptions about the language being taught in order to understand how it works. L2 learners cannot practice the language or improve their linguistic skills if they do not have opportunities to produce it.

Soler-Urzúa (2011) defines oral production (output) as "the ability to appropriately pronounce the target sound" (p. 50) and hypothesizes that enhancing the auditory input in the L2 (e.g., via slow speech, repetition) might lead to an improvement in oral production because this move provides L2 learners with ample opportunities to perceive and process the target L2 features. Accordingly, output practice can help learners develop automaticity, noticing abilities (including awareness), and hypothesis testing, allowing learners to understand how the target L2 phonological system functions (Swain, 1999).

Many scholars claim that corrective feedback can benefit EFL learners in both perception and production of pronunciation (Baker & Burri, 2016; Couper, 2019) by directing learners' attention to the form and consequently helping them notice the differences between their own output and what they should produce (Darcy, 2018).

Similarly, Darcy (2018) emphasizes that feedback is a predictor of self-awareness pronunciation improvement, primarily because it alerts the learner to specific difficulties as they occur. However, because of the many sorts of feedback, Darcy suggests that explicit feedback is preferred when pronunciation characteristics are taught as an integrated component of a lesson. As

a result, explicit feedback assists in clarifying that the correction is about a specific form rather than meaning. It is feasible to focus the learners' attention on their production in comparison to what they were supposed to produce in this manner. In response, students might concentrate on monitoring their pronunciation to achieve more understandable and comprehensible speaking.

In general terms, input is referred to the language incorporated in the communication contexts to which learners are exposed during their learning process (Vanpatten,2008 ; Benati, 2017). In particular, input is required for the development of learners' perceptual skills since it provides the phonological data needed to recognize L2 sounds.

Nevertheless, language input in instructional language learning environments is somehow constrained in terms of quality or/and quantity. According to Collins and Muñoz (2016) exposure to the target language is limited to short class sessions and is mostly provided by the teacher and peers, who typically have the same L1. As a result of the lack of frequent language input and limited opportunities for practicing oral skills (Carlet; Kivist-de Souza, 2018), L2 learners may not be able to properly improve their L2 pronunciation skills depending only on the limited contact with the target language provided in class.

Considering the discussion above, pronunciation aspects should be explicitly taught in the learning process, along with exercises aimed at strengthening learners' awareness of the specific forms and their own oral performance (Darcy,2018; Derwing,2018). According to Darcy (2018), the effectiveness of pronunciation instruction relies on three main components. These components include explicit and communicative exercises, attention to perception, and explicit feedback. Through explicit and communicative exercises, learners receive clear instructions on producing specific sounds or phonetic features of the target language, engaging in meaningful communication to develop a deeper understanding of pronunciation rules

and patterns. Attention to perception involves training learners to recognize and discriminate between different sounds, enhancing their ability to accurately produce the target language's features. Additionally, explicit feedback plays a crucial role, providing learners with specific and constructive input on their pronunciation performance. By incorporating these three components, pronunciation instruction can offer a comprehensive and structured approach to improving learners' pronunciation skills (Darcy, 2018).

As a result, explicit instruction combined with practice opportunities will help learners improve speech production, and given the importance of pronunciation to spoken communication and pronunciation teaching to SLA, all resources available to help achieve the goal of pronunciation development are supported. As discussed in the following section, digital tools can be valuable to both teachers and students in this regard.

In conclusion, the first section of this literature review provided an overview of English pronunciation, highlighted the challenges encountered by EFL learners in mastering pronunciation skills, and examined pedagogical approaches for teaching English pronunciation. The literature suggested that integrating technology, specifically speech synthesis, offered potential solutions to overcome these challenges. Moving forward, the subsequent section delved into the specific use of Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) technologies in EFL pronunciation instruction

Section Two

Introduction

In the age of advancing information technology, a multitude of speech tools have emerged as valuable resources for assisting students in improving their pronunciation skills. Among these tools, Text-to-Speech Synthesizers (TTS) and Automatic Speech Recognition (ASR) technologies

have gained considerable attention for their potential to facilitate input and output practice. Therefore, the second section of this study focuses specifically on the use of Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) technologies in the realm of pronunciation training within the context of English as a Foreign Language (EFL). This section aims to provide a comprehensive review of the existing literature on the application of TTS and ASR technologies in EFL pronunciation instruction.

2. Text-to-Speech Technology: An overview

In her book English Language and Technology, Chapelle (2003) supports the idea that computer-assisted language learning (CALL) can provide L2 learners with opportunities to receive enhanced input, interact with and produce the target language, all of which are known to be crucial for language acquisition. Text-to-speech (TTS) technology has emerged as a notable tool due to its remarkable capacity to offer enhanced language input.

Text-to-Speech (TTS) is a computer program that converts text input into speech output. Handley (2013, p. 5846) defines TTS as "the process of making the computer talk". In other words, TTS allows text input on personal computers or mobile devices to be converted into an oral version of speech. These programs were initially created for individuals with visual impairments or disabilities, as they enable users to listen to spoken language on a computer. Nowadays, TTS technology is widely used in applications such as desktop voice systems, audiobooks, and electronic dictionaries. Many modern computers come with integrated TTS functions, such as Apple's Siri, Amazon's Alexa, and Google Translate.

There are various TTS programs available, ranging from free to paid versions, which may differ in their functions. Updated versions often offer voice options, such as male or female voices, native or non-native accents, and different language varieties. They may also provide more user

interaction, such as highlighting each word being read aloud, and better access to other types of files, such as PDFs, e-books, and web documents.

According to Dutoit (1997), Text-to-Speech (TTS) systems consist of two main modules: Natural Language Processing (NLP) and Digital Signal Processing (DSP). The former is responsible for creating a voice recording by combining tone and intonation, while the latter converts the received symbol information into speech. Whereas, Zhu (2011) provided a more detailed explanation of the four primary modules of TTS. The initial model is Textual Analysis, which analyzes the syntax and meaning of the text in order to convert it into language-specific parameters. The subsequent model, known as the Rhythm Generator, generates the appropriate rhythm for each syllable in the given text. Following that, the Synthesis Unit Generator generates the synthesized unit by utilizing speech waveform samples of monosyllabic phonemes from the speech database. Lastly, the Text-to-Speech Synthesizer selects acoustic parameters from the sound database to accurately reproduce the desired sounds.

Text-to-speech (TTS) technology, an assistive technology, which employs artificial intelligence (AI), converts written information in a human-readable format from one language into audio or speech output, mimicking human accent. These systems utilize AI-driven algorithms as input to convert text into audible speech, often referred to as "real aloud technology" as it reads text aloud (Scott, 2022). This technology has gained attention from scholars as a potential educational tool, particularly in L2 classrooms, due to its potential benefits for various aspects of language learning. The following section presents different studies examining the efficacy of TTS technology as a pedagogical learning tool in L2 context representing related literature concerning the use of text to speech to improve L2 pronunciation.

2.1. Text-To-Speech Evaluation: Speech Quality Improvement

The issue of whether the quality of TTS synthesis is adequate for the L2 context has been raised in the Computer Assisted Language Learning (CALL) and Mobile Assisted Language Learning (MALL) research . Therefore, a crucial first step in evaluating synthetic speech is to identify the differences between synthetic and natural speech. In other words, how do synthetic voices compare to voices produced by humans? In general, previous studies have focused on listeners' reactions to non-native speech to evaluate L2 speakers' pronunciation. For instance, Derwing and Munro (2005) proposed three dimensions of oral speech: (1) comprehensibility, or the difficulty in understanding an utterance; (2) intelligibility, or the extent to which a message is understood by an interlocutor or group of listeners; and (3) accentedness, or the difference between an L2 accent and the L1, including the variations in accents that characterize native or fluent speech.

Over the past two decades, there have been a few evaluations of TTS systems and their voices. The preferred strategy has been to evaluate TTS and human speech using the criteria mentioned above. Nusbaum, Francis, and Henly (1995) compared TTS-produced English voices to their human counterparts for naturalness in both segmental and suprasegmental aspects. The findings indicated that even when the intelligibility variable was removed, participants judged human voices to be more natural than TTS. Supporting these results, Stevens, Lees, Vonwiller, and Burnham (2005) found that native English-speaking participants judged TTS sentences as less natural than human-produced sentences.

In another study that focused on gauging intelligibility using a French TTS system, Bailly (2003) discovered that participants performed better in shadowing tasks when they had human voice input rather than TTS-produced input. Interestingly, in a later study by Kang et al. (2008)

found no significant differences in text comprehension between participants' abilities to understand human and TTS voices.

Based on the existing studies, it appears that previous research has produced inconsistent results when comparing the quality of TTS systems to the human voice. One cause of this disparity is the adoption of incompatible or equivalent procedures. Previous studies have used different criteria in their evaluations of TTS-produced voice quality, rather than taking a comprehensive, holistic approach: while some focused exclusively on users' perceptions of the synthetic voice's naturalness (e.g., Stevens et al., 2005), others included only comprehension measures (e.g., Bailly, 2003). Additionally, most research used native speakers as TTS evaluators, which may have influenced their conclusions and made them inapplicable to L2 speakers.

Moreover, it is worth mentioning that the above-mentioned studies are dated. Text-to-speech synthesis has advanced significantly over the last two decades, especially since the introduction of voice-based personal assistants such as those found in GPS systems, cellphones (Siri, and Cortana), and smart speakers (Amazon Echo, Google Home). Finally, previous research, limited attention has been given to investigating the extent to which Text-to-Speech (TTS) technology can be optimized to focus on specific language forms. This aspect is essential in determining the effectiveness and suitability of TTS for second language (L2) instruction, which is an important factor in determining the effectiveness of any technology for L2 instruction.

Recently, Cardoso, Smith, and Garcia Fuentes (2015) conducted a study to evaluate the performance of Text-to-Speech (TTS) technology in four areas: comprehensibility, naturalness, pronunciation accuracy, and intelligibility, compared to human speech. They enlisted 15 undergraduate students to rate oral samples delivered by TTS and human recordings in two conditions: sentences and a story. Participants also undertook an identification task focusing on

the past form of regular verbs in English, determining the presence of the target grammar feature /t/, /d/, or /id/. The results showed a significant difference in the overall scores of human and TTS oral productions in both scenarios. However, there were no significant differences in the identification task. Despite not having the same baseline as human recordings, participants provided comparable high scores to the TTS samples in both scenarios for comprehensibility, pronunciation accuracy, and intelligibility. These findings suggest that technology-generated synthesized voices can currently serve as an adequate source of spoken input for L2 learners.

Similarly, Grimshaw, Bione Alves, and Cardoso (2018) examined the output of five different TTS applications and found similar outcomes. The results revealed that while participants' overall evaluations for comprehensibility were relatively high, ratings for naturalness were lower. Furthermore, this research suggests that as language users become more familiar with a specific synthesized voice, they can perceive speech as more understandable and natural.

In line with these findings, Liakin et al. (2017) found that the quality of synthesized voices can match that of human voices, making it suitable as a pronunciation model. These findings indicate that TTS technology is ready for integration in L2 classrooms, offering potential benefits for pronunciation practice, especially as a supplemental source of input that considers learners' individual needs and interests (Cardoso, 2018).

In conclusion, the effectiveness of Text-to-Speech (TTS) synthesis for second language (L2) instruction has been a topic of discussion in the literature. While previous studies have shown mixed results, comparing TTS to natural human speech, recent research indicates that TTS technology has improved significantly. Studies have evaluated dimensions such as comprehensibility, naturalness, pronunciation accuracy, and intelligibility. While participants generally perceive human voices as more natural, recent findings suggest that TTS technology can

provide adequate comprehensibility, pronunciation accuracy, and intelligibility for L2 learners. TTS can serve as a valuable tool in L2 classrooms, offering increased language input and supporting pronunciation practice. However, further research is needed to explore its effectiveness in addressing specific language forms and its overall impact on L2 instruction.

2.2. Text-To-Speech and Its Application in Language Learning

Text-to-speech (TTS) technology has been found to have significant benefits for improving second language (L2) pronunciation, according to various studies. These benefits include reinforcing the relationship between graphemes and phonemes in the target language enhancing writing and reading abilities, broadening vocabulary knowledge (Handley,2013), and improving the pronunciation of trainee teachers (Ekşi &Yesilçinar,2016). Studies have also shown that TTS can promote learner autonomy, improve competency, and make students' pronunciation intelligible and comprehensible (Khalid & Muhammad,2014). The present section provides a review of studies examining the effectiveness and benefits of Text-to-Speech (TTS) for L2 pronunciation improvement.

A quasi-experimental Study conducted by Soler Urzúa (2011) investigates the extent to which pedagogical instruction using text-to-speech (TTS) technology as a means to enhance the aural input assists learners in the acquisition of the English /ɪ/. Three groups of learners with the same L1 (Spanish) and similar English proficiency were pre-tested using different tasks on their ability to recognize and produce the target vowel (two for each ability). Each group received a distinct type of instruction: TTS-based instruction, non-TTS-based instruction, and ordinary classroom instruction. The TTS group completed tasks designed to improve their perception of the target forms using TTS; the non-TTS group completed thentical tasks but with input from the

researcher; and the third group completed listening comprehension tasks. After receiving treatment, the three groups were assessed on their producing and perceptual abilities regarding the target sound. The results revealed that the TTS group outperformed the non-TTS group in one of the pronunciation tests.

Handley (2013) suggests that Text-to-Speech (TTS) technology offers notable advantages in specific language areas. One such area is the reinforcement of the connection between graphemes (written symbols) and phonemes (speech sounds) in the target language. By using TTS, learners can hear the correct pronunciation of words and associate them with their written forms. This integration of visual and auditory information can enhance learners' writing and reading skills, as well as expand their vocabulary knowledge. TTS technology enables learners to improve their understanding of the phonetic aspects of the language, ultimately supporting their overall language development.

Similarly, a study by Khalid and Muhammad (2014) investigated the effectiveness of Text to Speech (TTS) software in improving the pronunciation of Graduation level students. They conclude that the use of Text to Speech Software in pedagogy promotes learner autonomy, improves competency, and makes students' pronunciation intelligible and comprehensible.

In addition, text-to-speech proved to be effective to improve the pronunciation of trainee teachers. The results of the study conducted by Ekşi and Yesilçinar (2016) demonstrate that the use of online text-to-speech tools had a positive impact on the pronunciation skills of EFL teacher trainees. The researchers compared the oral achievement test scores of the trainees before and after the introduction of text-to-speech tools, with a focus on their total grade and scores in two subsections of the weighted rubric: pronunciation and accent, and fluency. The results revealed a significant difference in all three dimensions. Trainees' fluency, pronunciation, and total scores on

the post-achievement test were higher after using the text-to-speech tools, as compared to their scores before using the tools. This suggests that incorporating online text-to-speech websites into self-study activities can effectively improve pronunciation skills in EFL teacher trainees. Furthermore, the findings from the reflection questionnaire administered to the trainees indicated that they perceived the text-to-speech websites as user-friendly and effective. Many trainees expressed their intention to continue using the tools in the future, which is noteworthy as it highlights the potential for long-term autonomous learning habits to be more effective than one-time attempts.

Thus, the study conducted by Ekşi and Yesilçinar (2016) suggests that online text-to-speech tools can be beneficial for EFL teacher trainees in improving their pronunciation skills, as evidenced by the improved test scores and positive feedback from the trainees regarding the user-friendliness and effectiveness of the tools.

In a mixed study conducted by Liakin, Cardoso, Liakina (2017) found that some learners reported an improvement in their pronunciation after using TTS. In addition, participants' responses from interviews acknowledged the potential benefits of using text-to-speech (TTS) technology, specifically in the areas of extensive listening and oral comprehension practices. The authors of the study explain this gain by stating that "TTS app increased the learners' exposure and access to a correct pronunciation model" (Liakin, Cardoso, & Liakina, 2017, p. 24).

In other words, the participants noticed that their pronunciation skills improved after using TTS, which the authors attribute to the increased exposure and access to accurate pronunciation models provided by the app. Additionally, some participants in the study recognized that TTS technology can be advantageous for activities such as listening extensively and improving their oral comprehension skills.

Based on Liakin et al. (2017) finding it is suggested that TTS technology has the potential to support language learners in various ways, including enhancing their pronunciation accuracy, listening skills, and comprehension abilities. The results highlight the value of incorporating TTS tools in language learning activities to facilitate learners' exposure to correct pronunciation models and potentially improve their overall language proficiency.

In terms of using TTS to improve phonological awareness De Araújo Gomes, Cardoso, and De Lucena (2018) conducted a study to explore the pedagogical potential of Text-to-Speech (TTS) as a tool for assisting English second language (L2) learners in developing their Phonological Awareness, with a focus on the morphophonological alternations involved in regular past tense marking in English (past -ed). The results revealed that TTS had a positive impact on the auditory perception and controlled production of the targeted phenomenon.

The study conducted by Ishikawa et al. (2021) in a university EFL flipped learning course in Japan aimed to investigate whether using TTS technology for speaking practice has an impact on students' anxiety and self-efficacy in making presentations in English. The study used pre- and post-investigation questionnaires. The results indicated that speaking practice using TTS technology decreased anxiety and increased self-efficacy in making presentations in English. These findings suggest that incorporating TTS technology in speaking practice activities may be beneficial in improving students' confidence and performance in delivering presentations in English.

In Van Duong's (2022) study which involved 60 high school students who utilized Text-to-Speech (TTS) tools to practice their English pronunciation. The results indicated a significant improvement in the pronunciation skills of English as EFL students, with participants reporting positive effects on their English pronunciation accuracy. The majority of participants expressed a

positive attitude toward the effectiveness of TTS tools, considering them as valuable aids for improving their pronunciation. These findings suggest that the use of TTS tools in the EFL context has potential implications for improving pronunciation skills.

In a study conducted by Al-Jarf (2022), the effects of using Text-to-Speech (TTS) software on decoding skills and pronunciation accuracy were investigated among freshman students enrolled in Vocabulary I and Reading I courses. The pre-test results indicated that there were no significant differences between the experimental and control groups, suggesting similar proficiency levels at the beginning. However, the experimental group utilized TTS software to practice listening to lessons from the textbook.

The findings of the study revealed significant improvements in decoding skills, reading fluency, and pronunciation accuracy after 8 and 12 weeks of using the TTS software. Furthermore, a positive correlation was observed between the number of practiced lessons and texts, weekly practice time, and posttest scores for decoding and pronunciation proficiency. The students reported positive attitudes towards the use of TTS for practicing decoding and pronunciation. However, no improvement was observed in vocabulary knowledge. Overall, the study suggests that incorporating TTS software in EFL contexts can be beneficial for enhancing pronunciation accuracy and decoding skills among freshman students. It highlights the potential of TTS as a valuable tool in improving specific language skills in EFL settings.

Overall, TTS synthesizers have proven to be useful in L2 pedagogy, particularly when it comes to pronunciation (Cardoso, 2018; Liakin et al., 2017; Soler-Urza, 2011). They can help students obtain adequate amounts of comprehensible input while also developing autonomy to learn on their own (Bione & Cardoso, 2020; Cardoso, 2018; Liakin et al., 2017; Ekşi & Yeşilçnar, 2016). For example, learners can listen to a target L2 word or phrase whenever and wherever they

want, as many times as they want, without requiring the support of a teacher (Liakin et al., 2017). As a result, TTS can assist teachers and students in expanding the reach of the classroom (Bione & Cardoso, 2020; Liakin et al., 2017b), reducing the time restrictions stated previously. Finally, TTS provides opportunities for Learners to focus on specific language forms. Cardoso (2018) focus on past -ed allomorphy in English, Liakin et al. (2017) focus on French liaison), and Soler-Urza (2011) focus on the acquisition of English /y/are examples of studies that have investigated the pedagogical use of TTS for learning L2 pronunciation.

To conclude, in addition to the enhanced input provided by TTS, Students need opportunities for output practice. Some researchers suggest combining TTS with another technology or technique to achieve better result. For example, Le and Tran (2022) employed a shadowing technique with Google Text-to-spoken Speech's text features to remove the learner's flat tone and achieve basic English intonation in the spoken text. Whereas, Amin (2022) recommended using Repeated- Reading and Listening –While- Reading via text-to-speech apps to help higher education students with their reading skills.

In essence, Combining TTS with additional technologies that encourage speech production on the side of the users would be more appropriate. Cardoso (2018) recommends employing TTS in addition to an automatic speech recognition tool to increase outcomes at the production level. Similarly, Liakin, Cardoso, and Liakina (2017, p.5) propose combining both ASR and TTS as an "anytime anywhere mobile learning setting," which appears to be a promising idea to assist pronunciation improvement at both level perception and production.

2.3. Automatic Speech Recognition and L2 Pronunciation

Automatic Speech Recognition (also known as "speech to text" or "computer speech recognition. An automatic speech recognition system involves voice recognition software

that processes human speech and turns it into text (Chazen, n.d.). ASR technology is a type of computer-based feedback tool that analyzes spoken language and provides learners with feedback on their pronunciation accuracy. Researchers propose employing ASR to teach pronunciation of a foreign language to assess students' oral production in the context of pronunciation. Whereas, some publications have questioned ASR technology in a CALL context during the last decade for its inability to interpret L2 speech effectively at the same rate as human listeners (Derwing , Munro, Carbonaro, 2000; Kim, 2006), as well as its insufficient or wrong feedback. Recent studies, on the other hand, have demonstrated that ASR technology has been improving in recent years (Ashwell, Elam,2017; Dizon; Tang, 2020; Bogach et al., 2021) The present section summarizes the evidence supporting the use of automatic speech recognition (ASR) technology in foreign language (FL) learning and teaching. Specifically, this section examines empirical studies that have utilized ASR technology in the context of teaching and learning, specifically with an emphasis on pronunciation.

Ashwell and Elam (2017) investigated how well the Google Web Speech API recognized the speech of Japanese learners of English as a foreign language (EFL). In an elicited imitation test, participants produced 13 sentences with specific grammatical features. They discovered that the system had an overall recognition accuracy of 89.4%. The researchers concluded that, compared to native speaker input, the most difficult issue for systems to perform speech recognition is the pronunciation of some specific sounds. Furthermore, they state that pronunciation issues may not be a barrier for ASR systems and ASR technology could be used to assess students' grammatical ability). Furthermore, Ashwell and Elam (2017, p. 61) argue that “these systems are continually improving on their respective accuracy rates by constantly gathering acoustic information and utilizing machine learning”.

More specifically, Inceoglu, Lim, and Chen (2020) investigated the utility of ASR

pronunciation practice in terms of its impact on learner production at the segmental level as well as the learners' view of ASR as a learning tool. In a pretest and posttest study design, 19 Korean university students produced 28 minimal pair phrases with vowel contrasts. Acoustic research revealed a significant improvement in certain vowels but no change in others. However, the vast majority of participants highlighted that ASR is useful for practicing pronunciation.

Moreover, Mroz (2018) explored how 16 learners of French as a foreign language using ASR in Gmail achieved greater awareness of their intelligibility. The researcher used a qualitative approach to analyze participants' responses during semi-structured interviews. According to the findings, the majority of participants thought ASR was a useful diagnostic tool if they could assess the gaps and successes in their intelligibility using such technology.

In essence, Mroz's research demonstrates the potential advantages of employing ASR as a diagnostic tool for FLL. The study's qualitative methodology enabled a deeper understanding of participants' perspectives, indicating that ASR was viewed as a helpful tool for enhancing participants' knowledge of their own intelligibility and self-assessment of their speaking abilities.

ASR programs, in particular, were compared with conventional materials and techniques for foreign language (FL) learning and instruction in a study done by Golonka et al. (2014). Their results provided solid evidence for the beneficial effects of ASR programs on FL teaching and learning. The researchers also stated that ASR technology has the potential to facilitate pronunciation improvement to a greater extent than teachers, and that ASR programs hold great potential in FL learning (Golonka et al., 2014, p.88). The study also revealed that ASR can be used in a variety of contexts to improve learning both inside and outside of the classroom. It also implies that ASR technology can be a useful tool for FL learners, offering a variety of advantages and opportunities to enhance pronunciation in the language learning process. ASR provides learners

with a tool for independent practice, allowing them to engage in self-guided exercises and receive immediate feedback. This promotes self-directed learning and enables learners to practice their pronunciation skills anytime and anywhere. ASR systems can also track progress over time, allowing learners to monitor their improvement and set goals for further development (Golonka et al., 2014) Moreover, ASR is “especially significant for learners who have little to no access to other L2 speakers outside of class” (Dizon, Tang, 2020, p.108).

Overall, the potential for ASR to improve pronunciation in second language acquisition is substantial. In essence, teachers and students can use this technology's strengths to achieve specific learning objectives. ASR can give students the opportunity to practice their pronunciation and speaking skills by allowing them to produce oral output.

2.4 Integrating ASR and TTS into L2 Pronunciation Learning

The preceding sections examined studies implementing ASR and TTS technology and their affordances or potential in enhancing EFL pronunciation. Ultimately, it is reasonable to suppose that these technologies can promote pronunciation improvement both inside (under the guidance of the teacher) and outside (autonomously by the student). According to Liakin et al. (2017), researchers have only begun to investigate the pedagogical uses of TTS and ASR together in EFL pronunciation; nonetheless, available studies imply positive results as a classroom instruction complement after extensive use.

According to Golonka et al. (2014), technological advances can provide learners with interaction opportunities, feedback, and increased contact with the target language. In addition, to enhance their motivation and interest, these advantages align with students' perceptions of ASR and TTS programs, as they can enhance motivation and interest in language learning. Liakin et

al.(2017) reported comparative results when researching learners' perceptions of the use of both technologies. The researchers claim that the participants not only recognized the pedagogical value of ASR and TTS, but also appreciated the mobile-enhanced learning environment they provided.

Similarly, Levis and Suvorov (2013) state that the connections between ASR and text-to-speech software have not been well investigated, yet they can yield promising results for non-native speech applications.

Darcy, Rocca, and Hancock (2021) argue that even brief periods of training with ASR and TTS can yield significant benefits for learners' pronunciation skills. This means that learners do not need to spend an entire course on pronunciation education, as even a small amount of attention devoted to segmental and suprasegmental features of pronunciation within communication classes can lead to improvements in learners' pronunciation.

Moreover, incorporating ASR and TTS into language learning can make pronunciation training more engaging and interactive. Learners can receive immediate feedback on their pronunciation, allowing them to identify and correct errors quickly. Additionally, TTS can provide models for learners to mimic and imitate, allowing them to practice their pronunciation in a more natural and authentic way.

Finally, based on Darcy's (2018,p.31) definition of efficient pronunciation teaching, it is possible to fulfil all of those 'ingredients' by combining both technologies: ASR can be an integrative part of different explicit and communicative activities, providing learners with endless opportunities to produce oral output, TTS can be used to develop learners' perception, and ASR can provide automatic explicit feedback at the learner's pace. As a result, combining both technologies could be an attractive alternative for pronunciation improvement.

As the aforementioned studies indicate, TTS and ASR technologies can provide significant

benefits to learners, primarily as a supplement to pronunciation instruction. Nonetheless, these advantages can be achieved only if the teachers are aware of "what their students require, and if they use tools that have been proven to be effective" (Darcy, 2018, p. 326) As a result, the use of technology resources is not meant to take the place of the vital role of the instructor. Rather, it is a method of allowing students to focus on individual problems and receive personalized feedback while becoming more autonomous in their learning process, implying that "learning is not limited to the classroom context." (Carlet; Kivistö-de Souza, 2018, p. 104).

The study by Khademi and Cardoso (2022) examined the use of Google Translate (GT) and its speech capabilities, Text-to-Speech Synthesis (TTS), and Automatic Speech Recognition (ASR) in assisting second language (L2) learners in acquiring the pronunciation of English past -ed allomorphy (/t/, /d/, /id/). The focus was on three aspects of pronunciation development: phonological awareness, perception, and production. The results of the pre- and post-tests showed significant improvements in participants' awareness and perception of the English past -ed, but limited improvements in production, except for /id/. However, it is important to note that the study lacked a control group which could have provided a basis for comparison and helped to determine the specific effects of using GT, TTS, and ASR on learning pronunciation. In addition, the duration of the study was relatively short, consisting of only two hours. These limitations suggest that further research with a control group and a longer study duration would be beneficial to gain a more comprehensive understanding of the impact of both technologies TTS and ASR on teaching/learning pronunciation.

Therefore, the present study aims to fill the gap by assessing the effectiveness of TTS and ASR to improve EFL learners' pronunciation on segmental level, more specifically on improving their pronunciation of regular past tense "ed" ending in terms of their awareness, perception and

production by employing control and experiment group. In addition, the present study will examine EFL learners' attitude towards TTS and ASR as a learning tool to improve pronunciation.

Conclusion

To sum up, existing studies suggest positive results for both classroom instruction and autonomous learning. However, further investigation is needed to determine the effectiveness of combining TTS and ASR to improve EFL learners' pronunciation, and examine students' individual experiences with both technologies as pronunciation learning tools. These issues were empirically investigated in the following chapter to shed more light on the potential benefits and drawbacks of using TTS and ASR for improving EFL learners' pronunciation skills in autonomous learning. The first chapter provided an overview of the theories pertaining to teaching second language (L2) pronunciation. Subsequently, a review of the relevant literature on the utilization of text-to-speech (TTS) and automatic speech recognition (ASR) technologies was presented. In the forthcoming chapter, the research design, participant details, study procedures, and data collection and analysis methods were expounded upon.

Chapter II: Methodology

Introduction

The “Methodology” chapter details the field work and analysis of data collected for the present study, which aimed to evaluate the efficacy of Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) technology in enhancing the pronunciation of regular past “ed” among English as a Foreign Language (EFL) learners in Algeria. The study employed a pretest-posttest quasi-experimental design with a control group to assess the effectiveness of the intervention in comparison to a group of learners who did not receive any intervention. Additionally, a questionnaire was administered to investigate the learners' attitudes and perceptions towards the use of TTS and ASR technology as pronunciation learning tools. The chapter is divided into three main sections. The first section expounds on the rationale behind the chosen methodology, including the research approach, research design, data collection methods, population, and sample. The second section presents an analysis of the data collected. Lastly, the third section provides a discussion of the primary findings to address the research questions and hypotheses posed in the general introduction of this dissertation.

1. Research design

Research approach refers to the overall strategy employed to conduct a study and gather data. Creswell (2018) categorized three types of research approaches: qualitative, quantitative, and mixed methods. In this specific investigation, a quantitative approach was adopted. This selection is justified by the researcher's use of an experimental design to examine the impact of Text-To-Speech and Automatic Speech Recognition on EFL learners' pronunciation of regular past tense "-ed”. The pretest posttest control and experimental group was chosen because it allows for the comparison of the effectiveness of the intervention with a control group that did not receive the intervention.

In educational research settings, it is not always feasible to randomly select or assign participants. This was the case in the present study, as the participants were already divided into two classes. Therefore, a quasi-experimental design was employed. More specifically, the current study utilized a non-randomized control group pretest-posttest design.

This research study comprises two distinct groups: an experimental group and a control group. Both groups undergo pre-test and post-test assessments. The experimental group receives a treatment utilizing Text-To-Speech (TTS) and Automatic Speech Recognition (ASR) features incorporated within the Neural Read application, administered outside as autonomous learning tool. On the other hand, the control group does not receive any specific treatment. One commonly utilized quasi-experimental design in educational research that can be employed in this study is as follows:

Table 1: *Nonrandomized Control Group, Pretest–Posttest Design*

Group	Pre-test	Treatment	Post-test
Experiment	Y1	X	Y2
Control	Y1	-	Y2

Explanation:

- Y1: Pre-test score indicates the assessment conducted by the researcher before administering the treatment. It measures the students' pronunciation skills across three levels: awareness, aural discrimination, and production.
- X: Treatment refers to the intervention administered by the researcher, involving the use of TTS (Text-To-Speech) and ASR (Automatic Speech Recognition) features in the application.
- Y2: Post-test score signifies the assessment administered by the researcher after the

treatment. It aims to evaluate the impact of the treatment and measure the students' scores on pronunciation after utilizing TTS and ASR as part of the intervention.

2. Population and Sample

The present study was conducted at the Center for Intensive Training at Mohamed Kheider University during the academic year 2022-2023. The population and sample will be elaborated upon as follows:

2.1 Population

In this research, the researcher chooses learners who were registered for an English as a foreign language (EFL) A1 student course at the Language Centre of Intensive Training at the University of Biskra. The center administers a level test to distribute its prospective students to different courses. Population was chosen to achieve multiple objectives. The first objective was to ensure that the experiment represented the diversity of students attending the same course at the Language Centre of the University of Biskra. The second objective was to guarantee that all participants had the same level of English proficiency at the start of the study. The third objective was to work with a group of strictly selected A1 learners, who had very little or no prior training in English phonetics. The confirmation of this was obtained through the pre-test results. Furthermore, the selection of participants ensured the representation of the diversity of learners at the Language Centre of the University of Biskra, and that all participants had a similar level of English proficiency and limited prior training in English phonetics.

2.2 Sample

According to Ary (2010), a sample is a subset of a population that is selected for observation. It represents a smaller portion of the entire population. In this study, the researcher employed cluster sampling, which involves selecting groups of individuals rather than individual

participants (Ary, 2010). For this research, only two classes were chosen as the samples. Class A was designated as the experimental group, while class B served as the control group.

The study comprised two groups, each consisting of 14 students, who were enrolled in A1 student courses at the Language Centre of Intensive Training at the University of Biskra. The participants' ages ranged from 18 to 61 years, reflecting a diverse age distribution within the sample.

Table 2: *Distribution of Participants by Group and Gender*

Group	**Experiment**	**Control**
Female	10	8
Male	4	6
Total	14	14

As depicted in the table above the experiment group consisted of 14 students, 10 females and 4 males while the control group consisted of 8 females and 6 males.

Prior to their participation in the study, the researcher provided detailed information about the research study, including its purpose, procedures, potential risks and benefits, and confidentiality measures. Informed consent forms were distributed to the participants, clearly outlining their rights as participants and emphasizing their voluntary participation in the study. Additionally, all 28 students, including those in both the experimental and in-classroom groups, were explicitly informed and requested to refrain from engaging in any additional English-related activities during the one-week duration of the experiment. This included avoiding extra English lessons or conversation exchanges with native speakers or using other CAPT tools. Importantly, all participants agreed to comply with this condition, signifying their commitment to the study protocol (See appendix 2).

4.Data Collection Methods

The data collection methods for the present study comprised a pretest, a posttest, and a semi-structured questionnaire. The pretest and posttest allowed for a comparison of the effectiveness of the intervention (TTS and ASR technology) in improving pronunciation of regular past “ed”. The questionnaire was used to collect information on learners' attitudes and perceptions regarding using TTS and ASR technology to improve their pronunciation skills.

The pretest served as a baseline measure of learners' pronunciation skills while the posttest measured any improvements in pronunciation skills following the intervention. The control group did not receive any intervention and served as a comparison group to assess the effectiveness of the intervention in the experimental group. Both the pretest and posttest were administered through a JotForm platform.

The questionnaire consisted of a series of close-ended questions, such as a Likert scale and multiple-choice questions, which allowed for the collection of quantitative data on learners' attitudes and perceptions. Additionally, a few open-ended questions were included to to provide deeper insights into learners' experiences and perceptions of the use of TTS and ASR technology for improving their pronunciation skills.

5.Description of the Quasi-Experiment

5.1 Pretest

The pretest (out of 54) was administered to evaluate participants' level of awareness, aural discrimination, and production of the final "ed" sound (See appendix 3). The specific details of the pretest are as follows:

For awareness test (out of 18), the first test involved students responding to a set of open-ended questions to determine their knowledge of how past "ed" is pronounced (out of 3). In the

second test, participants were asked to match regular past-tense verbs to their corresponding inflected endings, such as associating "used" with /d/ and "added" with /id/, based on their perceived pronunciations (out of 15. five verbs for each allomorphs).

For aural discrimination test (out of 18, six verbs for each allomorphs), participants listened to an audio file played via the jotform platform. They were required to categorize the sounds they heard, identifying the three past "ed" allomorphs (PAST-ASKED for /t/, PAST-USED for /d/, and PAST-ADDED for /id/). There were six verbs for each variation. This test aimed to assess participants' ability to recognize the pronunciation of final "ed" within a speech connected (paragraph).

To assess production (out of 18), three tests were employed. Firstly, participants were asked to read aloud a list of six verbs in their past "ed" forms, including an equal number of items for each "ed" allomorph. Their reading was audio-recorded. Subsequently, participants were asked to read aloud a short paragraph containing regural verbs. Finally, a controlled task was conducted where participants read a story and answered questions about a fictional character. This test aimed to evaluate participants' ability to pronounce the final "ed" sound in a more spontaneous activity. The tests design was adapted from Khademi (2021).

5.2 Treatment

During the practice phase, learners utilized the "Neural Reader" application individually to improve their listening and pronunciation skills of the past "-ed" forms. They were instructed to practice pronouncing past "-ed" in a self-directed environment, treating it as a homework assignment, for approximately one week. To accomplish this, they utilized the app's Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) features for practicing listening (including aural discrimination) and speaking skills, respectively **(See appendix 5).**

To enhance their "aural discrimination" of past "-ed," participants engaged in two sets of activities facilitated by the app's TTS. In the first activity, they worked with two stories and dialogues containing a balanced distribution of the three "-ed" allomorphs, with four instances of each. They were required to copy and paste each text into the app and listen to it.

The second types of listening activity consisted of a word list containing regular simple past tense verbs. Participants were instructed to copy and paste the verbs (30 in total, 10 for each allomorph) into the app and listen to their synthesized pronunciations. Subsequently, they had to match the verbs based on their perception of the inflected endings, associating them with existing forms such as "used" for /d/, "added" for /id/, and "asked" for /t/.

For oral production practice and orthographic feedback, two activities were assigned to participants using the app's ASR feature. Firstly, they were presented with pairs of verbs in both present and past tense (10 items for each alternation) and asked to read both forms aloud, ensuring that the intended form appeared on the screen. In case the app displayed a different form than what they intended to say, they were required to repeat until the correct verb form was shown.

The second activity involved a list of short sentences in both present and past tense (10 items for each variation). Similar to the first production activity, participants had to read both forms aloud using the ASR feature to verify if the intended form appeared on the screen. Again, if the app generated a different text output than what they intended to say, they needed to repeat their utterance until the ASR accurately recognized their speech.

Throughout the designated practice time, participants had the freedom to employ their preferred learning strategies using the app. This included the ability to listen to any part of the story or specific words as many times as needed, take notes on pronunciation observations related

to past "-ed," and select their preferred voice and speed settings to ensure comfort during the practice sessions.

5.2.1 Neural Reader Application

The Neural Reader application, developed by Chenghang Zheng in 2022 (version 2.3.15), served as a valuable pronunciation learning tool during the treatment phase of this study. This application combines text-to-speech and automatic speech recognition functionalities, offering a range of voice options and the ability to regulate reading speed. With its user-friendly interface, users can effortlessly convert text into audio and record their own voices for playback and analysis.

The selection of the Neural Reader app for this study was based on its notable attributes, including text-to-speech and automatic speech recognition capabilities. The availability of multiple voice options and speed control feature empowered students to personalize their learning experiences according to their individual needs and preferences. Another advantageous feature of the app was its ability to store data in both text and audio formats, facilitating sharing of progress with the researcher. This functionality was particularly crucial during the treatment phase, enabling the collection of valuable data on students' pronunciation skills.

Moreover, the study found that the free version of the app met the requirements of this investigation adequately. Although a premium version of the Neural Reader app offers additional features, the free version proved sufficient for the study's objectives.

In summary, the Neural Reader app emerged as a suitable and efficient tool for the treatment phase of this study. Its integration of text-to-speech and automatic speech recognition capabilities, along with multiple voice options and data storage functionalities, made it an optimal choice for enhancing EFL learners' perception and production of plosive consonants.

5.3 Posttest

A posttest was administered in class at the end of the treatment. Identical test formats were used in the posttest as in the pretest **(See appendix 4)**, but with different content. By using identical test formats in the pretest and posttest, any changes in the experimental group's test scores could be attributed to the intervention. To avoid any learning or practice effects, different content was used in the posttest. This ensured that any differences in test scores between the pretest and posttest could be confidently attributed to the intervention itself, rather than to any familiarity with the test content. In conclusion, utilizing identical test formats with different content in the pretest and posttest was a valid and effective method for evaluating the effectiveness of the intervention in the study.

5.4 Questionnaire

The questionnaire was the second data collection tool of the under-investigated study. It was a semi-structured questionnaire administrated to Fourteen (14) participant of the experiment group. It was submitted in written form to collect data about students' experiences, attitudes, perceptions, and opinions about the use of TTS and ASR as pronunciation learning tools to improve pronunciation. It comprised 22 questions that include Likert scale questions, yes and no questions multiple-choice questions, as well as open-ended questions **(See appendix 6).** Accordingly, the questionnaire to included five sections structured as follow:

5.4.1 **General information** (two items)

First section collected data about the students' information. This section encompassed two items that is concerned with the gender, and age of the participants.

5.4.2 Part One: Perceptions of Text-To-Speech (TTS)

In the first part of the questionnaire, participants were asked to provide their perceptions of Text-To-Speech (TTS) technology. This section included four items, focusing on the clarity and naturalness of the TTS input (Q1), ease of use of the TTS tool (Q2), the effectiveness of TTS in identifying and distinguishing different variations of the final "-ed" sound (Q3), and whether participants would recommend the TTS tool to other English language learners for improving pronunciation skills (Q4). Participants were provided with response options ranging from "Strongly agree" to "Strongly disagree."

5.4.3 Part Two: Perceptions of Automatic Speech Recognition (ASR)

The second part of the questionnaire aimed to gather participants' perceptions of Automatic Speech Recognition (ASR) technology. This section included five items. Participants were asked to evaluate the accuracy of the output provided by ASR technology (Q6), assess the helpfulness of ASR as a tool for pronunciation practice (Q7), compare its effectiveness to traditional methods (Q8), indicate the ease of using ASR for pronunciation practice (Q9), and express their willingness to recommend ASR as a pronunciation learning tool to other learners (Q10).

5.4.4 Part Three: Perceptions of Text-To-Speech (TTS) and Automatic Speech Recognition (ASR)

In this part, participants were instructed to indicate their level of agreement or disagreement with a set of statements regarding the combined use of TTS and ASR. The statements covered topics such as self-evaluation of pronunciation (Q1), comfort in practicing pronunciation (2 and 3), engagement beyond the classroom (4), and becoming a more autonomous learner (5). Additionally,

participants were asked about the helpfulness of using both TTS and ASR together to improve perception and production of the final "-ed" sound in English verbs (Q16), their confidence in perceiving and producing correct pronunciation after using TTS and ASR tools (Q17), and whether they would recommend pronunciation learning with TTS and ASR to other learners (Q18). Participants were also asked to rank the benefits of using TTS and ASR tools for pronunciation improvement (Q19) and specify which features of the app they utilized (Q20).

5.4.6 Part Four: Suggestions and Comments

In this section, participants were encouraged to provide suggestions, comments, or recommendations based on their experience with TTS and ASR technology (Q21). They were also asked about the potential usefulness of TTS and ASR in improving pronunciation. Finally, participants were given the opportunity to provide any additional suggestions, comments, or recommendations they may have (Q22).

6.Validity and Piloting the Questionnaire

To assess the validity and reliability of the questionnaire, it was initially evaluated by five teachers, who provided feedback to refine and update the questionnaire. Subsequently, a pilot test was conducted with three A1-level students to further assess its reliability. The teachers' input enhanced the validity of the questionnaire, while the pilot testing allowed for identification and resolution of potential issues. These steps collectively contributed to improving the questionnaire's quality and ensuring it effectively measures the intended construct.

7. Data Collection Procedures

Once the researcher granted access by the center, the researcher involved two teachers and students from both groups. The study was conducted from April 25th to May 3rd, 2023. The researcher implemented the following procedures to collect the necessary data:

1. Pretest with Both Groups: The researcher conducted a pretest with both the experimental and control groups to establish a baseline for comparison.
2. Introduction to TTS and ASR: The experiment group received two short presentations about Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) technologies. Additionally, the researcher shared a recorded video explaining how to sign up and effectively utilize the application along with its features.
3. Initial Application Usage in the Classroom: In the first session of the treatment phase, the researcher introduced the application to the students in the classroom, guiding them on its proper usage.
4. Semi-Autonomous Treatment Phase: During the treatment phase, the participants autonomously used the application. The researcher created a support group online, where they were available to answer any questions and resolve technical issues that arose.
5. Follow-Up Session: Approximately one week later, the researcher conducted another session to ensure that all students had familiarized themselves with the application. Students were given about 15 minutes in the classroom to utilize the app and confirm their proficiency in using TTS and ASR functionalities.
6. Posttest and Questionnaire Administration: In a subsequent session, the researcher administered the posttest in the classroom using the jotform platform, following the same procedure as the pretest. After completing the posttest, participants were asked to submit a questionnaire to gather additional information.

By implementing these procedures, the researcher aimed to collect relevant data for the study.

8. Data Analysis

The data collected in this quasi-experimental research study, was analyzed using descriptive and inferential statistical techniques. For the descriptive analysis, several descriptive measures were computed using SPSS 23. Frequency counts were conducted to determine the number of participants in each experimental condition (TTS and ASR group, control group). Measures of central tendency, including the mean, median, and mode, were calculated for the pre-test and post-test scores separately in each group to provide an overview of participants' initial and final performance. Measures of variability, such as standard deviation and range, were computed to assess the consistency or variability of participants' pronunciation improvement within each group.

In terms of inferential analysis, an independent t-test was performed to compare the pre-test and post-test scores between the control and experimental groups. This analysis aimed to determine whether there were any significant differences in pronunciation improvement between the two groups after the intervention. Additionally, the quantitative data from the questionnaire administered to the participants was analyzed using descriptive statistics to summarize their responses and provide insights into their perceptions of the effectiveness of TTS and ASR technologies in improving their pronunciation.

The statistical software SPSS 23 was utilized for all data analyses, providing the necessary tools to perform the descriptive and inferential statistical tests. The significance level (α) chosen for the inferential statistics was set at 0.05, indicating that any p-value below this threshold would be considered statistically significant. The results of the data analysis are presented and interpreted in the subsequent sections to determine the effectiveness of TTS and ASR technologies in enhancing EFL learners' pronunciation of regular past "-ed" forms.

8.1 Hypothesis Testing

The purpose of employing an independent sample t-test is to identify significant distinctions between an experimental group and a control group. The process involves conducting hypothesis testing and establishing the alpha (α) level at 0.05 for a two-tailed test. The hypotheses can be formulated as follows:

- Ho: There is no significant impact on students' scores after utilizing TTS and ASR between the control group and the experimental group.
- Hi: Learners who utilize TTS and ASR as pronunciation learning tools will demonstrate significantly higher scores compared to the control group who does not use these tools.

To determine the t-value, the independent sample t-test is employed, and the calculated result is compared to the predetermined significance level. If the p-value (sig.) obtained from the output is greater than 0.05, the researcher should accept the null hypothesis (Ho), indicating that there is no significant difference in students' scores between the groups. Conversely, if the p-value is less than 0.05, the researcher can reject the null hypothesis (Ho), implying that the alternative hypothesis (Hi) is accepted, indicating a significant difference in students' scores between the control and experimental groups.

Chapter III: Results and Discussion

Introduction

The study focuses on investigating the effectiveness of Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) technologies in improving the pronunciation skills of English as a Foreign Language (EFL) learners. This chapter presents the analysis of the collected data using appropriate statistical procedures. Through this analysis, the formulated hypothesis will be tested, and the research questions posed in this study will be addressed.

1.Analysis of Pretest and Posttest results and findings

1.1 Results of Pretest

1.1.1 Descriptive Statistics of Pretest

Table 3: *Description Statistics of pretest scores for both groups*

	Groups	*N*	*Mean*	*Std. Deviation*	*Std. Error Mean*
Pretest	*Control*	*14*	*24.8571*	*4.14835*	*1.10869*
	Experiment	*14*	*23.3571*	*4.04983*	*1.08236*

Note. The table presents the descriptive statistics for two groups: the control group and the experimental group. The statistics provided include the group size (N), mean, standard deviation, and standard error of the mean.

For the control group, which consists of 14 participants, the mean pretest score is 24.8571, with a standard deviation of 4.14835. The standard error of the mean, which measures the precision of the estimated mean, is 1.10869. Similarly, for the experimental group, also consisting of 14 participants, the mean pretest score is 23.3571, with a standard deviation of 4.04983. The standard error of the mean is 1.08236. based solely on the mean pretest scores, the control group appears to have a slightly higher average score than the experimental group. However, statistical analysis

such as independent t-test is required to determine if this difference is significant or merely due to random variation.

1.1.2 Statistical Consideration

Before conducting an independent t-test, there are several statistical considerations should be taking into account:

- Independent samples: The two groups being compared should consist of independent observations. This means that the participants in one group should not be related or matched to the participants in the other group which is the case in the present study.
- Normality: The data within each group should follow a normal distribution. This assumption is necessary for conducting a valid t-test. normality can be assessed using statistical tests like the Shapiro-Wilk test.
- Homogeneity of variances: The variances of the two groups being compared should be approximately equal (homogeneity of variances assumption). Violation of this assumption can impact the validity of the t-test results. homogeneity of variances can be assessed using tests like Levene's test.
- **Normality test:**

Table 4 : *Normality Tests for Pretest Scores in Control and Experimental Groups*

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pretest scores for control	.201	14	.132	.920	14	.220
pretest scores for experiment	.132	14	.200*	.972	14	.908

Note. The table provides the results of normality tests conducted on the pretest scores for the control group and the pretest scores for the experimental group. The tests used are the Kolmogorov-Smirnov test and the Shapiro-Wilk test.

The normality tests conducted on the pretest scores for both the control and experimental groups suggest that the data is consistent with a normal distribution. There is no strong evidence to reject the assumption of normality in either group. Therefore, it is reasonable to assume normality when conducting further statistical analyses.

The results of the normality tests suggest that the pretest scores for both the control and experimental groups follow a normal distribution. In the control group, the p-values of the Kolmogorov-Smirnov and Shapiro-Wilk tests are .132 and .220, respectively, both of these value are greater than p=0.05 indicating no significant deviation from normality. Similarly, in the experimental group, the p-values are .200 and .908 for the respective tests. the p-value suggests no substantial departure from normality. These findings support the assumption of normality for further statistical analyses involving the pretest scores.

- **Homogeneity of variances:**

Table 5: *Levene's Test for Equality of Variances*

		Levene's Test for Equality of Variances	
		F	Sig.
Pretest	Equal variances assumed	.069	.795
	Equal variances not assumed		

Note. The table presents the results of Levene's Test for Equality of Variances for the variable "Pretest."

Bases on the results Levene's test indicates that the variances of the two groups are statistically equal. Levene's test for equality of variances indicates that the assumption of equal variances between the groups is not violated, as the p-value is greater than .05.

The results implies that the assumption of equal variances can be reasonably assumed for subsequent statistical tests, such as the t-test for equality of means.

Independent t-test.

After ensuring that the scores of the pretest variable are normally distributed and the assumption of equal variances between the groups is not violated, it is appropriate to proceed with conducting the independent t-test. The independent t-test compares the means of two independent groups and assesses whether the observed difference in means is statistically significant. By conducting the independent t-test, the researcher can determine if there is a significant difference in the means of the two groups on the pretest variable, further supporting the analysis of the data.

Table 6: *Comparison of Pretest Scores between Groups Using Independent Samples T-Test*

		t-test for Equality of Means						
							95% Confidence Interval of the Difference	
		t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Pretest	Equal variances assumed	.968	26	.342	1.50000	1.54942	-1.68488	4.68488
	Equal variances not assumed	.968	25.985	.342	1.50000	1.54942	-1.68497	4.68497

Note. This table presents the results of an independent samples t-test conducted to examine the equality of means between two groups on the pretest variable.

Based on the results presented in the table, the following can be inferred:

- The independent t-test results show that there is no statistically significant difference in means between the two groups on the "Pretest" variable. The t-value is .968, with a corresponding p-value of .342. The p-value is greater than the conventional significance level of .05, suggesting that the difference in means is not statistically significant.

In summary, based on the provided results, there is no significant difference in means between the two groups on the "Pretest" variable. This suggests that any observed differences or changes in the posttest scores between the control and experimental groups can be more confidently attributed to the treatment or intervention itself, rather than pre-existing differences in the groups' baseline levels on the "Pretest" variable.

1.2. Results of Posttest

1.2.1 Descriptive Statistic for Posttest

Table 7: *Descriptive Statistics of Posttest Scores in Experiment and Control Groups*

	Group	N	Mean	Std. Deviation	Std. Error Mean
Posttest	Experiment group	14	42.0000	4.35007	1.16260
	Control group	14	31.6429	4.21731	1.12712

The experiment group, consisting of 14 participants, achieved a higher mean posttest score of 42.0000 compared to the control group's mean score of 31.6429. The standard deviations indicate slightly higher variability in the experiment group (4.35007) compared to the control group (4.21731), while the standard errors of the mean are relatively similar. These descriptive statistics suggest that the experiment group performed better on the posttest, but further statistical analysis is needed to determine the significance of this difference

1.2.2 Inferential Statistics

Table 8: *Independent Samples T-Test for Equality of Means in Posttest Scores.*

		t-test for Equality of Means						
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Posttest	Equal variances assumed	6.396	26	.000	10.35714	1.61928	7.02867	13.68561
	Equal variances not assumed	6.396	25.975	.000	10.35714	1.61928	7.02852	13.68577

The given results from the independent samples t-test indicate that the experimental group outperformed the control group on the "Posttest" variable.

The mean difference between the two groups is 10.35714. This means that, on average, the scores of the experimental group are higher than those of the control group. The standard error of the mean difference is 1.61928, suggesting that the observed mean difference is reliable and not due to random variation.

The 95% confidence interval for the difference in means ranges from 7.02867 to 13.68561. This means that we can be 95% confident that the true difference in means between the experimental and control groups falls within this range. As the confidence interval does not include

zero, it further supports the conclusion that the experimental group performed significantly better than the control group.

The t-value of 6.396 indicates the magnitude of the difference relative to the variability within the groups. The associated p-value of .000 (significantly less than .05) suggests that the observed difference in means is highly unlikely to occur by chance alone. Since p-value is less than 0.05, it indicates that the observed difference is statistically significant, and the the null hypothesis is rejected . In this case, the alternative hypothesis (Hi) is accepted, implying that there is a significant difference in students' scores between the control and experimental groups.

Taken together, these findings provide evidence that the experimental group had significantly higher scores on the "Posttest" variable compared to the control group. This finding suggests that the treatment or intervention implemented in the experimental group had a positive effect on their performance on the overall test, resulting in better outcomes compared to the control group. The upcoming sections will provide a detailed analysis of the effects of the treatment on three levels: awareness, perception, and production.

1.3 Results of Awareness Test

1.3.1 Descriptive Statistics ofAwareness Pretest

Table 9: *Descriptive Statistics for Awareness Pretest and Posttest Scores in Experiment and Control Groups*

	Groups	N	Mean	Std. Deviation	Std. Error Mean
Awareness Pretest	experiment group	14	7.0000	2.14834	.57417
	Control group	14	7.2143	2.11873	.56625
Awareness Posttest	experiment group	14	12.1429	2.21384	.59167
	Control group	14	9.9286	2.58589	.69111

Note. The table provides the group statistics for the "Awareness Pretest" and "Awareness Posttest" variables, separately for the experiment group and the control group.

In Awareness Pretest**,** the mean pretest score for the experiment group is 7.0000, with a standard deviation of 2.14834. The standard error of the mean is .57417. These statistics describe the distribution of scores within the experiment group on the "Awareness Pretest" variable.

Control group: The mean pretest score for the control group is 7.2143, with a standard deviation of 2.11873. The standard error of the mean is .56625. These statistics describe the distribution of scores within the control group on the "Awareness Pretest" variable.

For Awareness Posttest ,the mean posttest score for the experiment group is 12.1429, with a standard deviation of 2.21384. The standard error of the mean is .59167. These statistics describe the distribution of scores within the experiment group on the "Awareness Posttest" variable.

Control group: The mean posttest score for the control group is 9.9286, with a standard deviation of 2.58589. The standard error of the mean is .69111. These statistics describe the distribution of scores within the control group on the "Awareness Posttest" variable.

Interpretation

The mean pretest scores for both the experiment and control groups on the "Awareness Pretest" variable are relatively similar, with the control group having a slightly higher mean of 7.2143 compared to the experiment group's mean of 7.0000. In terms of the posttest scores on the "Awareness Posttest" variable, the experiment group had a higher mean of 12.1429, while the control group had a lower mean of 9.9286.

These findings suggest that, after the treatment or intervention, the experiment group showed a greater improvement in scores compared to the control group. The higher mean posttest

score in the experiment group indicates a positive effect of the treatment on increasing awareness, as compared to the control group where the improvement was relatively lower.

The standard deviations and standard errors provide information about the variability and precision of the mean estimates within each group. However, further inferential statistics is needed to get in depth understanding of data.

1.32 Inferential statistics of Awareness Posttest

Table 10: *Results of Levene's test for equality of variances and t-test for equality of means*

		Levene's Test for Equality of Variances	
		F	Sig.
Awareness Pretest	Equal variances assumed Equal variances not assumed	.065	.800
Awareness Posttest	Equal variances assumed Equal variances not assumed	.742	.397

Note. Levene's test was conducted to assess the equality of variances between groups.

In the case of the Awareness Pretest, the results of Levene's test show that the p-value is 0.800, which is greater than the significance level of 0.05. This suggests that there is no significant difference in the variances of the two groups. Therefore, the assumption of equal variances between the groups is reasonable for the Awareness Pretest.

Similarly, for the Awareness Posttest, the p-value from Levene's test is 0.397, which is also greater than 0.05. This indicates that there is no significant difference in the variances of the two groups. Hence, the assumption of equal variances is reasonable for the Awareness Posttest as well. Therefore, both Levene's test results suggest that the assumption of equal variances between the groups is reasonable for both the Awareness Pretest and the Awareness Posttest. This means that

when conducting subsequent statistical analyses, such as independent t-tests, it is appropriate to assume equal variances between the groups.

Table 11: *Comparison of Awareness Levels Before and After Treatment: Independent Samples Test*

		t-test for Equality of Means						
							95% Confidence Interval of the Difference	
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Awareness Pretest	Equal variances assumed	-.266	26	.793	-.21429	.80642	-1.87191	1.44334
	Equal variances not assumed	-.266	25.995	.793	-.21429	.80642	-1.87192	1.44335
Awareness Posttest	Equal variances assumed	2.434	26	.022	2.21429	.90978	.34420	4.08437
	Equal variances not assumed	2.434	25.397	.022	2.21429	.90978	.34203	4.08654

Awareness Pretest. The t-test for equality of means examines whether there is a significant difference in the mean scores of the two groups on the Awareness Pretest. The p-value of 0.793 is not significant, indicating that there is no significant difference in mean scores between the groups. The mean difference is -0.21429, indicating a slight decrease in scores for one group compared to the other. However, the confidence interval (-1.87191 to 1.44334) includes zero, further supporting the lack of statistical significance.

Interpretation. Based on the results, there is no significant difference in the Awareness Pretest scores between the control and experimental groups. This suggests that any initial differences in awareness levels between the groups were not statistically significant.

In Awareness Posttest, the t-test for equality of means examines whether there is a significant difference in the mean scores of the two groups on the Awareness Posttest. The p-value of 0.022 is significant, suggesting that there is a significant difference in mean scores between the groups. The mean difference is 2.21429, indicating a noticeable increase in scores for the experimental group compared to the control group. The confidence interval (0.34420 to 4.08437) does not include zero, further supporting the statistical significance of the difference.

Interpretation. The results indicate that there is a significant difference in the Awareness Posttest scores between the control and experimental groups. The experimental group shows a significant increase in awareness compared to the control group.

Conclusion: Based on the analysis, the treatment was effective in improving awareness levels in the experimental group. The experimental group demonstrated a significant increase in Awareness Posttest scores compared to the control group. This suggests that the treatment had a positive impact on awareness levels.

1.4 The Results of Perception (Aural discrimination) Test

1.4.1 Descriptive statistics of Perception Test

Table 12: *Group Statistics for Aural Discrimination Pretest and Posttest*

	Groups	N	Mean	Std. Deviation	Std. Error Mean
Aural Discrimination Pretest	Experiment	14	7.2857	3.33809	.89214
	control group	14	7.5714	2.90131	.77541

Aural Discrimination Posttest	Experiment	14	14.7143	2.75761	.73700
	control group	14	9.4286	2.27746	.60868

Note. The group statistics table provides information on the mean, standard deviation, and standard error of the mean for the variables "Aural Discrimination Pretest " and "Aural Discrimination Posttest" in both the experiment group and the control group. Here is the interpretation:

In the experiment group, the participants had an average score of 7.2857 on the "Aural Discrimination Pretest " variable, with a standard deviation of 3.33809. The standard error of the mean is .89214, indicating the precision of the mean estimate.

Similarly, in the control group, the participants had a slightly higher average score of 7.5714 on the "Aural Discrimination Pretest" variable, with a standard deviation of 2.90131. The standard error of the mean is .77541.

After the treatment or intervention, the experiment group showed a significant improvement in their scores on the "Aural Discrimination Posttest" variable. They achieved an average score of 14.7143, with a standard deviation of 2.75761. The standard error of the mean is .73700. In contrast, the control group had an average score of 9.4286 on the "Aural Discrimination Posttest" variable, with a standard deviation of 2.27746. The standard error of the mean is .60868.

Interpretation:

The experiment group demonstrated a substantial improvement in aural discrimination skills from the pretest (Aural Discrimination Pretest) to the posttest (Aural Discrimination Posttest). The average posttest score of 14.7143 indicates a significant enhancement in their ability to discriminate sounds. On the other hand, the control group also exhibited an improvement, but to a lesser extent. Their average posttest score of 9.4286 suggests a relatively smaller enhancement in aural discrimination skills compared to the experiment group.

The difference in the mean scores between the experiment and control groups indicates that the treatment or intervention implemented in the experiment group likely played a role in the observed improvement.

However, further statistical analysis, such as an independent t-test, would be required to determine if the difference in posttest scores between the experiment and control groups is statistically significant.

1.4.2 Inferential Statistical of Perception Test

Table 13: *Levene's Test for Equality of Variances in Aural Discrimination Pretest and Posttest*

		Levene's Test for Equality of Variances	
		F	Sig.
Aural Discrimination Pretest	Equal variances assumed Equal variances not assumed	.935	.342
Aural Discrimination Posttest	Equal variances assumed Equal variances not assumed	.174	.680

Note. The table presents the results of an independent samples t-test and Levene's test for equality of variances for the variables "Aural Discrimination Pretest" and " Aural Discrimination Posttest"

The Levene's test results suggest that the assumption of equal variances is reasonable for both the Aural Discrimination Pretest (Sig. = .342) and the Aural Discrimination Posttest (Sig. = .680). This indicates that the variances of the perception scores in both the pretest and posttest are not significantly different between the groups. These findings provide confidence in the use of an appropriate t-test to analyze the mean differences between the groups, ensuring that any observed differences in perception scores can be more confidently attributed to the treatment or intervention rather than unequal variances.

Table 14: *Independent t-test for aural discrimination*

		t-test for Equality of Means						
							95% Confidence Interval of the Difference	
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Aural Discrimination Pretest	Equal variances assumed	-.242	26	.811	-.28571	1.18202	-2.71539	2.14397
	Equal variances not assumed	-.242	25.505	.811	-.28571	1.18202	-2.71769	2.14626
Aural Discrimination Posttest	Equal variances assumed	5.530	26	.000	5.28571	.95585	3.32093	7.25050
	Equal variances not assumed	5.530	25.103	.000	5.28571	.95585	3.31751	7.25392

Note. The table presents the results of an independent samples t-test for the variables "Aural Discrimination Pretest" and " Aural Discrimination Posttest" the following is the interpretation of the table:

- The t-test for equality of means shows that there is no statistically significant difference between the groups on the " Aural Discrimination Pretest " variable. The t-value is -0.242, with degrees of freedom (df) of 26. The p-value is .811, which is greater than the conventional significance level of .05. This suggests that any observed difference in the

means between the groups is likely due to chance, and there is no significant difference in the initial scores of aural discriminations between the groups.

- The t-test for equality of means reveals a statistically significant difference between the groups on the " Aural Discrimination Posttest " variable. The t-value is 5.530, with df = 26. The p-value is .000, which is smaller than .05, indicating a significant difference. This suggests that the observed difference in the means between the groups on the post-treatment scores of aural discriminations is unlikely to be due to chance and is more likely a result of the treatment or intervention.

Overall, there is no significant difference between the groups in their initial scores of aural discriminations (Aural Discrimination Pretest). However, there is a significant difference between the groups in their post-treatment scores of aural discriminations (Aural Discrimination Posttest). This indicates that the treatment or intervention had a significant impact on improving aural discrimination skills, with one group demonstrating higher scores than the other.

1.5 The Results of Production test

1.5.1 Descriptive statistics of Production Test

Table15: *Group Statistics for Production Pretest and Posttest in Experiment and Control Groups*

	Groups	N	Mean	Std. Deviation	Std. Error Mean
Production Pretest	Experiment group	14	9.0714	2.73058	.72978
	control group	14	9.1429	3.08488	.82447
Production Posttest	Experiment group	14	15.1429	1.91581	.51202
	control group	14	12.2857	2.70124	.72194

Note. The table provides information on the group statistics for the variables "Production Pretest" and "Production Posttest".

For the experiment group, which consists of 14 participants, the average score on the "Production Pretest" variable is 9.0714. The standard deviation indicates the variability in the scores within the group, which is 2.73058. The standard error of the mean represents the precision of the average score estimate, which is 0.72978. Similarly, for the control group of 14 participants, the mean score on the " Production Pretest " variable is 9.1429, with a standard deviation of 3.08488 and a standard error of the mean of 0.82447.

In the experiment group, the average score on the " Production Posttest " variable is 15.1429, indicating a higher performance level after the intervention. The standard deviation within the group is 1.91581, suggesting relatively less variability in the scores. The standard error of the mean is 0.51202.

For the control group, the mean score on the " Production Posttest " variable is 12.2857, with a standard deviation of 2.70124 and a standard error of the mean of 0.72194.

Before the intervention, both the experiment and control groups had similar mean scores on the " Production Pretest " variable, indicating comparable performance levels.

After the intervention, the experiment group showed a higher average score on the " Production Posttest " variable compared to the control group, suggesting that the treatment had a positive impact on improving production skills.

Overall, the data suggests that the intervention had a positive impact on production performance, particularly for the experiment group. The experiment group showed a significant improvement in

production scores compared to the control group. However, to determine the statistical significance of these differences, further analysis such inferential statistics is required.

1.52 Inferential Statistics of Production Test

Table16:

Levene's Test for Equality of Variances in Production Scores

		Levene's Test for Equality of Variances	
		F	Sig.
Production Pretest	Equal variances assumed	.190	.666
	Equal variances not assumed		
Production Postest	Equal variances assumed	3.332	.079
	Equal variances not assumed		

The results of the Levene's test for the Production Pretest indicate that the assumption of equal variances is met between the experiment and control groups, suggesting similarity in the variability of initial production scores. However, for the Production Posttest, the Levene's test suggests a potential violation of the equal variances assumption, implying that there may be a difference in the variability of production scores between the two groups after the treatment or intervention.

Table 17:

Independent t-Test Analysis: Comparing Production Scores between Experimental and Control Groups

		t-test for Equality of Means						
							95% Confidence Interval of the Difference	
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Production Pretests	Equal variances assumed	-.065	26	.949	-.07143	1.10106	-2.33468	2.19182
	Equal variances not assumed	-.065	25.622	.949	-.07143	1.10106	-2.33631	2.19345
Production Posttests	Equal variances assumed	3.228	26	.003	2.85714	.88508	1.03784	4.67644
	Equal variances not assumed	3.228	23.437	.004	2.85714	.88508	1.02811	4.68617

Note. The table presents the results of an independent samples t-test and Levene's test for equality of variances for the variables " Production Pretest " and " Production Posttest

- The t-test for equality of means shows that there is no statistically significant difference between the groups on the "Production Pretest " variable. The t-value of -0.065, with degrees of freedom (df) of 26, suggests that any observed difference in the means between the groups is likely due to chance. The p-value of .949, which is greater than the conventional significance level of .05, further supports this finding. In simpler terms, the initial production scores between the experiment and control groups were not significantly different, and any slight differences observed can be attributed to random variation rather

than a meaningful distinction.

- The t-test for equality of means reveals a statistically significant difference between the groups on the " Production Posttest " variable. The t-value of 3.228, with df = 26, indicates a significant difference in the means between the groups. The p-value of .003, which is smaller than .05, confirms the significance of the observed difference. This finding suggests that the treatment or intervention had a significant impact on improving production skills, with one group demonstrating higher scores than the other.

In summary, the initial production scores (Production Pretest) did not show a significant difference between the groups, indicating that they were comparable before the intervention. However, after the treatment (Production Posttest), there was a significant difference in the production scores between the experiment and control groups, suggesting that the intervention had a positive effect on improving production skills.

1.6 Questionnaire results

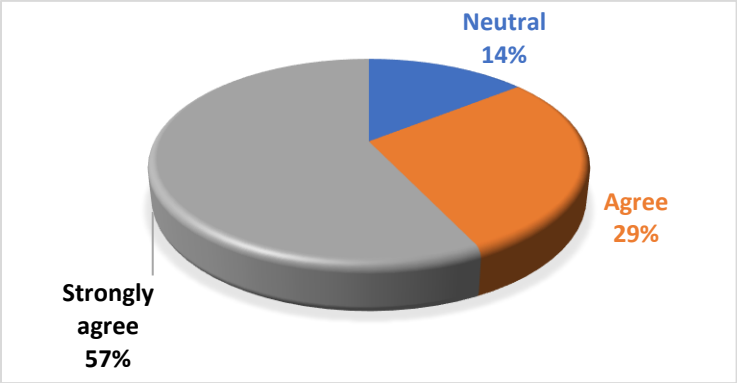
Part One: Perceptions of Text-To-Speech (TTS).

1. The input provided by TTS technology was clear and natural.

Table 18: *Students Evaluation of TTS Technology Input: Perceived Clarity and Naturalness*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Neutral	2	14.3%	14.3	14.3
	Agree	4	28.6%	28.6	42.9
	Strongly agree	8	57.1%	57.1	100.0
	Total	14	100.0	100.0	

Figure 2: Students Evaluation of TTS Technology Input: Perceived Clarity and Naturalness



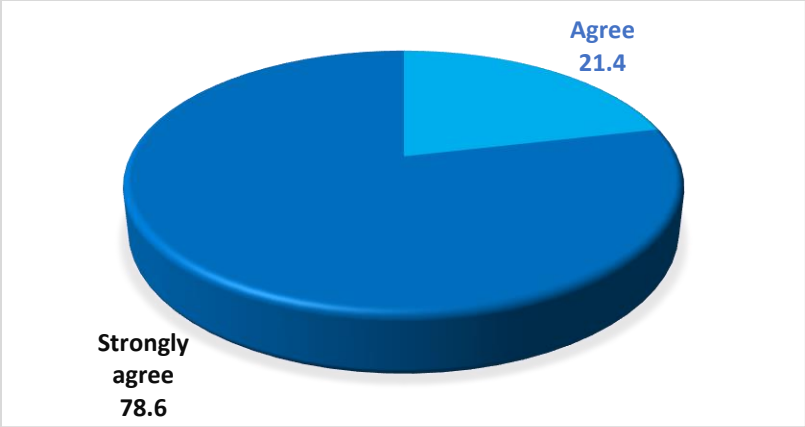
The aim of the question (Q1) was to gather feedback on the clarity and naturalness of the input provided by TTS (Text-to-Speech) technology. The data presented shows that a significant majority of respondents (85.7%) either agreed or strongly agreed that the TTS technology's input was clear and natural. Among the respondents, 57.1% strongly agreed, indicating a high level of satisfaction and positive perception of the TTS technology's performance. A small portion of respondents (14.3%) chose the neutral option, indicating a neutral opinion which could be attributed to individual preferences for voice selection. Overall, the analysis suggests a positive evaluation of the TTS technology's ability to generate understandable and natural speech output, effectively fulfilling the aim of assessing clarity and naturalness.

Item 2 : The TTS tool was easy to use.

Table 19: *Participants Perception of Ease of Use for the TTS Tool*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	3	21.4%	21.4	21.4
	Strongly agree	11	78.6%	78.6	100.0
	Total	14	100.0%	100.0	

Figure 3: *Participants Perception of Ease of Use for the TTS Tool*



The goal of asking question two was to assess the ease of use of the TTS (Text-to-Speech) tool. The data from the table indicates that a significant majority of respondents (78.6%) strongly agreed that the TTS tool was easy to use, while a smaller portion (21.4%) agreed with the statement. This indicates a high level of satisfaction and positive perception of the tool's usability. In summary, the analysis reveals that the TTS tool was perceived as easy to use by the majority of respondents, implying that it was user-friendly and intuitive, allowing users to interact with it effortlessly.

Item3: The TTS tool assisted me in accurately identifying and distinguishing between different variations of the final "-ed" sound.

Table 20: *Evaluation of TTS Tool's Assistance in Identifying and Distinguishing "-ed" Sound Variations*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Neutral	2	14.3	14.3	14.3
	Agree	4	28.6	28.6	42.9
	Strongly agree	8	57.1	57.1	100.0
	Total	14	100.0	100.0	

The aim of the question was to assess the effectiveness of the TTS (Text-to-Speech) tool in assisting users in accurately identifying and distinguishing between different variations of the final "-ed" sound. The data presented in the table reveals that the majority of respondents (85.7%) either agreed or strongly agreed that the TTS tool was helpful in this regard. Specifically, 57.1% of respondents strongly agreed that the TTS tool assisted them in accurately identifying and distinguishing between different "-ed" sound variations. A smaller portion of respondents (14.3%) chose the neutral option, indicating a neutral opinion. Overall, the analysis indicates a positive perception of the TTS tool's effectiveness in helping learners with the accurate identification and differentiation of "-ed" sounds, aligning with the aim of the question.

Item 4: Would you recommend the TTS tool to other English language learners for improving their pronunciation skills?

Table 21: *Recommendation of TTS Tool for English Language Learners' Pronunciation Improvement*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	14	100.0	100.0	100.0

Based on the data presented in the table, all respondents (100%) indicated that they would recommend the TTS (Text-to-Speech) tool to other English language learners for improving their pronunciation skills. The unanimous agreement among the participants suggests a strong consensus regarding the effectiveness of the tool in improving pronunciation. This collective recommendation highlights the perceived value and efficacy of the TTS tool as a valuable resource for English language learners seeking to refine their pronunciation abilities. Consequently, based on the unanimous agreement expressed by the respondents, it can be concluded that the TTS tool

is highly recommended for other English language learners aiming to enhance their pronunciation skills.

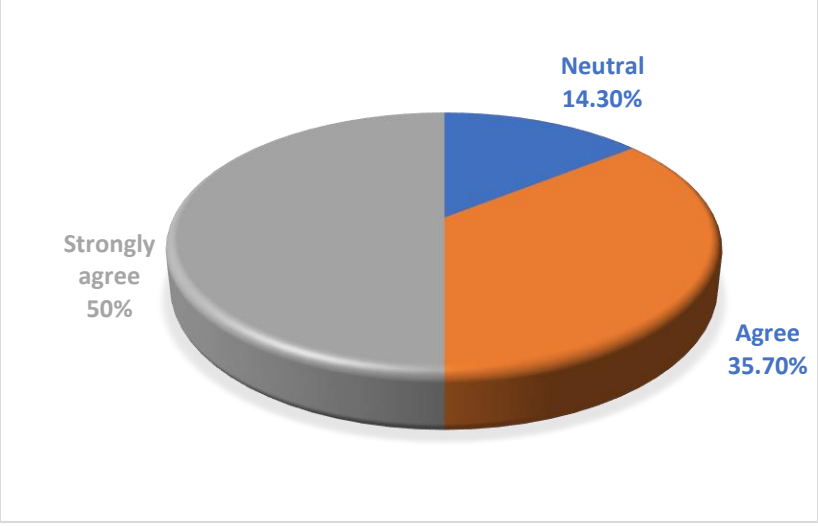
Part Two: Perceptions of Automatic Speech Recognition (ASR).

Item 5: The Output provided by ASR technology was accurate.

Table 22: *Evaluation of ASR Technology Output: Perceived Accuracy*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Neutral	2	14.3%	14.3	14.3
	Agree	5	35.7%	35.7	50.0
	Strongly agree	7	50.0%	50.0	100.0
	Total	14	100.0%	100.0	

Figure 4: *Evaluation of ASR Technology Output: Perceived Accuracy*



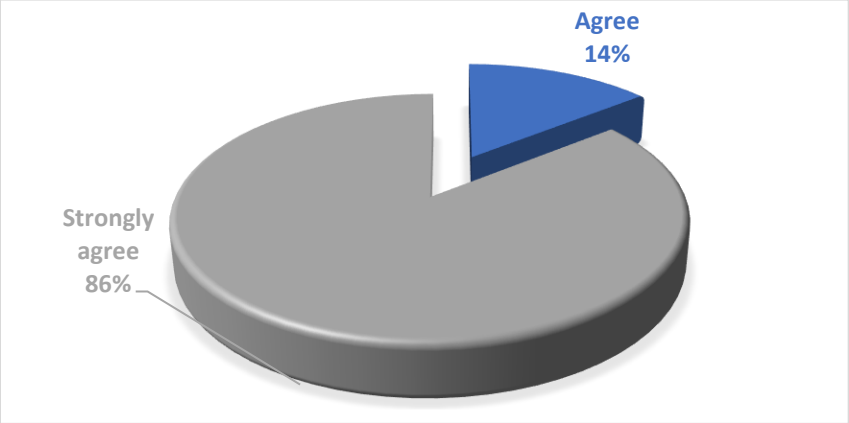
The aim of question 5 was to assess the ASR (Automatic Speech Recognition) technology's ability to accurately transcribe learners' speech. The interpretation of the data from the table suggests that the respondents had a generally positive perception of the accuracy of the output provided by ASR (Automatic Speech Recognition) technology. The fact that the majority of respondents either agreed or strongly agreed (85.7%) that the ASR technology's output was accurate indicates a favorable evaluation of its performance in accurately transcribing or interpreting speech. Additionally, the significant proportion of respondents (50.0%) who strongly agreed highlights a high level of satisfaction and confidence in the accuracy of the ASR technology's output. The small number of respondents who chose the neutral option (14.3%) may indicate a lack of strong opinion or uncertainty about the accuracy of the ASR technology's output. Overall, the interpretation suggests that the ASR technology demonstrated a satisfactory level of accuracy in producing output, as perceived by the respondents.

Item 6: ASR is a helpful tool to practice pronunciation

Table 23: Perception of ASR as a useful Tool for Pronunciation Practice

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	2	14.3%	14.3	14.3
	Strongly agree	12	85.7%	85.7	100.0
	Total	14	100.0%	100.0	

Figure 5: *Perception of ASR as a useful Tool for Pronunciation Practice*



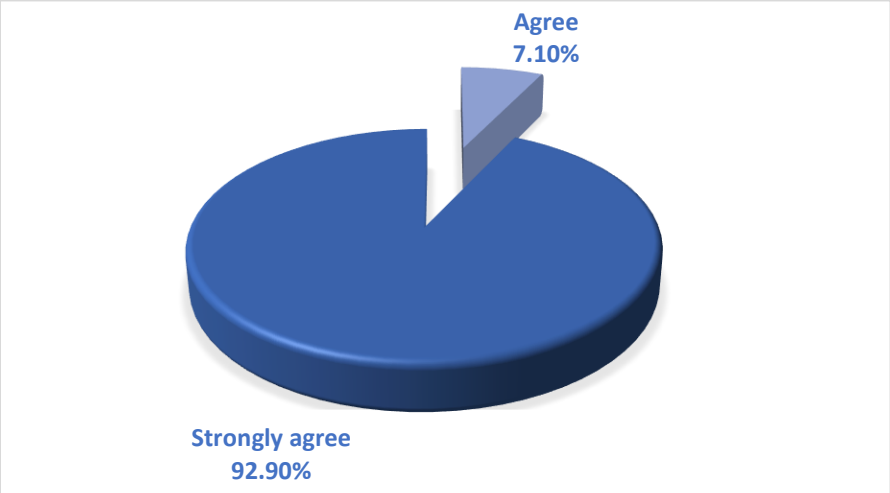
The aim of the statement/question was to gather feedback on the perception of the respondents regarding the usefulness of ASR (Automatic Speech Recognition) as a tool for practicing pronunciation. Based on the data, it can be interpreted that the majority of respondents (85.7%) strongly agreed that ASR is a valuable tool for practicing pronunciation. This indicates a high level of satisfaction and positive perception of ASR's effectiveness in assisting with pronunciation practice. The small portion of respondents who agreed (14.3%) suggests that they also recognize the usefulness of ASR for pronunciation practice, although to a lesser extent. The analysis suggests that the majority of respondents perceive ASR as a helpful tool for practicing pronunciation. This implies that ASR technology can be a beneficial resource for learners seeking to improve their pronunciation skills. The high percentage of strong agreement indicates a strong consensus among the respondents regarding the effectiveness of ASR for pronunciation practice.

Item 6: Using ASR for pronunciation practice is more effective than traditional methods (e.g., repeating after a teacher)?

Table 24: *Perception of ASR's Effectiveness for Pronunciation Practice Compared to Traditional Methods*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	1	7.1%	7.1	7.1
	Strongly agree	13	92.9%	92.9	100.0
	Total	14	100.0%	100.0	

Figure 6: *Perception of ASR's Effectiveness for Pronunciation Practice Compared to Traditional Methods*



The aim of the question was to assess the respondents' perception of the effectiveness of using ASR (Automatic Speech Recognition) for pronunciation practice compared to traditional methods like repeating after a teacher. The data from the table indicates that a significant majority of respondents (92.9%) strongly agreed that ASR is more effective for pronunciation practice, while a small portion (7.1%) agreed. This suggests a high level of confidence and positive perception of ASR's superiority over traditional methods. This perception may be attributed to the benefits of ASR, such as its ability to provide immediate, objective feedback and enable personalized, self-paced learning. Additionally, ASR offers a private and accessible platform for practicing pronunciation without anxiety or time constraints.

Item 8: Was it easy to use ASR to practice your pronunciation?

Table 25: Users' Perception of Ease of Use for ASR in Pronunciation Practice

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	14	100.0%	100.0	100.0

Based on the data presented in the table, all respondents (100%) indicated that it was easy to use ASR (Automatic Speech Recognition) to practice their pronunciation. The unanimous agreement among the participants suggests a positive experience with the usability of ASR for pronunciation practice. This implies that the ASR technology was user-friendly and accessible, allowing learners to engage in pronunciation practice comfortably and without difficulty. In summary, the analysis indicates a high level of ease in using ASR for pronunciation practice, as indicated by the unanimous agreement among the respondents.

Item 9: Would you recommend ASR as a pronunciation learning tool to other learners?

Table 26: *Recommendation of ASR as a Pronunciation Learning Tool for Other Learners*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	14	100.0	100.0%	100.0

Based on the data presented, all participants (100%) indicated that they would recommend ASR (Automatic Speech Recognition) as a pronunciation learning tool to other learners. The unanimous agreement among the respondents emphasizes their collective belief in the effectiveness of ASR for pronunciation improvement. This suggests that the participants found ASR to be a valuable and beneficial tool for enhancing their pronunciation skills. Therefore, based

on the unanimous recommendation from the participants, it can be concluded that ASR is highly recommended as a pronunciation learning tool for other learners.

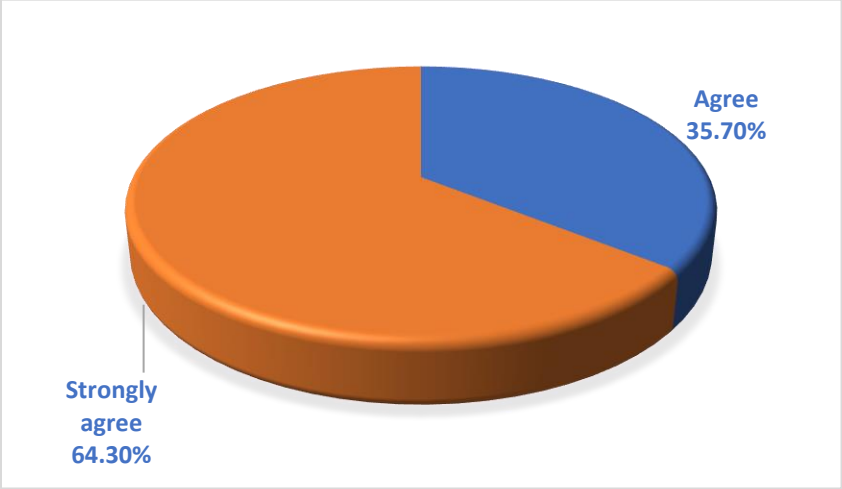
Part Three: Perceptions of Text-To-Speech (TTS) and Automatic Speech Recognition (ASR).

Item 10: TTS and ASR allowed me to evaluate my own pronunciation (for example, to decide whether my pronunciation was correct or incorrect).

Table 27: *Evaluation of TTS and ASR for Self-Assessment of Pronunciation Accuracy*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	5	35.7%	35.7	35.7
	Strongly agree	9	64.3%	64.3	100.0
	Total	14	100%	100.0	

Figure 7: *Evaluation of TTS and ASR for Self-Assessment of Pronunciation Accuracy*



The aim of the question was to gather feedback on the extent to which TTS (Text-to-Speech) and ASR (Automatic Speech Recognition) technologies enabled the respondents to evaluate their own pronunciation. The question specifically aimed to assess whether these tools provided the respondents with the ability to determine the correctness or incorrectness of their pronunciation. Based on the data presented, a significant majority of respondents (64.3%) strongly

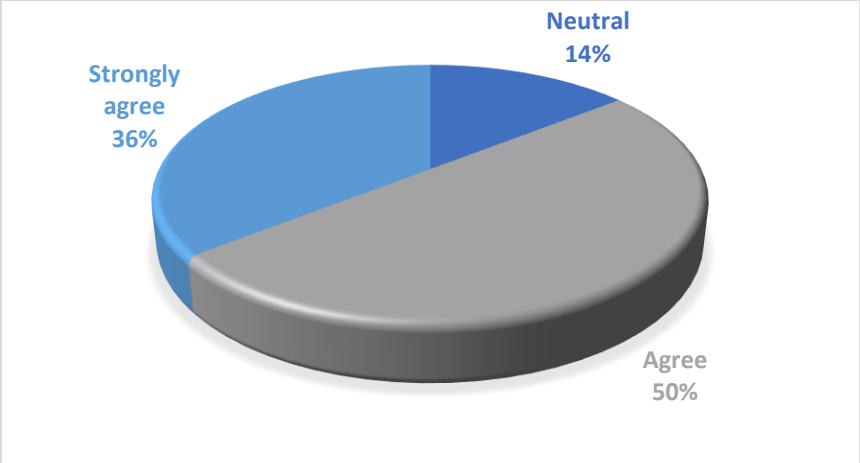
agreed and 35.7% agreed that both TTS (Text-to-Speech) and ASR (Automatic Speech Recognition) allowed them to evaluate their own pronunciation. This indicates that the respondents found TTS and ASR tools useful for assessing the correctness of their pronunciation. The high percentage of strong agreement suggests a positive perception of these technologies' ability to provide self-evaluation opportunities and support in improving pronunciation skills. Overall, the data highlights the effectiveness of TTS and ASR in helping learners assess their pronunciation accuracy.

Item 11: I felt more comfortable practicing pronunciation with TTS and ASR than I would in front of other students.

Table 28: *Comfort Level in Pronunciation Practice with TTS and ASR Compared to Practice in Front of Other Students*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Neutral	2	14.3%	14.3	14.3
	Agree	7	50.0%	50.0	64.3
	Strongly agree	5	35.7%	35.7	100.0
	Total	14	100.0%	100.0	

Figure 8: *Comfort Level in Pronunciation Practice with TTS and ASR Compared to Practice in Front of Other Students*



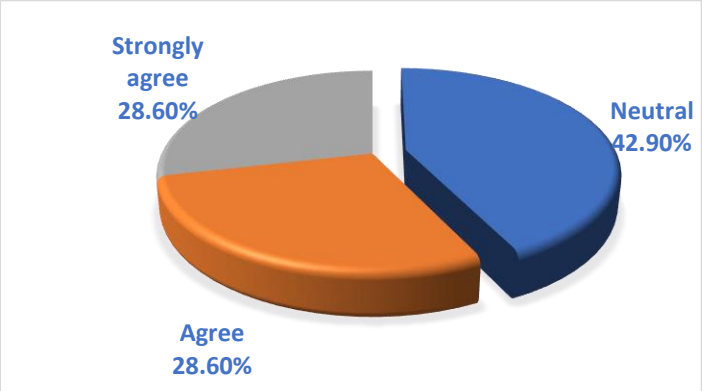
The aim of the question was to evaluate the respondents' comfort level while practicing pronunciation using TTS (Text-to-Speech) and ASR (Automatic Speech Recognition) technologies compared to practicing in front of other students. The data from the table indicates that a majority of respondents (85.7%) either agreed or strongly agreed that they felt more comfortable using TTS and ASR for pronunciation practice. This suggests that TTS and ASR technologies provide a supportive and non-judgmental environment that enhances learners' comfort during pronunciation practice. The small portion of respondents who chose the neutral option suggests a neutral stance on the comfort aspect. In summary, the analysis indicates a positive perception of TTS and ASR technologies in creating a comfortable space for learners to practice pronunciation, alleviating potential discomfort that may arise when practicing in front of others.

Item 12 : I felt more comfortable practicing pronunciation with TTS and ASR than I would in front of the teacher.

Table 29: *Response Percentages and Frequencies for Item 12.*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Neutral	6	42.9%	42.9	42.9
	Agree	4	28.6%	28.6	71.4
	Strongly agree	4	28.6%	28.6	100.0
	Total	14	100.0	100.0	

Figure 9: *Percentage of Students Agreement or Disagreement With Item 12*



The aim of the question was to assess the respondents' level of comfort while practicing pronunciation using TTS (Text-to-Speech) and ASR (Automatic Speech Recognition) technologies compared to practicing in front of a teacher. The data from the table indicates that a significant proportion of respondents (57.1%) either agreed or strongly agreed that they felt more comfortable practicing pronunciation with TTS and ASR than in front of a teacher. This suggests that TTS and ASR technologies provide a more relaxed and less intimidating environment for learners to practice their pronunciation skills. On the other hand, a considerable number of respondents (42.9%) chose the neutral option, indicating that they neither agreed nor disagreed

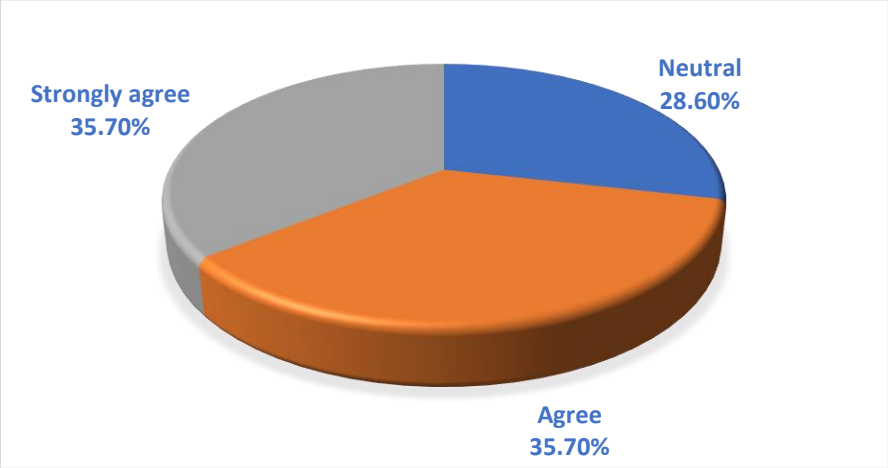
with the statement. This implies that there is a variation in individual experiences and preferences when it comes to practicing pronunciation, and some learners may still prefer the guidance and presence of a teacher despite the benefits of using TTS and ASR. In summary, the analysis suggests that a majority of respondents felt more comfortable practicing pronunciation with TTS and ASR compared to in front of a teacher, indicating the potential benefits of these technologies in creating a supportive learning environment.

Item 13: TTS and ASR enhanced my engagement beyond the classroom(facilitating learning and practice pronunciation at any time and any location).

Table 30: *Response Percentages and Frequencies for Item 13.*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Neutral	4	28.6%	28.6	28.6
	Agree	5	35.7%	35.7	64.3
	Strongly agree	5	35.7%	35.7	100.0
	Total	14	100.0%	100.0	

Figure 10: *Percentage of Students Agreement or Disagreement with Item 13.*



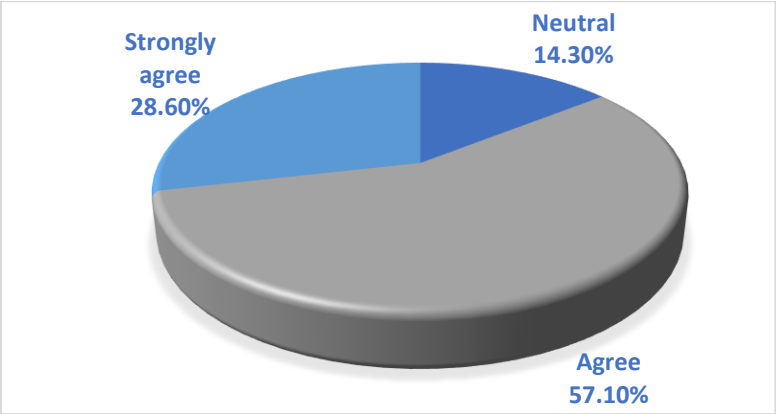
The aim of the question was to assess the respondents' perception of whether TTS (Text-to-Speech) and ASR (Automatic Speech Recognition) technologies enhanced their engagement beyond the classroom or not , allowing for learning and practicing pronunciation at any time and any location. The data from the table indicates that a significant proportion of respondents (71.4%) either agreed or strongly agreed that TTS and ASR facilitated their engagement beyond the classroom. This suggests that these technologies provide learners with the flexibility and convenience to continue learning and practicing pronunciation outside of traditional classroom settings. Additionally, a portion of respondents (28.6%) chose the neutral option, indicating a neutral opinion on the statement. This may imply that some individuals may not have fully utilized or experienced the benefits of TTS and ASR in enhancing their engagement outside the classroom. Overall, the analysis suggests a positive perception among the respondents regarding the ability of TTS and ASR to enhance engagement beyond the classroom, providing learners with opportunities for continuous learning and practice.

Item 14 : I feel confident in my ability to perceive and produce the correct pronunciation of final 'ed' after using the TTS and ASR tools

Table 31: *Response Percentages and Frequencies for Item 14.*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Neutral	2	14.3%	14.3	14.3
	Agree	8	57.1%	57.1	71.4
	Strongly agree	4	28.6%	28.6	100.0
	Total	14	100.0%	100.0	

Figure 11: *Percentage of Students Agreement or Disagreement with Item 14.*



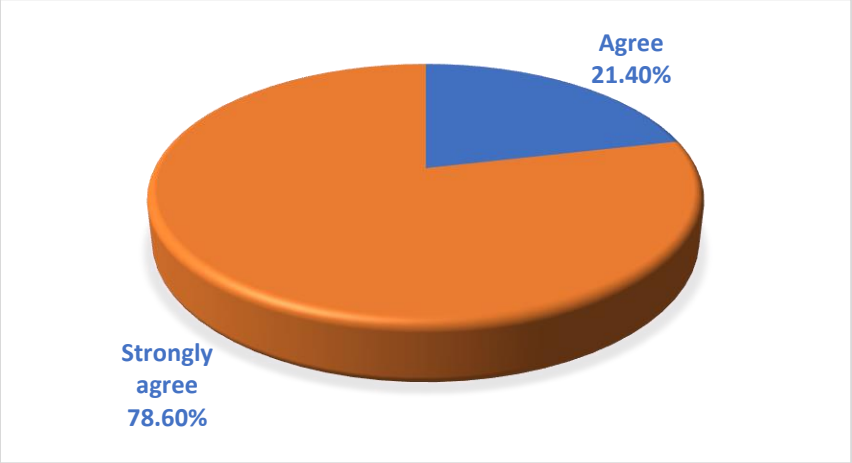
The aim of the question was to assess the respondents' confidence in perceiving and producing the correct pronunciation of the final 'ed' sound after using the TTS (Text-to-Speech) and ASR (Automatic Speech Recognition) tools. The data presented in the table indicates that a majority of respondents (85.7%) either agreed or strongly agreed that they felt confident in their ability to perceive and produce the correct pronunciation of the final 'ed' sound. This suggests that the use of TTS and ASR tools has positively influenced their confidence in mastering this particular pronunciation aspect. The small portion of respondents (14.3%) who chose the neutral option indicates a neutral stance or uncertainty in their confidence level. Overall, the analysis reveals a positive impact of TTS and ASR tools on the respondents' confidence in perceiving and producing the correct pronunciation of the final 'ed' sound.

Item 15 : TTS and ASR tools helped me become a more autonomous learner.

Table 32: *Response Percentages and Frequencies for Item 15.*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	3	21.4%	21.4	21.4
	Strongly agree	11	78.6%	78.6	100.0
	Total	14	100.0	100.0	

Figure 12 : Percentage of Students Agreement or Disagreement with Item 15.



The aim of the question was to examine the extent to which the use of TTS (Text-to-Speech) and ASR (Automatic Speech Recognition) tools contributed to the respondents' autonomy as language learners. The data from the table indicates that a significant majority of respondents (78.6%) strongly agreed and 21.4% agreed that TTS and ASR tools helped them become more autonomous learners. This suggests that these tools have played a crucial role in fostering self-directed learning by providing learners with the means to practice pronunciation and receive immediate feedback independently. The high percentage of strong agreement highlights the perceived value of TTS and ASR tools in promoting autonomy and empowering learners to take control of their own learning process. Overall, the analysis reveals a positive impact of TTS and ASR tools on the respondents' autonomy as language learners.

Item 16: Did you find it helpful to use both TTS and ASR together to improve your perception and production of the final 'ed' sound in English verbs?

Table 33: *Students Answers of Item 16.*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	14	100.0%	100.0	100.0

Based on the data provided, all respondents (100%) found it helpful to use both TTS (Text-to-Speech) and ASR (Automatic Speech Recognition) together to improve their perception and production of the final 'ed' sound in English verbs. This unanimous agreement suggests that the combination of TTS and ASR technologies has proven to be beneficial in enhancing learners' understanding and ability to accurately pronounce the final 'ed' sound. The positive response indicates that utilizing these tools in conjunction can provide valuable support and guidance for learners in mastering this aspect of English pronunciation.

Item 17: Would you recommend the TTS tool to other English language learners for improving their pronunciation skills?

Table 34: *Students Answers of Item 17.*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	14	100.0%	100.0	100.0

Based on the data provided, all participants (100%) expressed their recommendation for using TTS (Text-to-Speech) and ASR (Automatic Speech Recognition) technologies to learn pronunciation. This unanimous agreement signifies the positive perception of these tools in pronunciation improvement. The unanimous recommendation suggests that TTS and ASR can be effective resources for learners seeking to enhance their pronunciation skills. The respondents' positive feedback supports the idea that TTS and ASR offer valuable opportunities for practice, feedback, and self-assessment in language learning.

Item 18: Do you think you should spend more time using Neural reader app to improve your pronunciation outside the class?

Table 35: *Students answers of item 18.*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	14	100.0%	100.0	100.0

Based on the data presented, all participants (100%) agreed that spending more time using the Neural reader app to improve pronunciation outside of the class would be beneficial. This unanimous agreement indicates a shared belief in the effectiveness of the app for enhancing pronunciation skills beyond classroom instruction. The participants' consensus suggests that utilizing the app for additional practice and exposure to pronunciation is considered valuable and can contribute to pronunciation improvement.

Item 19: Do you think Neural reader application (ASR and TTS) matches your learning style, considering some students are visual learners while others prefer auditory learning?

Table 36: *Response Percentages and Frequencies for Item 19.*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	2	14.3%	14.3	14.3
	yes	12	85.7%	85.7	100.0
	Total	14	100.0%	100.0	

Based on the data provided, the majority of respondents (85.7%) indicated that the Neural reader application (ASR and TTS) matches their learning style. This suggests that a significant proportion of students find the combination of auditory and visual components offered by the

application to be suitable for their learning preferences. However, it is worthnoting that a small portion of respondents (14.3%) disagreed or felt that the application does not align with their learning style. This implies that there may be variations among individuals in terms of their preferred learning modalities and the effectiveness of the Neural reader application in meeting their specific needs.

Item 20: Which features did you utilize while using the application.

Table 34: *Features used in the application.*

		Responses		
		N	Percent	Percent of Cases
Application Features	change the playback	8	33.3%	57.1%
	opt for a different voice	11	45.8%	78.6%
	Select an alternate accent	5	20.8%	35.7%
Total		24	100.0%	171.4%

The results indicate that several features were of importance to the respondents: 33.3% of respondents utilized the feature to change the playback settings, indicating their desire for control over the audio speed or pace to enhance their learning experience. 45.8% of respondents preferred the option to choose a different voice, suggesting their inclination towards personalization and variation in the auditory component of the application. 20.8% of respondents expressed an interest in selecting an alternate accent, recognizing the value of exposure to diverse English accents for their learning journey.

It should be noted that respondents were able to choose multiple features, resulting in a total frequency count exceeding the number of respondents. These findings highlight the specific features which respondents found valuable and actively used in the application.

Discussion

The present study investigated the effectiveness of Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) as tools for improving the pronunciation of English past -ed, focusing on awareness, perception, and production. Additionally, the study sought to explore participants' perceptions and attitudes towards both tools.

Regarding the first research question, which examined participants' development across the three levels of testing (awareness, perception, and production). The results of the t-test analysis indicated significant improvements in the post-test scores of the experimental group compared to the control group. Specifically, the experiment group demonstrated significant improvements in all three levels: awareness, perception, and production. These findings support that TTS and ASR can effectively aid learners in acquiring the target pronunciation feature and improve their L2 pronunciation.

In terms of awareness, the observed improvements in awareness suggest that participants in the experimental group became more conscious of the pronunciation differences associated with English past -ed then the control group. This heightened awareness can be attributed to the exposure and feedback provided by the TTS and ASR tools, allowing learners to recognize and differentiate the target pronunciation. This finding aligns with recent research (Moussalli & Cardoso, 2021; Khademi & Cardoso, 2022) indicating that TTS and ASR can assist learners in developing their phonological and morphophonemic awareness.

Considering that awareness is considered crucial in the three main stages of pronunciation development, as outlined by Celce-Murcia et al. (2010), it can be concluded that using TTS and ASR as pedagogical tools has the potential to initiate and facilitate the subsequent stages of

perception and production in developing the morphophonemics of the English past tense marker "-ed."

Regarding perception, aural discrimination, the results suggest that access to synthesized L2 input contribute to a better perception of sentential verbal tense, as well as the identification of regular past -ed in English. The findings indicate that the utilization of TTS and ASR significantly contributed to the improvement of participants' perception skills in identifying and distinguishing the three allomorphs of past -ed. These results align with the assertion made Van Lieshout and Cardoso (2022) that TTS and ASR speech capabilities can effectively support the development of aural skills.

One possible explanation for the positive impact on perception is the nature of TTS synthesis. The synthesized L2 input provided by TTS may facilitate better perception of past “ed” tense and enhance learners' ability to identify regular past -ed forms in English. The quality of current TTS synthesis, combined with the inherent benefits of the technology, such as promoting repetition, practice, and creating a stress-free learning environment, may explain the significant improvement observed in participants' ability to discriminate between past and non-past constructions, as well as among the three -ed allomorphs. The findings suggest that TTS and ASR technologies have a positive impact on learners' perception skills, enabling them to better distinguish and identify the allomorphs of past -ed. The quality of TTS synthesis and the advantages offered by the technology contribute to improved discrimination abilities and create a conducive learning environment. Integrating TTS and ASR into pronunciation instruction can effectively support learners in developing their aural skills and enhancing their perception of target language features.

Another point made in the discussion section highlights the significant progress observed in the experimental group regarding production skills. This finding suggests that the use of TTS and ASR technologies in pronunciation instruction can effectively support learners in producing the target sounds accurately.

The implementation of TTS and ASR provides learners with opportunities for practice and feedback on their pronunciation. The immediate feedback from ASR systems allowed learners to identify and correct their pronunciation an error, thereby enhancing their ability to produce the target sounds more accurately.

Overall, the use of TTS and ASR technologies in present study positively influenced participants' performance in all three stages, providing empirical evidence for the effectiveness of these tools in pronunciation instruction. These findings have important pedagogical implications, as they suggest that TTS and ASR can be valuable resources for learners to enhance their pronunciation skills and extend the learning environment beyond the confines of the classroom.

Moreover, valuable insights were obtained regarding the perceptions of learners regarding the effectiveness and usability of TTS and ASR technologies for improving pronunciation skills. The respondents expressed a positive perception of both TTS and ASR tools, highlighting their potential as effective resources for pronunciation practice and self-assessment.

Regarding TTS technology, the majority of respondents agreed that the input generated by TTS was clear and natural. This indicates a high level of satisfaction with the clarity and naturalness of the speech output, suggesting that TTS technology is successful in generating understandable and natural voices. Additionally, the respondents found TTS easy to use, highlighting its user-friendly nature. The positive perception of the ease of use indicates that

learners can interact with the TTS tool effortlessly, making it accessible for learners with varying technological skills.

Furthermore, the TTS tool was found to be effective in assisting learners with accurately identifying and distinguishing between different variations of the final "-ed" sound. The majority of respondents agreed that TTS helped them in this regard, indicating its value in supporting learners' understanding and practice of this specific pronunciation aspect. This finding aligns with the aim of the question, suggesting that TTS can be a useful tool for learners striving to improve their pronunciation accuracy.

The unanimous agreement among respondents to recommend the TTS tool to other English language learners for improving pronunciation skills highlights the perceived value and efficacy of TTS technology. This collective recommendation emphasizes the confidence and satisfaction of the respondents with the TTS tool, indicating its potential as a valuable resource for learners aiming to enhance their pronunciation abilities.

Moving on to the perceptions of ASR technology, the results indicate a generally positive perception of its accuracy in transcribing learners' speech. The majority of respondents agreed that the output provided by ASR was accurate, suggesting a favorable evaluation of its performance in accurately interpreting speech. This perception of accuracy is crucial, as learners rely on the ASR tool to provide reliable feedback on their pronunciation.

Furthermore, ASR technology was perceived as a helpful tool for pronunciation practice. The majority of respondents agreed that ASR is a valuable tool for practicing pronunciation, indicating its usefulness in supporting learners' pronunciation improvement. This finding highlights the potential benefits of ASR, such as providing immediate and objective feedback,

enabling personalized and self-paced learning, and offering a private and accessible platform for practice.

Interestingly, the respondents strongly believed that using ASR for pronunciation practice is more effective than traditional methods, such as repeating after a teacher. This perception may be attributed to the benefits of ASR, including its ability to provide individualized feedback and facilitate autonomous learning. The perceived effectiveness of ASR over traditional methods underscores the potential of technology in enhancing pronunciation practice and fostering learners' confidence.

The unanimous agreement among respondents to recommend ASR as a pronunciation learning tool to other learners further reinforces its perceived effectiveness. This collective recommendation signifies the participants' belief in the value of ASR for pronunciation improvement and their positive experience using it.

Moving on to the perceptions of TTS and ASR used together, the results indicate that these technologies allowed learners to evaluate their own pronunciation accurately. The majority of respondents agreed that TTS and ASR facilitated self-evaluation, providing learners with the ability to determine the correctness or incorrectness of their pronunciation. This finding suggests that the combination of TTS and ASR tools supports learners in monitoring and assessing their own pronunciation accuracy.

The respondents also felt more comfortable practicing pronunciation with TTS and ASR compared to in front of other students or teachers. This perception highlights the supportive and non-judgmental environment created by TTS and ASR technologies, allowing learners to practice pronunciation with increased comfort and confidence. The comfort factor is crucial for learners to

engage in meaningful practice and overcome potential anxiety or self-consciousness associated with pronunciation practice.

Additionally, the participants felt that TTS and ASR tools enhanced their engagement beyond the classroom. This finding suggests that the use of these technologies extends learners' practice opportunities outside formal instructional settings, enabling continuous and independent learning. By providing accessible and interactive resources, TTS and ASR technologies can potentially motivate learners to engage in more frequent and consistent pronunciation practice.

The majority of respondents also expressed confidence in their ability to perceive and produce the correct pronunciation of the final 'ed' sound after using TTS and ASR tools. This perceived improvement in pronunciation skills further supports the effectiveness of TTS and ASR technologies in facilitating learners' progress in mastering specific pronunciation aspects.

The participants indicated that TTS and ASR tools helped them become more autonomous learners. This finding suggests that the combination of these technologies empowers learners to be more autonomous in their pronunciation practice, fostering their independence and self-directed learning skills. The perception of increased autonomy aligns with the broader goal of language learning, as learners who can independently identify and address their pronunciation errors are better equipped to continuously improve their speaking skills.

The findings of the questionnaire indicate positive perceptions and high satisfaction with both TTS and ASR technologies for improving pronunciation skills. The participants recognized the clarity and naturalness of the TTS output, the accuracy of the ASR transcriptions, and the ease of use of both tools. They found TTS and ASR valuable in enhancing pronunciation practice, supporting self-evaluation, and fostering autonomy. The combination of TTS and ASR technologies was seen as particularly beneficial, providing learners with comfortable and engaging

practice opportunities outside the classroom. The unanimous agreement and recommendations from the respondents underscore the potential of TTS and ASR technologies as effective resources for pronunciation improvement in language learning contexts.

Conclusion

This chapter was devoted to the data collection and analysis of the current study. The researcher employed two data collection methods: a pretest-posttest and a participants' questionnaire. The results derived from the tests were analyzed using both descriptive and inferential statistics. The analysis of the results revealed that the alternative hypothesis is accepted, as the calculated t-value was less than the critical value. Consequently, the researcher concluded that TTS and ASR are effective in improving the pronunciation of EFL learners. Overall, based on the findings and the statistical analysis, it can be concluded that the use of TTS and ASR technologies as pronunciation learning tools have a positive impact on learners' pronunciation skills, specifically in the context of English past -ed. The findings highlight the effectiveness of TTS and ASR in promoting awareness, perception, and production of the target pronunciation feature. Learners' perceptions of these technologies were overwhelmingly positive, indicating their potential as valuable resources for pronunciation instruction and self-directed practice.

Implications of the study

Implications for Students:

- Improved Pronunciation: The combined use of TTS and ASR technologies provides students with a valuable resource for improving their pronunciation skills. By engaging in targeted practice and receiving immediate feedback, students can enhance their awareness, perception, and production of specific pronunciation features, such as the English past -ed allomorph. This can lead to greater clarity and accuracy in their spoken English.
- Enhanced Self-Evaluation: TTS and ASR technologies enable students to assess their own pronunciation accurately. Through the use of these tools, students can identify their strengths and areas for improvement, allowing for self-directed learning and continuous progress. This empowers students to take ownership of their learning process and actively work towards improving their pronunciation skills.
- Increased Autonomy: The use of TTS and ASR technologies fosters learner autonomy by providing students with resources for independent practice outside of the classroom. Students can engage in pronunciation practice at their own pace and convenience, which promotes self-motivation and a sense of control over their learning. This autonomy extends beyond pronunciation skills and cultivates important self-directed learning skills that can benefit students in various areas of language acquisition.

Implications for ESP Teachers:

- Time Optimization: By incorporating TTS and ASR technologies into pronunciation instruction, teachers can optimize classroom time. These tools allow students to engage in focused practice outside of class. Teachers can allocate more time to activities that require face-to-face interaction and personalized instruction.
- Individualized Feedback: ASR technology provides teachers with an objective and

immediate assessment of students' pronunciation. This enables teachers to identify specific areas of improvement for individual students and provide targeted feedback. With access to accurate feedback, teachers can tailor their instruction to meet students' unique needs, ensuring more effective pronunciation instruction.

Implications for Materials Designers:

- Integration of TTS and ASR: Materials designers can incorporate TTS and ASR technologies into pronunciation learning materials to enhance learners' engagement and practice opportunities. By including TTS-generated audio and interactive ASR-based activities, materials can provide learners with authentic models and immediate feedback, creating a dynamic and interactive learning experience.
- Authenticity and Naturalness: TTS technology's ability to generate clear and natural speech output can be leveraged by materials designers to provide learners with high-quality L2 input. This enhances the authenticity of the learning materials and exposes learners to natural pronunciation patterns, contributing to their overall language development.

Overall, the implications of this study for students, teachers, and materials designers highlight the potential benefits of incorporating TTS and ASR technologies into pronunciation instruction. These technologies can enhance students' pronunciation skills, promote learner autonomy, optimize classroom time, enable individualized feedback, and enrich learning materials. By leveraging these technologies, stakeholders in language learning can create an engaging and effective pronunciation learning environment.

Limitations and Future Directions

The present study, focused on assessing the effectiveness of TTS and ASR in improving EFL learners' pronunciation; nevertheless, it acknowledges some difficulties and limitations:

1. The participants in this study were limited to Elementary level learners. Including learners from additional proficiency levels would provide a more comprehensive understanding of the effectiveness of TTS and ASR technologies across different proficiency levels.
2. The duration of the study was relatively short, with the instructional treatment period lasting only one week. Extending the duration of the study could provide a more thorough evaluation of the long-term effects of TTS and ASR on pronunciation improvement.
3. The study focused exclusively on the segmental level of pronunciation, neglecting other important aspects such as suprasegmental features (e.g., stress, intonation). Future studies should consider incorporating a broader range of pronunciation elements.
4. Due to technical issues like internet availability and outdated equipment, the study was not conducted in a language lab. Conducting future studies in a controlled lab environment could enhance the reliability of the findings.
5. The study employed a specific TTS application, "Neural Reader." Exploring the effectiveness of different TTS and ASR applications would provide a more comprehensive understanding of the range of tools available and their respective impact on pronunciation improvement.

General Conclusion

The present study aimed to investigate the effectiveness of Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) technologies in improving English as a Foreign Language (EFL) learners' awareness. The findings shed light on the potential benefits of these technologies in enhancing learners' pronunciation and listening skills, as well as their overall language awareness.

The study consisted of three chapters. The first chapter involved a literature review that focused on English pronunciation challenges faced by English as a Foreign Language learners. It examined various pedagogical approaches employed in English pronunciation instruction. Additionally, this chapter explored the utilization of Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) technologies in EFL pronunciation training through an examination of existing literature. The second chapter discussed the study's methodology, which utilized a quantitative approach. To collect relevant data and test the research hypothesis, two data collection methods were employed: a pretest-posttest design and a questionnaire. Both tools were carefully chosen to gather necessary data and provide insights into the research question.

After the analysis of the findings, the research questions were answered and the hypothesis has been verified. Thus, the first question, which is does the use of TTS and ASR helps EFL learners to improve their pronunciation of the final "ed" sound in terms of awareness, perception, and production, is supported by the results of the experiment. During the pretest, the participants showed a lack of knowledge regarding the pronunciation of "ed" in all aspects - awareness, perception, and production. However, after being exposed to TTS and ASR as a treatment for one week, the participants in the experimental group were able to enhance their awareness, perception, and production of the regular past "ed" sound. This improvement was confirmed in the posttest.

As a result, the alternative hypothesis stating that the experimental group would show improvement in at least one of the three levels (awareness, perception, or production) of the pronunciation of "ed" has been confirmed. This provides evidence for the effectiveness of using TTS and ASR as pronunciation learning tools to enhance the pronunciation skills of EFL learners.

The second question, aimed at exploring the learners' attitude towards the use of TTS and ASR as pronunciation learning tools, was answered through a questionnaire. The results of the questionnaire indicated that the participants had a positive attitude towards both tools.

In conclusion, this study provides empirical evidence supporting the effectiveness of TTS and ASR technologies in improving EFL learners' pronunciation, listening skills, and overall language awareness. The findings suggest that integrating TTS and ASR into language learning environments can offer learners valuable opportunities for practice, feedback, and self-reflection. The results have implications for language educators and curriculum designers, highlighting the potential of these technologies as valuable tools for promoting effective language learning and teaching. Future research should explore the optimal integration strategies, the role of learner characteristics, and the long-term effects of TTS and ASR interventions on language proficiency.

LIST OF REFERENCES

Al-Jarf, R. (2022). Text-To-Speech Software for Promoting EFL Freshman Students' Decoding Skills and Pronunciation Accuracy. *Journal of Computer Science and Technology Studies*, *4*(2), 19–30. https://doi.org/10.32996/jcsts.2022.4.2.4

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. https://doi.org/10.5430/wjel.v12n1p211

Anthony, J. L., & Francis, D. J. (2005). Development of Phonological Awareness. *Current Directions in Psychological Science*, *14*(5), 255–259. https://doi.org/10.1111/j.0963-7214.2005.00376.x

Ary, D. (2010). *Introduction to Research in Education*. Cengage Learning.

Ashwell, T., & Elam, J. R. (2017). How accurately can the Google Web Speech API recognize and transcribe Japanese L2 English learners' oral production? *The JALT CALL Journal*, *13*(1), 59–76. https://doi.org/10.29140/jaltcall.v13n1.212

Bailly, G. (2003). Close Shadowing Natural Versus Synthetic Speech. *International Journal of Speech Technology*, *6*(1), 11–19. https://doi.org/10.1023/a:1021091720511

Baker, A., & Burri, M. (2016). Feedback on Second Language Pronunciation: A Case Study of EAP Teachers' Beliefs and Practices. *Australian Journal of Teacher Education*, *41*(6), 1–19. https://doi.org/10.14221/ajte.2016v41n6.1

Benati, A. (2017). The role of input and output tasks in grammar instruction: Theoretical, empirical and pedagogical considerations. *Studies in Second Language Learning and Teaching*, *7*(3), 377–396. https://doi.org/10.14746/ssllt.2017.7.3.2

Benitez, C., Cabrera, P., Solano, L., & Espinoza, S. (2020). Improving Past Tense Pronunciation of Regular Verbs Through the Use of Audacity. *ResearchGate*. https://www.researchgate.net/publication/339737917_IMPROVING_PAST_TENSE_PRONUNCIATION_OF_REGULAR_VERBS_THROUGH_THE_USE_OF_AUDACITY_A_CASE_STUDY_OF_EFL_UNDERGRADUATE_STUDENTS_IN_ECUADOR

Bione, T., & Cardoso, W. (2020). Synthetic Voices in the Foreign Language Context. *Language Learning & Technology, 24*(1), 169–186. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44715/1/24_1_10125-44715.pdf

Bogach, N., Boitsova, E., Chernonog, S., Lamtev, A., Lesnichaya, M., Lezhenin, I., Novopashenny, A., Svechnikov, R., Tsikach, D., Vasiliev, K., Pyshkin, E., & Blake, J. (2021). Speech Processing for Language Learning: A Practical Approach to Computer-Assisted Pronunciation Teaching. *Electronics*, *10*(3), 235. https://doi.org/10.3390/electronics10030235

Cardoso, W. (2018). Learning L2 pronunciation with a text-to-speech synthesizer. *Future-proof CALL: Language Learning as Exploration and Encounters – Short Papers From EUROCALL 2018*, 16–21. https://doi.org/10.14705/rpnet.2018.26.806

Cardoso, W., Smith, G., & Garcia Fuentes, C. (2015). Evaluating text-to-speech synthesizers. *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy*. https://doi.org/10.14705/rpnet.2015.000318

Carlet, A., & Kivistö-de Souza, H. (2018). Improving L2 pronunciation inside and outside the classroom. *Ilha Do Desterro a Journal of English Language, Literatures in English and Cultural Studies, 71*(3), 99–124. https://doi.org/10.5007/2175-8026.2018v71n3p99

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010). *Teaching Pronunciation: A Course Book and Reference Guide (Cambridge Teacher Training and Development)* (2nd ed.). Cambridge University Press.

Chapelle, C. A. (2003). English Language Learning and Technology. *Language Learning and Language Teaching*. https://doi.org/10.1075/lllt.7

Chazen, D. (n.d.). ASR Transcription Software. *Verbit*. https://verbit.ai/asr-and-the-next-generation-of-transcription/

Collins, L. C., & Muñoz, C. (2016). The Foreign Language Classroom: Current Perspectives and Future Considerations. *The Modern Language Journal*, *100*(S1), 133–147. https://doi.org/10.1111/modl.12305

Collins, L., & Muñoz, C. (2016). The Foreign Language Classroom: Current Perspectives and Future Considerations. *The Modern Language Journal*, *100*(S1), 133–147. https://doi.org/10.1111/modl.12305

Couper, G. (2019). Teachers' cognitions of corrective feedback on pronunciation: Their beliefs, perceptions and practices. *System*, *84*, 41–52. https://doi.org/10.1016/j.system.2019.04.003

Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.

Crystal, D. (1991). A Dictionary of Linguistics and Phonetics. *Oxford: Blackwell*. https://doi.org/10.1002/9781444302776

Darcy, I. (2018). Powerful and Effective Pronunciation Instruction: How Can We Achieve It?. *The CATESOL Journal*, *30*(1), 13–45.

Darcy, I., Rocca, B., & Hancock, Z. (2020). A Window into the Classroom: How Teachers Integrate Pronunciation Instruction. *RELC Journal*, *52*(1), 110–127. https://doi.org/10.1177/0033688220964269

De Araújo Gomes, A. A., Cardoso, W., & De Lucena, R. M. (2018). Can TTS help L2 learners develop their phonological awareness? In *Future-proof CALL: language learning as exploration and encounters* (pp. 29–34). EUROCALL.

DeCapua, A. (2017). Grammar for Teachers. In *Springer eBooks*. https://doi.org/10.1007/978-3-319-33916-0

Derwing, T. M. (2018). The Efficacy of Pronunciation Instruction. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge Handbook of Contemporary English Pronunciation* (pp. 320–334). Milton Park: Routledge.

Derwing, T. M., & Munro, M. J. (1997). ACCENT, INTELLIGIBILITY, AND COMPREHENSIBILITY: Evidence from Four L1s. *Studies in Second Language Acquisition*, *19*(1), 1–16. http://www.jstor.org/stable/44488664

Dizon, G., & Tang, D. (2020). Intelligent personal assistants for autonomous second language learning: An investigation of Alexa. *The JALT CALL Journal*, *16*(2), 107–120. https://doi.org/10.29140/jaltcall.v16n2.273

Dubin, F., & Robinett, B. W. (1981). Teaching English to Speakers of Other Languages: Substance and Technique. *Language*, *57*(1), 222. https://doi.org/10.2307/414299

Dutoit, T. (1997). An Introduction to Text-to-Speech Synthesis. In *Text, speech and language technology*. Springer Nature (Netherlands). https://doi.org/10.1007/978-94-011-5730-8

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

Flege, J. E. (2011). Assessing constraints on second-language segmental production and perception. In *De Gruyter eBooks*. https://doi.org/10.1515/9783110895094.319

Frazier, L. L., & Mills, R. (2015). *Northstar Listening Speaking 2 Sb W/ Interactive Sb and Myenglishlab*. Pearson Education ESL.

Gilakjani, A.P. (2012). A Study of Factors Affecting EFL Learners' English Pronunciation Learning and the Strategies for Instruction.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2012). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer Assisted Language Learning*, *27*(1), 70–105. https://doi.org/10.1080/09588221.2012.700315

Grim, F., & Sturm, J. L. (n.d.). Where does pronunciation stand? *Pronunciation in Second Language Learning and Teaching*, 51–65.

Grimshaw, J., Bione, T., & Cardoso, W. (2018). Who's got talent? Comparing TTS systems for comprehensibility, naturalness, and intelligibility. In *Future-proof CALL: language learning as exploration and encounters* (pp. 83–88). https://doi.org/10.14705/rpnet.2018.26.817

Handely, Z. (2013). Text-to-Speech Synthesis in Computer-Assisted Language Learning. In C. A. Chapell (Ed.), *The encyclopedia of applied linguistics* (pp. 5846–5851). New York: Wiley-Blackwell.

Hismanoglu, M., & Hismanoglu, S. (2010). Language teachers' preferences of pronunciation teaching techniques: traditional or modern? *Procedia - Social and Behavioral Sciences*, *2*(2), 983–989. https://doi.org/10.1016/j.sbspro.2010.03.138

Hornby, A. S. (1995). *Oxford Advanced Learner's Dictionary.* Oxford: Oxford University Press.

Inceoglu, S., Lim, H., & Chen, W. H. (2020). ASR for EFL Pronunciation Practice: Segmental Development and Learners' Beliefs. *The Journal of AsiaTEFL*, *17*(3), 824–840. https://doi.org/10.18823/asiatefl.2020.17.3.5.824

Ishikawa, Y., Takamichi, S., Umemoto, T., Aikawa, M., Sakamoto, K., Yui, K., Fujiwara, S., Suto, A., & Nishiyama, K. (2021). Japanese EFL Learners' Speaking Practice Utilizing Text-to-Speech Technology Within a Team-Based Flipped Learning Framework. *Lecture Notes in Computer Science*, 283–291. https://doi.org/10.1007/978-3-030-77889-7_19

Kang, M., Kashiwagi, H., Treviranus, J., & Kaburagi, M. (2008). Synthetic speech in foreign language learning: an evaluation by learners. *International Journal of Speech Technology*, *11*(2), 97–106. https://doi.org/10.1007/s10772-009-9039-3

Khademi, H. (2021). *Learning Pronunciation with Google Translate: Focus on English Past -ed* [MA thesis]. Concordia University.

Khalid, A., & Muhammad, A. I. (2014). Text to Speech Software for Improving Pronunciation of Graduation Level Students. *EUROPEAN ACADEMIC RESEARCH*, *2*(1), 870–887.

Kiliçkaya, F. (2011). *Improving Pronunciation via Accent Reduction and Text-to- Speech Software*. 102–113. https://doi.org/10.4324/9780203831762-15

Kim, I. (2006). Automatic Speech Recognition: Reliability and Pedagogical Implications for Teaching Pronunciation. *Educational Technology & Society*, *9*(1), 322–334.

Larsen-Freeman, D. (2000a). *Techniques and Principles in Language Teaching*. Oxford University Press.

Larsen-Freeman, D. (2000b). *Techniques and Principles in Language Teaching*. Oxford University.

Le, P., Vu, H., & Tran, M. (2022). Improving EFL Students' Intonation In-Text Using Shadowing Technique with the Implementation of Google Text-to-Speech. *AsiaCALL Online Journal*, *13*(1), 93–121.

Lee, S. T. (2008). *Teaching Pronunciation of English Using Computer Assisted Learning Software: An Action Research Study in an Institute of Technology in Taiwan* [Phd dissertation]. Australian Catholic University.

Levis, I., & Suvorov, R. (2013). Automatic speech recognition. In C. A. Chapell & C. Chapell (Eds.), *The encyclopedia of applied linguistics*. New York: Wiley-Blackwell.

Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, *32*(1), 1–25. https://doi.org/10.1558/cj.v32i1.25962

Liakin, D., Cardoso, W., & Liakina, N. (2017). The pedagogical use of mobile speech synthesis (TTS): focus on French liaison. *Computer Assisted Language Learning*, *30*(3–4), 325–342. https://doi.org/10.1080/09588221.2017.1312463

Linebaugh, G. (2015, March 6). *Evidence that L2 production training can enhance perception*. https://journal.aall.org.au/index.php/jall/article/view/326

Lluïsa, G. V. M. (2015, November 24). *Teaching Pronunciation through ICT Tools: A Case Study*. https://repositori.uji.es/xmlui/handle/10234/143535

Most, T., & Peled, M. (2007). Perception of Suprasegmental Features of Speech by Children With Cochlear Implants and Children With Hearing Aids. *Journal of Deaf Studies and Deaf Education*, *12*(3), 350–361. https://doi.org/10.1093/deafed/enm012

Moussalli, S., & Cardoso, W. (2021). Intelligent personal assistants and L2 pronunciation development: focus on English past -ed. In *CALL and professionalisation: short papers from EUROCALL 2021* (pp. 226–231). https://doi.org/10.14705/rpnet.2021.54.1337

Mroz, A. (2018). Seeing how people hear you: French learners experiencing intelligibility through automatic speech recognition. *Foreign Language Annals*, *51*(3), 617–637. https://doi.org/10.1111/flan.12348

Nusbaum, H. C., Francis, A. L., & Henly, A. S. (1995). Measuring the naturalness of synthetic speech. *International Journal of Speech Technology*, *1*(1), 7–19. https://doi.org/10.1007/bf02277176

Proctor, M. (2021). Consonants. In *Cambridge University Press eBooks* (pp. 65–105). https://doi.org/10.1017/9781108644198.004

Saito, K., & Lyster, R. (2011). Effects of Form-Focused Instruction and Corrective Feedback on L2 Pronunciation Development of /ɹ/ by Japanese Learners of English. *Language Learning*, *62*(2), 595–633. https://doi.org/10.1111/j.1467-9922.2011.00639.x

Schoenberg, I. (2016). *Focus on Grammar 2*. Pearson Education ESL.

Scott, A. (2022). A Guide To How Text-To-Speech Works. *Data Science Central*. https://www.datasciencecentral.com/a-guide-to-how-text-to-speech-works/

Soler-Urzúa, F. (2011). *The acquisition of English /ɪ/ by Spanish speakers via text-to-speech synthesizers: A quasi-experimental study* [MA thesis]. Concordia University.

Stevens, C., Lees, N., Vonwiller, J., & Burnham, D. (2005). On-line experimental methods to evaluate text-to-speech (TTS) synthesis: effects of voice gender and signal quality on intelligibility, naturalness and preference. *Computer Speech &Amp; Language*, *19*(2), 129–146. https://doi.org/10.1016/j.csl.2004.03.003

Svalberg, A. M. (2012). Teaching for Language Awareness. In *The Encyclopedia of Applied Linguistics*. https://doi.org/10.1002/9781405198431.wbeal1157

Swain, M. (1999). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantof (Ed.), *Sociocultural Theory and Second Language Learning* (pp. 97–114). Oxford University Press.

Van Duong, T. (2022). The Effects of using Online Text-To Speech Tools on EFL Students' Perceptions in Learning Pronunciation. *International Journal of Science and Management Studies*, 270–276. https://doi.org/10.51386/25815946/ijsms-v5i4p129

Van Lieshout, C., & Cardoso, W. (2022). Google Translate as a tool for self-directed language learning. *Language Learning & Technology*, *26*(1), 1–19. http://hdl.handle.net/10125/73460

VanPatten, B., Smith, M., & Benati, A. (2020). Key Questions in Second Language Acquisition: An Introduction. *Instructed Second Language Acquisition*, *4*(2).

Wahyuni, N., Dj, M. Z., & Nur, H. (2021). Identification the students' pronunciation problems in pronouncing –ed ending at English Study Program of IAIN Bone. *International Journal of Research on English Teaching and Applied Linguistics*, *1*(2), 83–91. https://doi.org/10.30863/ijretal.v1i2.1231

Wang, X. Z. (2020). Segmental versus Suprasegmental: Which One is More Important to Teach? *RELC Journal*. https://doi.org/10.1177/0033688220925926

Williamson, G. (2018). Vowel Parameters. *SLT Info*. https://www.sltinfo.com/ess101-vowel-parameters/

Yenkimaleki, M., & Van Heuven, V. J. (2021). Effects of attention to segmental vs. suprasegmental features on the speech intelligibility and comprehensibility of the EFL learners targeting the perception or production-focused practice. *System*, *100*, 102557. https://doi.org/10.1016/j.system.2021.102557

Zhu, S.-G. (2011). Speech synthesize note. Retrieved from http://irw.ncut.edu.tw/peterju/speech.html#categories

APPENDICES

APPENDIX 1

INFORMATION AND CONSENT TO PARTICIPATE IN A RESEARCH STUDY

Study Title: Assessing the Effectiveness of Text-to-Speech and Automatic-Speech-Recognition Tools in Improving EFL Learners' Pronunciation of Final "Ed"

the Case of A1 Learners' Level at CEIL Center

Researcher: Kouidri Halima Saadia (Master's student in Applied Linguistics)

Researcher's Contact Information: kouidrihalima20@gmail.com

Faculty Supervisors: Dr.Ilham Tigane Applied Linguistics

Source of funding for the study: None

You are invited to participate in the research study mentioned above. This form provides information about what participating would mean. Please read it carefully before deciding if you want to participate or not. If there is anything you do not understand, or if you want more information, please ask the researcher.

A. PURPOSE

The purpose of the research is to investigate the pedagogical role of Text-to-Speech synthesizers and Automatic-Speech-Recognition in improving your pronunciation in English, and your perceptions of the technology and learning experience.

B. PROCEDURES

If you participate, you will be asked to complete the following tasks in class and remotely (times are approximate):

• Fill out the demographic survey (about 3 minutes)

• Take a set of pre-tests on English pronunciation (about 25 minutes).
• Watch a short video on how to use the app to practice pronunciation (about 5 minutes).

• Practice your pronunciation using the TTS and ASR features in the app (about 15 minutes in your official session) and the rest as homework for one week

• Take a set of online post-tests on English pronunciation (about 25 minutes).

•Answer a questionnaire (about 15-20 minutes)

As mentioned earlier most aspects of the study will be conducted in class (pretest and posttest) and only the treatment phase will be in class and remotely and recorded for further data analysis. Neither your voice nor your name will appear on the recording or will be published.

C. RISKS AND BENEFITS

There are no known risks associated with participating in this research. One benefit could be that you will enjoy practising English using applications. As well, the study will introduce you to a potentially beneficial technology that can help you improve your pronunciation in English.

D. CONFIDENTIALITY

We will gather the following information as part of this research:

- Through a questionnaire, we will collect data on your learning strategies and perceptions of using Neural Reader as a learning tool. We will also record and gather data about your level in the pretest and posttest .so that it can be analyzed.

- We will not allow anyone to access your information, except people directly involved in the research. We will only use the information for the purposes of the research described in this form.

- The information gathered will be coded and your name will be removed and replaced by a pseudonym which means your name will not appear anywhere in the written study. No one else will have access to your information.

- We intend to publish the results of the research. However, it will not be possible to identify you in the published results.

F. CONDITIONS OF PARTICIPATION

- You do not have to participate in this research. It is purely your decision. If you do participate, you can stop at any time. You can also ask that the information you provided not be used, and your

choice will be respected. If you decide that you don't want us to use your information, you must tell the researcher before April 30, 2023.

- There are no negative consequences for not participating, stopping in the middle, or asking us not to use your information.

G. PARTICIPANT'S DECLARATION

I have read and understood this form. I have had the chance to ask questions and any questions have been answered. I agree to participate in this research under the conditions described.

NAME _____

SIGNATURE _____

DATE _____

If you have questions about the scientific or scholarly aspects of this research, please contact the researcher. Their contact information is on page 1.

If you have concerns about ethical issues in this research, please contact the Manager of Research.

APPENDIX 2

Demographic Questionnaire

Please fill out the following form.

Name : .....

Email: .....

Age : .....

Gender: Female ( ) Male ( )

Questions :

1. “I would really like to have better pronunciation.”

Strongly Agree( ) Agree ( ) neutral ( ) Disagree ( ) Strongly disagree ( )

2. Do you believe that pronunciation can be improved outside of the classroom?

.....

3. On a scale from 1 (No knowledge) to 9 (Very advanced knowledge), rate your knowledge level when using technology in general:

(No knowledge) 1 2 3 4 5 6 7 8 9 (Very advanced knowledge)

4. How often do you use text-to-speech feature to improve your English pronunciation?

Always ( ) Often ( ) Sometimes ( ) rarely ( ) Never ( )

5. How often do you use automatic-speech-recognition feature of GT to improve your English pronunciation?

Always ( ) Often ( ) Sometimes ( ) rarely ( ) Never ( )

6. Do you use any websites or applications to practice pronunciation? Do you use any? If yes, which ones?

.....

7. Have you ever struggled with pronouncing -ed endings in English, and if so, what strategies have you used to improve your pronunciation?

.....

8. Do you know how past tense -ed is pronounced? Explain.

.....

APPENDIX 3

Awareness Pretest

1. **what are the most common ways that -ed endings are pronounced in English, and when is each pronunciation used?**

……

Instruction : Select the verbs that have similar ending pronunciations.

Verbe	**Asked**	**Used**	**Added**
● Dreamed			
● Risked			
● Amazed			
● Trusted			
● Belonged			
● Painted			
● Used			
● Demanded			
● Kicked			
● Forced			
● Changed			
● Landed			
● Discovered			
● Included			
● Mixed			
● Fished			

• Judged			
• Rested			

Pretest Perception (Aural discrimination) and Production:

https://form.jotform.com/231133537517048

APPENDIX 4

Awareness Posttest

1. Do you know how past tense -ed is pronounced? Explain.

……

……

Instruction : Select the verbs that have similar ending pronunciations.

Verbe	**Asked**	**Used**	**Added**
• Laughed			
• Voted			
• Minded			
• Arrived			
• Hepled			
• Damaged			
• Used			
• Sounded			
• Danced			
• Forced			
• Asked			
• Adviced			
• Discovered			
• Invited			

• Burned			
• Interrupted			
• Fined			
• Realized			

Posttest Perception (Aural discrimination) and Production:

https://form.jotform.com/231134268994564

APPENDIX 5

Treatment Materials

1- Download the APP (Neural Reader Humanlike TTS).
2- Watch the video on how to set app the app and how to use it.

Activity one :

1. Copy and paste the story into the app, and use the Text-to-Speech feature to listen to it carefully. Pay close attention to the pronunciation of the verbs.

2. Use the Automatic Speech Recognition tool to read the story out loud, and pay particular attention to the final "ed" sound in past tense verbs.

Lily's Day at the Park

A seven-year-old girl named Lily went to the park with her father. They played catch with a ball and then sat on a bench. Lily watched the people around her. She saw a man walking his dog and another man riding his bicycle. She was fascinated by the different activities. Suddenly, she saw a group of children playing a game she didn't know. She stood up and walked toward them. The children were playing tag. Lily wanted to join but she was too shy to ask. Finally, a little girl came up to her and asked, “Do you want to play with us?” Lily nodded her head and smiled. The girl explained the rules of the game to her. Lily was excited to play. She ran and dodged the other children, trying not to get tagged. After a while, Lily got tired and sat down on the grass. She watched the children playing and felt happy to have made new friends. Suddenly, she saw her father walking toward her. “Time to go, Lily,” he said. Lily got up and waved goodbye to her new friends. She felt happy to have had such a fun day at the park with her father.

A Day in the Park with Tom

It was a beautiful day in the park. People walked their dogs, kids played on the swings, and the sun shone brightly overhead. Tom enjoyed walking through the park. He noticed a group of children

laughing and chirping near the pond. Tom reached the pond and saw a duck paddling in the water. He walked closer and noticed a fish darting through the water. The duck reached its beak into the water and plucked the fish out. Tom felt a pang of sadness for the fish. He walked to a nearby bench and sat down. He remembered a time when he caught a fish with his grandpa. They packed their fishing gear and drove to a nearby lake. Tom put the bait on the hook and waited patiently. After a while, Tom noticed the bobber moving. He took hold of the rod and reeled the line in slowly. Tom laughed as he saw the fish at the end of the line. He took the fish off the hook and put it in a bucket. Tom was happy and proud to catch his first fish.

Activity two:

1. Copy and paste the verbs from the list in the app.Use the app's Text-to-Speech feature to listen to the pronunciation of the verbs.

2. Read the verbs aloud from a piece of paper using the Automatic Speech Recognition feature (name may vary depending on the app).Pay attention to the pronunciation of the verbs to improve your spoken English skills.

walked - talked - watched - kicked - picked – worked - helped - passed - washed - brushed jumped - laughed - looked – loved - played - climbed - typed - zipped - stopped - added - carried - closed - decided - divided - ended - folded - guided - included - landed - melted - nodded - opened - rained - saved - tasted - waited - wanted - needed - faded - hated - painted - rented - repeated - started - united - visited - voted - rewarded – sounded

Activity Three: Practice Materials (Production #1)

Using the dictation feature (Automatic Speech Recognition) in the app, read the following verbs:

face, faced	**dance, danced**	**rain, rained**
start, started	**drip, dripped**	**decorate, decorate**
attend, attended	**burn, burned**	**divide, divide**
plant, planted	**drop, dropped**	**repeat, repeated**
color, colored	**cheat, cheated**	**fix, fixed**
manage, managed	**escape, escaped**	**connect, connected**
develop, developed	**worry, worried**	**advise, advise**
cross, crossed	**suggest, suggested**	**argue, argued**
label, labeled	**dislike, disliked**	**force, force**
learn, learned	**test, tested**	**answer, answered**
paint, painted	**count, counted**	**end, ended**

walk, walked	**love, love**	**help, helped**
clean, cleaned	**decide, decide**	**visit, visited**
drag, dragged	**wait, waited**	**watch, watched**
use, used	**sniff, sniffed**	**shave, shaved**
stuff, stuffed	**want, wanted arrest,**	**complete, complete**
close, close	**arrested**	**work, worked**
welcome, welcome	**agree, agree**	**allow, allowed**
add, added	**thank, thanked**	**search, searched**
wash, washed	**type, type**	**avoid, avoided**
accept, accepted	**analyze, analyze**	**shop, shopped**
kill, killed	**correct, corrected**	**arrange, arrange**

Activity four:

Read the following sentence using Dictation feature in the app (Automatic speech recognition)

The classes **started** at 5 o'clock. The classes start at 5 o'clock.

The dogs **jumped** over the fence. The dogs jump over the fence.

Mary and Kate **described** the man to the Police. Mary and Kate describe the man to the Police.

They **posted** a video on Youtube. They post a video on Youtube.

You **ordered** a bag from Amazon. You order a bag from Amazon.

I **decided** to learn French. I decide to learn French.

They **walked** in the park. They walk in the park.

I **closed** my store. I close my store.

John and Mathew **cleaned** everywhere. John and Mathew clean everywhere We

wished her a happy birthday. We wish her a happy birthday.

Peter and his friend **joined** the team. Peter and his friend join the team.

They **waited** for 2 hours. They wait for 2 hours.

I **asked** my questions in the class. I ask my question in the class.

You **fixed** your car. You fix your car.

They **offered** me a good job. They offer me a good job. They

repeated their questions. They repeat their question. We **loved** the

book. We love the book.

They **arrested** the robber. They arrest the robber.

You **brushed** your teeth. You brush your teeth.

We **planted** a tree in the yard. We plant a tree in the yard. They

played soccer. They play soccer.

We **danced** with the music. We dance with the music.

Carol and John **corrected** my mistakes. Carol and John correct my mistakes.

We **stayed** in his house. We stay in his house.

I **called** you at 4 O'clock. I call you at 4 O'clock.

They **painted** the house. They paint the house.

You **knocked** on the door. You knock on the door.
They **attracted** a lot of tourists. They attract a lot of tourists. I **helped** him with his English. I

help him with his English. We **baked** a cake. We bake a cake.

APPENDIX 6

Students' Questionnaire

Dear student,

We would like to invite you to take part in research entitled **" Assessing The Effectiveness of Text-To-Speech (TTS) and Automatic Speech Recognition (ASR) in improving EFL learners' Pronunciation"**

The purpose of this questionnaire is to collect your perceptions and attitudes towards the use of TTS and ASR tools in improving your pronunciation of plosive consonants. In This questionnaire you will be asked about your experience using these tools, their effectiveness in helping you improve your pronunciation, the ease of use, and any other relevant feedback you may have. Your participation in this research is entirely voluntary, and all responses will be kept confidential.

*We would like to **thank you** for your participation in advance, and your feedback will be greatly appreciated.*

The researcher's email: kouidrihalima20@gmail.com

Supervisor's email: ilham.tigane@univ-biskra.dz

Key terms:

TTS (Text-To-Speech): refers to technology that converts written text into spoken words. It involves a computerized system that reads text aloud in a natural-sounding voice (read aloud feature in the Neural Reader application)

ASR (Automatic Speech Recognition): refers to technology that uses machine learning algorithms to convert spoken words into text. It involves a computerized system that analyzes audio recordings of speech and transcribes them into written text. (Dictation feature in the Neural Reader application)

Autonomous learner: An autonomous learner is an individual who is responsible for his/her education. Autonomous means are self-directed or capable of working independently.

Instruction: Please tick (√) the appropriate answer choice for the question above.

General Information:

1. What is your age?

2. What is your gender? ☐ Male ☐ Female

Questionnaire:

Part One: Perceptions of Text-To-Speech (TTS).

1. The input provided by TTS technology was clear and natural.

☐ Strongly agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly disagree

2. The TTS tool was easy to use.

☐ Strongly agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly disagree

3. The TTS tool assisted me in accurately identifying and distinguishing between different variations of the final "-ed" sound.

☐ Strongly agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly disagree

4. Would you recommend the TTS tool to other English language learners for improving their pronunciation skills?

☐ Yes ☐ No

Part Two: Perceptions of Automatic Speech Recognition (ASR).

5. The Output provided by ASR technology was accurate.

☐ Strongly agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly disagree

6. ASR is a helpful tool for practising pronunciation.

☐ Strongly agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly disagree

7. Do you think using ASR for pronunciation practice is more effective than traditional methods (e.g., repeating after a teacher)?

☐ Yes ☐ No

8. Was it easy to use ASR to practice your pronunciation?

☐ Yes ☐ No

9. Would you recommend ASR as a pronunciation learning tool to other learners? ☐ Yes ☐ No

Justify your answer

……

……

……

Part Three: Perceptions of Text-To-Speech (TTS) and Automatic Speech Recognition (ASR). Instruction: How much do you agree or disagree with the following statements

N	Statements	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
10	TTS and ASR allowed me to evaluate my own pronunciation (for example, to decide whether my pronunciation was correct or incorrect).					
11	I felt more comfortable practicing pronunciation with TTS and ASR than I would in front of other students.					

12	I felt more comfortable practicing pronunciation with TTS and ASR than I would in front of the teacher.					
13	TTS and ASR enhanced my engagement beyond the classroom(facilitating learning and practice pronunciation at any time and any location) .					
14	I feel confident in my ability to perceive and produce the correct pronunciation of final 'ed' after using the TTS and ASR tools					
15	TTS and ASR tools helped me become a more autonomous learner.					

16. Did you find it helpful to use both TTS and ASR together to improve your perception and production of the final 'ed' sound in English verbs.?

☐ Yes ☐ No

Explain:……
……

17. Would you recommend leaning pronunciation with TTS and ASR to other learners?

☐ Yes ☐ No

Why?……
…… …….

18-Do you think you should spend more time using Neural reader app to improve your pronunciation outside the class?

☐ Yes ☐ No

Justify

……

……

19. Do you think Neural reader application (ASR and TTS) matches your learning style, considering some students are visual learners while others prefer auditory learning? how?

……

…….

20. Which features did you utilize while using the app? (You can tick more than one option)

a.☐ Change the playback speed.

b. ☐opt for a different voice.

c. ☐Select an alternate accent.

☐ Others, please specify ……

Part 4: Suggestions and Comments

21. Based on your experience, in what ways can Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) be helpful in improving pronunciation?

……

……

……

……

22. In case you have further suggestions, comments or recommendations, you are most welcome to add them below.

……

……

THANK YOU

Appendix 7:

Students' scores

	Awareness	Perception	Production	Pretest sco	Awareness	Perception	Production	Postest
Control 1	11	9	7	27	13	11	15	39
Control2	6	4	12	22	6	10	16	32
Control3	5	8	8	21	8	10	10	28
Control4	10	9	9	28	12	11	14	37
Control 5	7	12	16	35	11	12	15	38
Control6	11	8	6	25	14	9	11	34
Control 7	7	8	13	28	10	8	14	32
Control 8	8	5	8	21	12	5	11	28
Control 9	7	3	12	22	12	7	13	32
Control 10	7	8	8	23	11	9	13	33
Control11	6	5	10	21	9	6	14	29
Control 12	4	13	5	22	7	13	7	27
Control3	6	9	7	22	7	11	8	26
Control 14	6	5	7	18	7	10	11	28
Exp 1	10	6	11	27	15	11	18	44
EXP 2	10	5	11	26	14	14	17	45
EXP 3	7	8	8	23	12	18	15	45
EXP 4	8	9	3	20	11	14	14	39
EXP 5	7	6	9	22	15	16	14	45
EXP 6	5	9	10	24	10	15	15	40
EXP 7	11	5	6	22	14	17	17	48
EXP 8	6	13	13	32	13	13	15	41
EXP 9	8	3	11	22	14	18	16	48
EXP 10	6	3	7	16	12	16	16	44
EXP 11	5	10	10	25	9	17	14	40
EXP 12	6	3	11	20	13	8	15	36
EXP 13	4	10	6	20	9	15	16	40
EXP 14	5	12	11	28	9	14	10	33

Le résumé

Au cours des dernières années, un intérêt croissant s'est porté sur les outils d'entraînement à la prononciation assistée par ordinateur (CAPT), et l'application de la synthèse vocale et de la reconnaissance automatique de la parole dans les applications d'apprentissage des langues a été remarquable. Cette étude quasi-expérimentale visait à évaluer l'efficacité de la synthèse vocale (TTS) et de la reconnaissance automatique de la parole (ASR) pour améliorer la prononciation du passé en anglais (-ed) chez les apprenants de langues étrangères de niveau élémentaire. L'enquête visait à renforcer la conscience, la perception et la production des apprenants en utilisant une conception de groupes prétest-posttest avec un groupe expérimental et un groupe témoin. De plus, les perceptions et attitudes des participants à l'égard de ces outils ont été explorées à l'aide d'un questionnaire. Les résultats de l'analyse t-test ont révélé des améliorations significatives dans les scores du post-test du groupe expérimental par rapport au groupe témoin. La section de discussion ultérieure a souligné l'impact positif des technologies TTS et ASR sur la conscience, la perception et les compétences de production des apprenants. Les résultats suggéraient que les technologies TTS et ASR ont le potentiel de faciliter les différentes étapes du développement de la prononciation, aidant ainsi les apprenants à acquérir les caractéristiques de prononciation souhaitées. De plus, les résultats du questionnaire ont révélé des perceptions positives et une satisfaction élevée à l'égard des technologies TTS et ASR, soulignant leur efficacité en tant que ressources précieuses pour la pratique de la prononciation et l'auto-évaluation. L'étude a conclu en mettant l'accent sur les implications pédagogiques de l'intégration des technologies TTS et ASR dans l'enseignement de la prononciation afin d'améliorer les compétences de prononciation des apprenants et d'étendre l'environnement d'apprentissage au-delà des limites de la salle de classe traditionnelle.