People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

Mohamed Khider University of Biskra

Faculty of Letters and Languages

Department of English Language and Literature

Thesis Presented to the Department of English Language and Literature in Candidacy for the LMD Doctorate Degree in Applied Linguistics

The Effect of Automatic Speech Recognition Technologies on Students' Pronunciation

A Case Study of First-year EFL Learners at Biskra University

Submitted by: *Supervised by:*

Miss Roumaissa SOUICI   Pr. Saliha CHELLI

The Board of Examiners

Chair: Pr. Ahmed Chaouki HOADJLI	Biskra University
Supervisor: Pr. Saliha CHELLI	Biskra University
Examiner: Pr. Sabrina BAGHZOU	Khenchela University
Examiner: Pr. Leila DJAAFRI	Batna 2 University
Examiner: Dr. Mostapha MEDDOUR	Biskra University
Examiner: Dr. Hanane SAIHI	Biskra University

2023 – 2024

Dedication

To those who believed in me, even when I did not believe in myself,

I dedicate this work to my dearest mother, **Fairouz**, I'm forever grateful for everything you have done for me,

I dedicate this work to my beloved father, **Okba**, who pushed me to strive for excellence, encouraged me to follow my dreams, helped and supported me every step of the way,

To my beautiful sisters, **Selssabil** and **Mouna**, who were my inspiration and the reason I aspire to be someone they could look up to,

I hope that this accomplishment brings you joy and pride.

Acknowledgments

My teacher and supervisor, **Pr. Saliha CHELLI** made this humble work possible. I owe her a great debt of gratitude for her guidance, advice, and encouragement. I also want to thank her for spending much of her valuable time proofreading and correcting my mistakes, adding her expertise and knowledge to this work, and helping me become better at what I do.

I would like to express my sincere gratitude and appreciation to the members of the jury, **Pr. Ahmed Chaouki HOADJLI, Pr. Sabrina BAGHZOU, Pr. Leila DJAAFRI, Dr. Mustapha MEDDOUR,** and **Dr. Hanane SAIHI** who have dedicated their time, expertise, and valuable insights to evaluate and assess my work.

A debt of gratitude is also owed to **Dr. Ben Abdelrrezak ABDELNACER** for his guidance, help, and support. Without him, this work would have not been possible.

A special thanks to **Dr. Moustafa AMRATE,** for his comments that greatly helped me improve my work, his assistance, and help to me and all the other candidates.

I would like to also acknowledge all the teachers at the Department of English Language and Literature at Mohamed Khider Biskra University, my colleagues, my friends, and everyone who was part of this journey.

I am also grateful to the students who took part in this study, who were helpful, motivated, kind, and understanding, and without whom this work would not have been possible; I owe this work to each one of you.

Declaration

I, **Roumaissa SOUICI**, hereby declare and affirm that the work presented in this thesis entitled

"The Effect of Automatic Speech Recognition Technologies on Students 'Pronunciation" is

entirely my own, conducted under the supervision and guidance of my supervisor **Pr. Saliha**

Chelli. I have acknowledged all the sources of information and ideas that have been used in this

work. I have appropriately cited and referenced all the sources, including direct quotations,

paraphrases, and ideas borrowed from other authors or publications. Further, this thesis is

submitted only to Mohamed Khider University of Biskra (Algeria) as a partial requirement for the

degree of LMD Doctorate in Applied Linguistics.

Souici Roumaissa

Biskra

April 2023

Abstract

Effective English pronunciation instruction requires exposure to authentic input, practice, and feedback. Nowadays, advanced technologies can be employed to teach and assess pronunciation. Therefore, this mixed-method research aimed to investigate the effect of Automatic Speech Recognition (ASR) technology on EFL students' pronunciation at the Department of English and Literature at Biskra University. For this purpose, a semi-structured questionnaire was administered to a sample of four teachers of oral expression and phonetics in the initial phase, in addition to using a quasi-experimental design involving a control and an experimental group with twelve participants each (n=12). The intervention was followed by a semi-structured interview conducted with the participants of the experimental group. The four teachers agreed that incorporating technology into the classroom is important, and they emphasized that adequate planning and training are essential before introducing a technological tool to the students. Besides, the results of the intervention revealed the significance of the use of ASR to improve students' pronunciation. Further, the positive responses to the interview supplemented the results and also showed that the participants enjoyed using ASR technology. The research findings highlighted the potential benefits of integrating ASR technology into pronunciation instruction to facilitate the learning process and lead to better learning outcomes.

Keywords: Pronunciation; Automatic Speech Recognition; Technology; English as a Foreign Language (EFL) Learners

List of Figures

List of Diagrams

List of Pictures

List of Graphs

List of Tables

List of Abbreviations

4G LTE	Fourth Generation Long-Term Evolution
AI	Artificial Intelligence
API	Application Programming Interface
ASR	Automatic Speech Recognition
CALL	Computer-Assisted Language Learning
CAPT	Computer-Assisted Pronunciation Training
EFL	English as a Foreign Language
ESL	English as a Second Language
GVC	Google Voice Search
HMM	Hidden Markov Model
ICT	Information and Communication Technology
OS	Operating System
PC	Personal Computer
SPSS	Statistical Package for the Social Sciences
TTS	Text-to-Speech
WSR	Windows Speech Recognition

Table of Contents

Appendix B: Participants 'Information Card

Appendix C: Consent Form

Appendix D: Speaking Test Answers

Appendix E: Control Group Scores

Appendix F: Experimental Group Scores

Appendix G: A Sample of Participants 'Recordings

Appendix H: Screenshots of WhatsApp Messages

Appendix I: Post-Training Interview

Résumé

الملخص

General Introduction

General Introduction

General Introduction

Introduction

The widespread use of English in a variety of fields, including business, education, technology, and entertainment, has increased the need for non-native speakers to become proficient in the language. While mastering grammar, vocabulary, and pronunciation are critical components of learning English, it's vital to bear in mind that language learning is a multifaceted process that includes developing skills in listening, speaking, reading, and writing. Furthermore, language learning involves not only acquiring linguistic knowledge but also comprehending cultural nuances and communicative strategies.

Although pronunciation is a pervasive aspect of successful communication and a crucial feature of spoken language, Kelly (1969) argued that Pronunciation is considered to be the "Cinderella Area" of foreign language teaching. Pronunciation has been a neglected aspect for decades and it was pointed out by many scholars and researchers (Lin, Fan, & Chen, 1995; Wei & Zhou, 2002; Dalton, 2002; Gilbert, 2008).

Pronunciation helps learners attract the listener's attention and its importance relies on the knowledge being transmitted and its delivery. The speaker needs a practical mastery of the language's sounds, rhythms, and cadences and needs to learn how to fit them together in connected speech. Traditionally, the main aim of pronunciation teaching was to enable EFL learners to achieve native-like pronunciation. EFL learners intended to attain a native-like pronunciation; however, as English has become the most spoken language in the world, the focus of teaching pronunciation has shifted to only achieving a comprehensible pronunciation. Therefore,

intelligibility is considered to be the primary goal when it comes to learning a foreign language rather than a native-like pronunciation. EFL and ESL learners aim to be comfortably intelligible.

The momentum to teach pronunciation has led to looking for new technological techniques and resources to fulfill teachers' and learners' desire to achieve intelligibility. Richey (2008) stated that technology changes the role of language teachers from the traditional way of transmitting knowledge to having a new part of a facilitator and delegator inside the classroom. For instance, this study investigated the effectiveness of Automatic Speech Recognition (ASR) technologies in enhancing EFL learners' pronunciation.

1. Statement of the Problem

Pronunciation is considered to be one of the most challenging skills to teach and learn. Despite the challenges, effective pronunciation instruction can help learners improve their communicative competence and confidence in using English. Guilbert (2008) inserted that learning a foreign language like English differs from learning the first language. EFL learners may feel uneasy while learning English since their first language may affect their pronunciation. This is because there are different sounds and rhythms in English. Moreover, learners may fear sounding "foreign," which may cause them to feel anxious and embarrassed.

EFL learners at Biskra University are introduced to the Phonetics module in their first year. In this module, they learn about the basic features of phonology and the different sounds of English. Although technology has recently witnessed many advances, the use of technological materials or tools to teach pronunciation is rare because the university lacks materials and equipment. Furthermore, exposing learners to authentic language input inside the language classroom is

considered one of the essential ingredients for successful language learning. A meaningful input is more relevant to the learners' needs and can stimulate motivation. Practicing the language is also a vital aspect of improving pronunciation since using English to communicate with others is the primary goal of learning it in the first place.

The world around us now is controlled by technology; therefore, this study demonstrated the importance of including technology inside the EFL classroom. Nowadays, advances in machine learning have led to significant improvements in ASR technology making it reliable and effective. Automatic Speech Recognition (ASR) offers many advantages to human life and enables people to perform many tasks on their phones or computers. Hence, it may be included inside the EFL classroom as well and may help address individual problems and allow learners to train individually at any time and place. This technology provides learners with authentic language input, instant corrective feedback, and flexibility to practice anytime and anywhere. Further, this technology allows learners to enhance their self-confidence and minimizes classroom anxiety.

2. Research Aims

This study investigated the effect of ASR technologies on EFL learners' pronunciation. The study's objective was to shed light on the potential of using ASR-based technology to promote learners' pronunciation. It also sought to highlight its practicality and relevance in language teaching.

Furthermore, it tried to offer insights into how ASR technology can be used as an additional teaching instrument to make the learning process more efficient. Finally, this study proposed a set of recommendations and suggestions on the appropriate use of ASR-based technology and its impact on EFL learners.

3. Research Questions

The current study aimed to answer the following questions:

a. Main Research Question

Does the use of Automatic Speech Recognition Technologies improve EFL learners' pronunciation?

b. Subsidiary Questions

- What are teachers' attitudes toward using technology to improve learners' pronunciation? And their perceptions of ASR technology?

- What are learners' attitudes toward using ASR technology to improve their pronunciation?

- What difficulties and challenges may learners encounter when using ASR technology to improve their pronunciation?

4. Research Hypotheses

Based on the previous research questions, the researcher hypothesized that:

o **Alternative Hypothesis (H1):** Using ASR technologies as a means to improve learners' English pronunciation affects learners' pronunciation.

o **Null Hypothesis (H0):** Using ASR technologies to teach English pronunciation does not affect learners' pronunciation.

5. Significance of the Study

Pronunciation is considered to be an essential aspect when learning a foreign language. Many scholars and researchers showed interest in how to teach it and how to help EFL learners improve their pronunciation (Elimat & AbuSeileek, 2014; Moedjito, 2017; Pardede, 2018; Chen, Inceoglu & Lim, 2020). Therefore, teachers tried many tools and methods to fulfill this important task. Many studies revealed that technology could be used to support pronunciation teaching when it is applied correctly and appropriately inside the language classroom. In addition, EFL learners are more than ever exposed to technology everywhere and anytime.

Therefore, this study holds significant value as it attempted to demonstrate the effectiveness of allowing technology into the language classroom. It tested the usefulness of an ASR technology-based tool on EFL learners' pronunciation. Further, it tried to draw teachers' attention to the importance of incorporating such technology as an additional teaching tool to gain effort and time. It also aimed to help EFL learners improve their pronunciation, fluency, and speaking skills.

6. Research Methodology

6.1. Method

Based on the research's purpose and nature, the mixed method was used including the quasi-experimental design to measure and test the effect of using ASR technology on EFL learners' pronunciation. Further, the research questions necessitate a multi-dimensional analysis that goes

beyond mere statistical results. Hence, a questionnaire and an interview were employed as additional qualitative data-gathering techniques.

This study was carried out following these stages:

- The first stage included a semi-structured questionnaire addressed to four teachers of Oral Expression and Phonetics modules at Mohamed Khider Biskra University.
- The second stage, the experimental phase, featured pre- and post-speaking tests and a training period using an ASR-based technological tool.
- The last phase was about interviewing the study sample.

6.2. Population and Sampling

The study included an experiment on EFL first-year learners at Biskra University. This population was chosen based on the fact that EFL learners are new to practicing their speaking skills. The investigation took place in the second semester of the academic year 2021 after allowing learners to have a background on the basics of English phonetics.

The experiment was carried on as additional instruction/training to enable the learners to practice their pronunciation and fluency. The study sample was divided randomly into two groups, a control group and an experimental group, each group consisted of 12 participants. The experimental group was exposed to training sessions through the ASR-based website "Speechace." The researcher was a facilitator and a guide during the training. The intervention was introduced to the sample and the researcher explained to the participants (i.e., the experimental group) how they can access the website and utilize its different features and options. The research also helped

the learners when they faced obstacles or had ambiguities. The other group, the non-ASR technology user, the control group, was not exposed to the technology or any separate training

6.3. Data Gathering Tools

Both quantitative and qualitative tools were employed to gather and analyze data. In the initial phase, four teachers of Phonetics and Oral Expression modules at the University of Biskra were requested to participate in a questionnaire that aimed to discover their perceptions and opinions on teaching and assessing pronunciation. Following the quasi-experimental method, the researcher conducted a pre-test and a post-test for both groups (i.e., the control and experimental groups) The purpose of the pre-test was to determine participants' pronunciation levels as well as their weaknesses. Then, the post-test was conducted to check whether any improvements were made by the experimental group in comparison to the control group. The tests were speaking tests that were executed on the ASR website.

The study sample took a test answering some questions about a particular topic, their answers were recorded and saved on the website, and then they were provided with a score on their pronunciation and fluency proficiency. During the experimental phase, the experimental group took part in training using the ASR website. After the intervention, the experimental group participants who participated in the training participated in a semi-structured interview to discover their perceptions and attitudes toward using the ASR website.

6.4. Analysis Procedures

First, teachers' replies to the questionnaire were collected and compiled for analysis. Content analysis was employed to look for recurring themes. It involved counting the number of

responses in each category to determine how prevalent each theme is among the teachers. This analysis helped to identify any patterns that may exist within the teachers' answers. Then, the findings were presented, including the themes identified, differences between teachers' answers, and their overall attitudes toward technology and pronunciation teaching. Second, the pre-test and post-test scores were taken from the website and compared.

Then, statistical analyses were conducted on SPSS Statistics to check if there was any significant difference in the test scores between the two groups (i.e., the experimental and control groups). Finally, the interviews conducted with the experimental group participants aimed to gather qualitative data on their perceptions, attitudes, and experiences with the ASR technology. Their responses were gathered and classified into different themes. Some questions aimed to explore participants' opinions and feelings towards the technology, and the specific aspects of pronunciation that the technology helped them improve. The analysis involved detecting underlying themes, emotions, and attitudes expressed by participants in their responses.

7. Thesis Structure

This thesis introduces a new approach to pronunciation instruction by proposing the integration of Automatic Speech Recognition ASR-based technologies as an additional teaching instrument. It sought to enhance the effectiveness of pronunciation learning for English as a Foreign Language (EFL) learners. This innovative approach holds promise for the future of pronunciation instruction at Biskra University, offering opportunities for more interactive, personalized, and technologically driven learning experiences. It provided a thorough analysis of historical practices, current trends, and the potential benefits of ASR technology. It also explored

how ASR-based technologies can be integrated, the challenges and opportunities associated with their implementation, and the impact on EFL learners' pronunciation proficiency.

The thesis consists of five chapters, the first chapter titled "**Pronunciation Beyond Listen and Repeat"** offers a comprehensive explanation of English phonology. It emphasizes the importance of accurate pronunciation in effective communication and language learning. Furthermore, this chapter explores the historical evolution of English pronunciation instruction, tracing its development over the decades. It investigates various methodologies and approaches used in teaching pronunciation, analyzing their effectiveness and identifying successful strategies as well as areas for improvement. In addition, it addresses the current landscape of pronunciation instruction, particularly in light of technological advancements. It explores the role of technology in teaching and assessing pronunciation, highlighting the use of innovative tools and methods for measurement and practice.

"**Automatic Speech Recognition Technologies**" is the title of the second chapter which provides a comprehensive overview of ASR technologies and their relevance to language learning and teaching. It explains the underlying principles and components of ASR systems, describes the process of speech recognition, and discusses the algorithms and machine learning techniques used in ASR. Moreover, it provides an overview of the history and evolution of ASR technology, and milestones in ASR development, from early speech recognition systems to modern applications. This chapter also provides an overview of some popular ASR systems and compares their features and applications to offer readers valuable insights into the diverse landscape of ASR technology and its relevance in language learning and teaching contexts. Further, it discusses the potential

benefits of using ASR technology in educational settings and provides strategies for integrating ASR into pronunciation instruction. It identifies potential challenges and limitations associated with ASR technology as well as strategies for addressing and overcoming these challenges in educational contexts.

The third chapter "**Research Methodology**" serves as a guide to the methodology employed in the study to investigate the effectiveness of an ASR-based technology in improving EFL learners' pronunciation. It explains how the methodology aligns with the research objectives, justifies the chosen research method based on the research question and objectives, describes the target population and the criteria used for sample selection, and provides an overview of the ASR technology used in the study, including its features, capabilities, and limitations. In addition, this chapter explains the procedures employed to analyze the collected data, ensures transparency in the research process, and helps readers understand how the study was conducted and why certain methodological choices were made.

The following and final chapter "**Analysis and Interpretation of the Results**" focuses on analyzing and discussing the results obtained from various data sources, including teachers' questionnaires, participants' online accounts, WhatsApp groups, and semi-structured interviews conducted with the experimental group participants. This chapter synthesizes the findings from different data sources to provide a comprehensive understanding of the impact of ASR technology on pronunciation instruction from both teachers' and students' perspectives. It combines quantitative analysis with qualitative insights to offer a nuanced interpretation of the research outcomes and address the research questions effectively. Further, this chapter provides a

comprehensive reflection on the study's outcomes, offering insights into its significance, implications, limitations, and avenues for future research.

Chapter One

Pronunciation Beyond Listen and Repeat

Chapter One: Pronunciation Beyond Listen and Repeat

Chapter One: Pronunciation Beyond Listen and Repeat

Introduction

Pronunciation is a crucial aspect of successful communication. It is a crucial feature of spoken language since it is not merely required for speaking but also for communicating and making sense to others. Pronunciation ensures the clarity of the speech produced by the speaker as it also helps the hearer to interpret it appropriately. Moreover, it is an essential feature in second or foreign language teaching and learning. A growing body under the realm of research shows that pronunciation is an aspect of the language that has been neglected for decades and requires more attention in education. Therefore, teachers should aim to improve their learners 'phonological competence and other skills. Pronunciation is essential because it does not matter how good a learner's vocabulary or grammar is if they cannot be understood when they speak.

This first chapter portrays a general image of the different sounds of English. It involves a thorough explanation of the sound system of this language and its main features, and how to apply them appropriately. It also sheds light on the history of teaching English pronunciation throughout the previous decades while trying to answer questions about how it was taught and what worked and what did not. Further, it discusses the status of pronunciation instruction recently, especially with the development of technology through the use of different types and tools of measurement, along with the current practices that include technology.

1.1. The Sounds of English

Every language that has ever existed is unique in its sound system. Sounds help humans distinguish different meanings. Pronunciation is the initial layer of talk through which the speaker builds a set of utterances and sounds that are then decoded and interpreted by the listener. How we say, something is different from how we write it (Pennington and Rogerson-Revell, 2018). Hence, studying pronunciation and its other components encompasses two main fields: phonetics and phonology. Both of them describe and analyze the language from two different perspectives.

1.1.1. Phonetics

Phonetics and phonology are concerned with the ways humans produce and interpret speech. Speaking and listening to each other are viewed as normal human behavior, but these two behaviors encompass many unremarkably occurring mechanisms. Delahunty & Garvey (2010) define phonetics as "a system for describing and recording the sounds of language objectively" (p. 89). Phonetics studies the different speech sounds of the language.

Rogers (2000) explained that speech could be analyzed on various levels. It is a matter of anatomy and physiology; it studies organs and their functions. It is interpreted through speech sounds and the different units of speech. Besides, it can be analyzed by investigating the other properties of sound waves. Since speech is intended to be heard and perceived, it can be interpreted from the speaker's perspective and how he processes the sounds. Phonetics can be viewed as a group of phonetic sciences that is divided by Clark and Yallop (1995) into Articulatory Phonetics (Anatomy and Physiology of Speech), Acoustic Phonetics, and Auditory (Perceptual) Phonetics.

The first sub-field deals with identifying and classifying the different individual sounds of the language; it is interested in the other parts of the human anatomy that participate in speech production and helps demonstrate how these speech sounds are articulated and produced. The second concern is the instrumental analysis and measurement of the physical properties of the sound waves as transmitted between the mouth and the ear, as it also measures the frequency and pitch of the vibrations of the vocal cords. The third one studies how sounds are heard, perceived, or mediated by the ear and the auditory nerve or brain. The sub-fields above demonstrate that speech is a purposeful, organized, systematic human activity, not just wasted noise.

1.1.2. Phonology

Phonology concerns the sounds within a particular language, how they are produced, and how they work together. As Kelly (2000) stated, "If phonetics deals with the physical reality of speech sounds, then phonology, on the other hand, is primarily concerned with how we interpret and systematize sounds" (p.9). Collins and Mees (2013) stated that it concerns a language's sound system and patterns. Many get these two concepts mixed because of their close definitions, but the difference relies on the word "system." Thus, phonetics deals with speech sounds but is not connected to a specific language. However, phonology is related to a particular language since each language has its system, which distinguishes one language from the other; it looks at the different features of the language and studies them in depth.

According to Crystal (1985), phonology deals with the functional classification of the sounds of a particular language, whereas phonetics is concerned with human sounds in general. Kreidler (2004) confirmed that phonology relates to how sounds are organized into a system to

enable communication. Phonetics is about how the speaker articulates sounds and affects the hearer. Therefore, it is crucial to be conscious of those differences even if it is agreed upon that phonetics and phonology are two fields that work together. Still, their approaches are divergent (Kelly, 2000).

1.2. English Pronunciation Features

Written and spoken English are two contrasting things. "Letters are not sounds" ; there is a tremendous difference between what people write and talk about; letters and sounds are not the same. Letters are written symbols that can be read and may represent one or many sounds, while sounds are noises and vibrations produced from movements of particular speech organs in the human body. Schmitt (2010) describes pronunciation as the study of how speech sounds are used to enable humans to communicate, whereas Ur (2012) commented that the term "pronunciation" includes all of the sounds of the language. Therefore, it refers to the sounds articulated by the speaker and understood by the listener.

Pronunciation is essential, especially when it comes to teaching a foreign or second language teaching and learning. Gut (2009) explained that teachers of English need to be mindful of the different English speech sounds and how they are produced and perceived. EFL\ESL learners need to learn about this to improve their pronunciation. Hence, the following section will tackle the essential features of English pronunciation.

Articulatory phonetics is about analyzing the different speech organs before understanding the various features of English pronunciation. First, this part below examines the different body parts that are involved in the production of speech. The question here is about how and where these phonemes are articulated.

Speech is produced by two devices, the neurological device in the brain and the mechanical device. The former includes the neurons participating in the language programmatic, while the latter consists of the different speech organs. Furthermore, the human vocal tract can produce an unlimited number of noises. Learners should be aware that some noises are not considered speech sounds, for instance, coughs, snores, and grunts, but they are still counted as phonological productions since they are articulated and produced by the vocal tract (Kelly, 2000; Gut, 2009).

1.3. Speech Production Process

A source of energy and different processes are needed to produce any sound. The sounds are made from muscles contrasting together to create songs. Below are the other body parts that take part in speech production.

1.3.1. The Lungs

These organs initiate speech production by providing air that is pushed upwards. When the lungs are expanded, air goes in and then out when compressed. Clark and Yalop (1995) describe this process as "the respiratory system counts as the energy source […] the lungs are compressed by various respiratory forces […] as the lungs are compressed, air flows out" (p. 11). Furthermore,

Gut (2009) stated, "The lungs are a pair of organs consisting of soft sponge-like material that rests on the diaphragm. They are connected to the windpipe, or trachea, by two bronchial tubes" (p.14). These organs' primary function is breathing, and its main function is to produce speech by providing an airstream that helps the production of different sounds.

Figure 1

The Respiratory System (Gut, 2009, p.14)
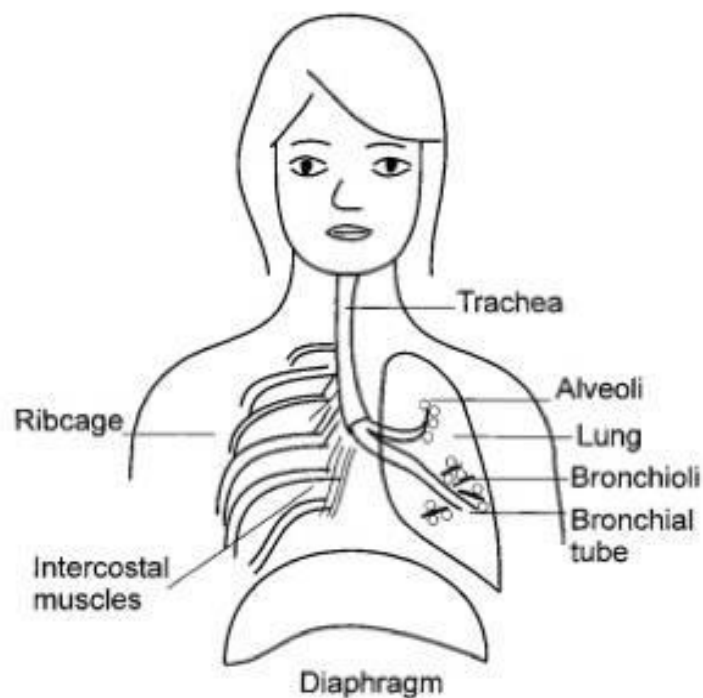
Figure 1 demonstrates the respiratory system. It displays the different organs that participate in speech production.

1.3.2. The Larynx

This organ is located behind Adam's apple, at the front of the throat. Its principal function is receiving the airstream to produce voice. Inside the larynx are the vocal cords, also called the

vocal folds, which are muscles covered by fibrous tissue. The opening between these two muscles is known as the glottis (Gut, 2009). These vocal cords can be adjusted to provide different acoustic features. Philip (2019) explained the changing positions of the vocal cords by stating that the vocal cords may be brought together and closed. This closure prevents air from escaping or entering the lungs, also known as a "glottal stop." However, if there is air pressure from the lungs, this massive muscular pressure will force the closure to open, and the air passes through the vocal cords.

Figure 2

The Vocal Cords (Philip, 2019, p. 36)

Vocal Cords Apart   **Vocal Cords Closed**
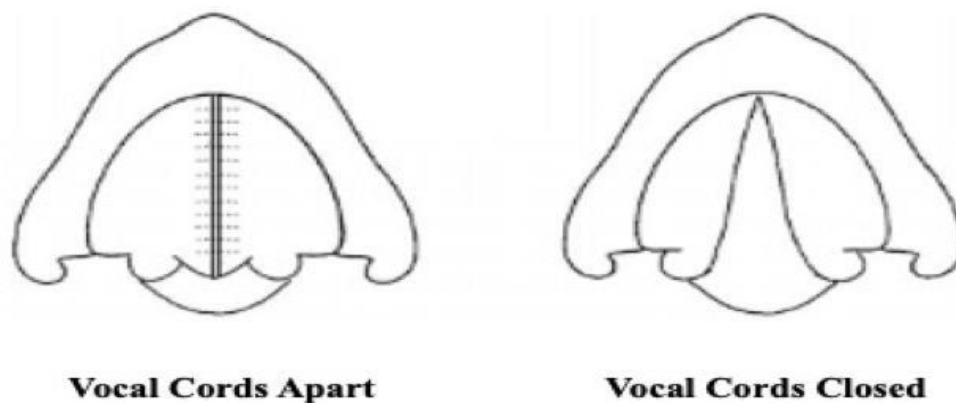
Figure 2 shows when the vocal cords are apart and closed to stop the air from escaping. Further, when the vocal cords open, they make a kind of vibration, and the sounds produced are called "voiced," but sometimes they are "voiceless".

Figure 3

The vocal cords during the production of voiced and voiceless sounds (Gut, 2009, p.21)

Voiced Sound   Voiceless Sound
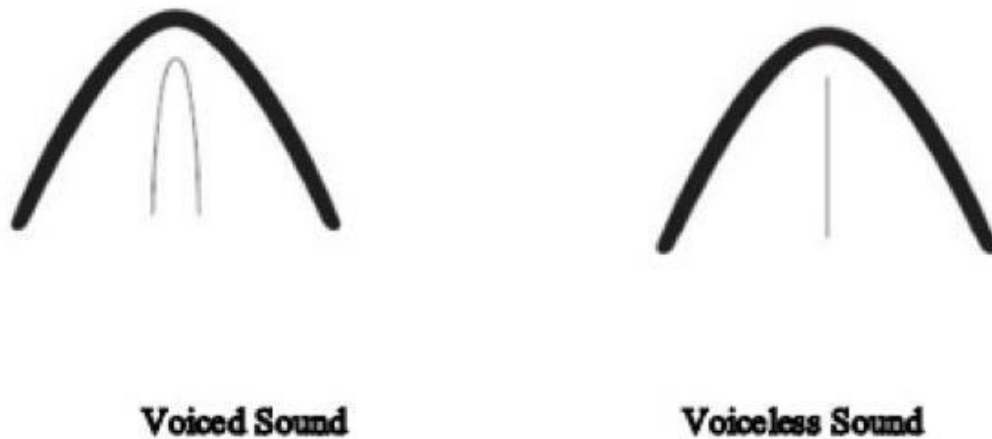
Gut (2009) showed in Figure 3, the position of the vocal cords while producing a voiced or a voiceless sound. It shows that the glottis is closed when producing a voiced sound, and the vocal folds touch each other lightly. However, in creating a voiceless sound, the glottis opens, and the vocal folds separate so the air can escape without making them vibrate.

1.3.3. The Pharynx

The pharynx is a speech organ which is also known as the throat. It goes up from the larynx and past the mouth to the nasal cavity. It is a tube muscle. Rogers (2000) and Gut (2009) suggested that the pharynx can be divided into two (3) different cavities which are: the pharyngeal cavity, which consists of the pharynx. The oral cavity, i.e., the mouth, is a crucial feature of speech production. The shape of the mouth helps proceed with many sounds. This cavity includes the lips,

teeth, alveolar ridge, palate, velum, uvula, tongue, and epiglottis. These are the main articulators used in speech production. The nasal cavity, which includes the velum, is a bony strutted soft palate forming the roof of the mouth, this area controls the production of sounds when it is lowered or raised, and the sounds made are called "nasal sounds." Moreover, Underhill, (2005) and *Cruttenden (1994*) illustrated the organs that take part in the production of these phonemes that happen in the vocal tract as follows:

Figure 4

Organs of Speech (Cruttenden, 1994, p.10)
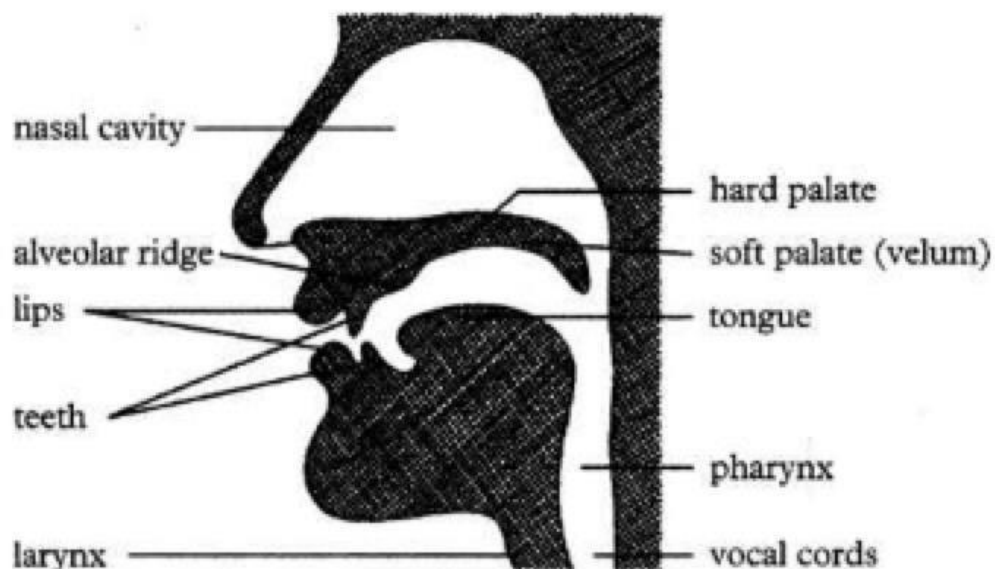
Figure 4 demonstrate the significant organs that help produce speech, including the lungs, the windpipe (trachea), the larynx, the pharyngeal cavity, the oral cavity, and the nasal cavity. The vocal cords are situated in the larynx (the glottis).

1.4. Speech Mechanisms

The production of speech happens due to the different speech organs mentioned above. However, knowing the organs cannot help learners distinguish between the different speech sounds. It is not easy to know a sound merely by its place of articulation, but it requires knowledge about its manner of articulation. Cruttenden (1994) explained how production occurs through these organs. The vocal cords are brought together because of the airflow pressure the lungs provide to make sound vibrations.

Similarly, Clark and Yallop (1995) provided in their book entitled "*An Introduction to Phonetics and Phonology*" an overview of how speech production is made. They asserted that airflow from the lungs is the energy source to produce a sound. The airflow pressure forces the vocal folds to vibrate at a particular frequency. Then, the air passes through the throat, mouth, or nasal cavity. Hence, Kreidler (2004) confirmed that Human speech is very much like the playing of a wind instrument" (p.18). Kelly (2000) added that sounds could be voiced or unvoiced. The voiced sounds are produced when the vocal cords vibrate, whereas there is no vibration when producing unvoiced or voiceless sounds.

Phonology is divided into two (2) areas: Segmental and Suprasegmental phonology. The former entails all aspects of speech and the individual sounds of the language, and the latter is about how humans speak to convey meaning.

1.4.1. Segmentation

One of the most crucial parts when learning the pronunciation of a foreign language is to know how its speech sounds are produced and articulated in a specific manner. EFL learners may

need help during that process due to their limited knowledge of the different sounds and their way of articulation. For learners to know these sounds and how they work, they must be aware of their small units. Kelly (2000) broke down the main features of pronunciation to show the different parts of the sounds of English.

Diagram 1

Segmental Features (Kelly, 2000, p.1)
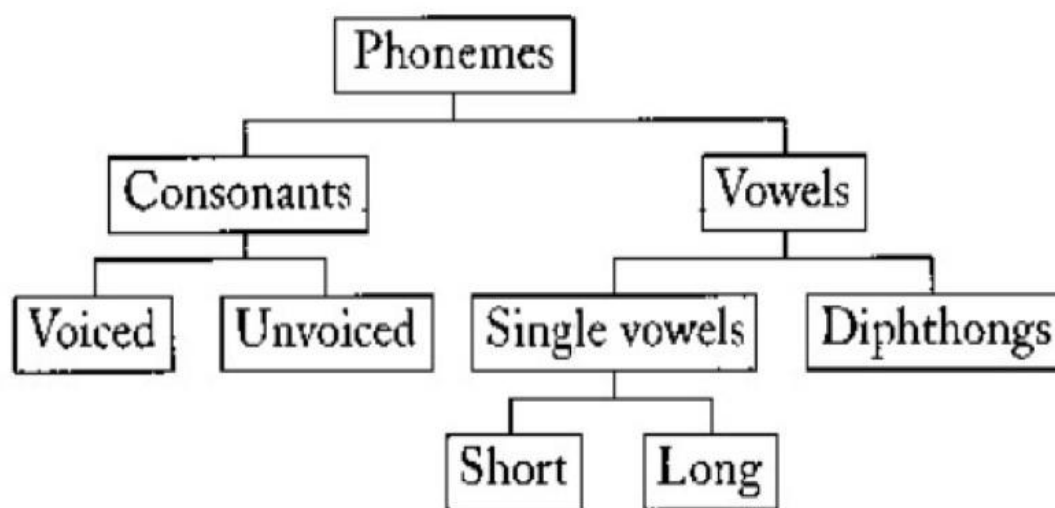
Diagram 1 demonstrates the segments of English pronunciation (i.e., the individual speech sounds). Phonemes are divided into two types: Vowels and Consonants. These phonemes can be distinguished merely based on their manner of articulation.

1.4.1.1. Consonants. Consonants are the sounds produced in a way where the air stream is sometimes squeezed or completely blocked because it involves the obstruction of the vocal tract. Rogers (2000) classified consonants into three dimensions: voiced or voiceless, their place of

articulation, and their manner of articulation. In the same way, Reed and Levis (2015) explained that three essential criteria might be used to describe consonants: whether they are voiced or voiceless, where in the vocal tract they are pronounced, and how they are pronounced. Furthermore, Brown (2014) confirmed, "Consonant segments can be divided into those that involve the vibration of the vocal cords (are voiced) and those that do not (are voiceless)" (p. 6). Therefore, consonants are segmented phonemes articulated when the vocal tract goes through a significant constriction.

1.4.1.2. Vowels. Vowels are sounds produced when the air stream comes from the lungs and moves smoothly without meeting any obstacles because there is no constriction in the vocal tract.

According to Gut (2009), "The manner of articulation of vowels is fairly restricted. The airstream passes the oral cavity relatively unhindered" (p.28). Accordingly, Celce-Murcia et al. (2010) clarified that vowel sounds differ according to the tongue position, the lip rounding, and the tense of the tongue muscles. Thus, they may be a single vowel sound or a combination of two vowel sounds, also known as a diphthong. A triphthong is another term when combining three vowel sounds (Kelly, 2000).

1.4.2. Suprasegmental

When EFL learners are introduced to pronunciation, they usually think it only conveys the different sounds of the language, vowels, and consonants. However, the knowledge and articulation of sounds are not the only aspects that may affect EFL learners 'pronunciation. Since the word segment refers to the language's phonemes, other pronunciation aspects are related to

speech. Ogden (2017) stated that "Suprasegmental are aspects of speech which persist over several segments, such as duration, loudness, tempo (speed), pitch characteristics and voice quality" (p.23). In other words, the musical aspects of speech include the voice quality, pitch, as well as posture of the tongue.

Diagram 2

Suprasegmental Features of Pronunciation (Kelly, 2000, p.1)
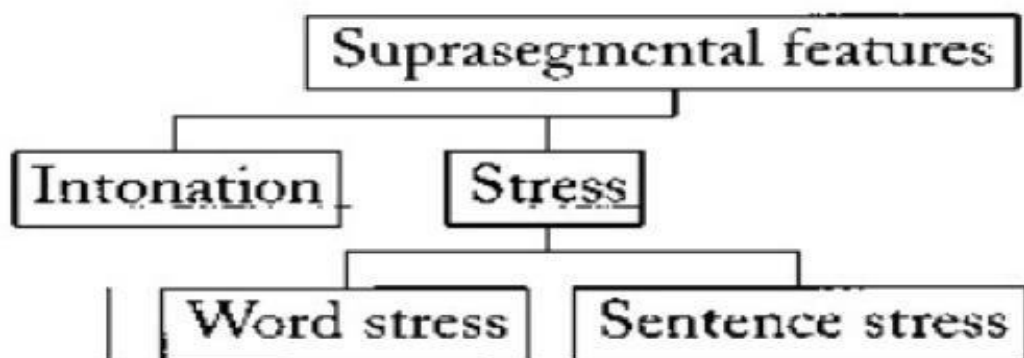
According to Kelly (2000), stress and intonation are the most critical suprasegmental features.

1.4.2.1. Stress. Stress is the main feature of suprasegmental related to conversational speech where words can be emphasized. Underhill (2005) defines stress as the term used to express the accent or emphasis placed on a specific word's syllable. Accordingly, Cruttenden (1994) emphasized that stress is related to the prominence of one syllable compared to other words in the same word. Roach (2000) added that "stress is a complex auditory impression which the listener perceives as making one syllable more prominent than its neighbors" (p. 94). In other words, stress

is about the prominence of more than one syllable of words. This syllable is produced with more energy and is mainly featured in English. When the speaker puts more emphasis on one syllable than the other, this stressed or prominent syllable may be distinguished by its loudness, higher pitch, or length, which gives English its rhythmic musical uniqueness.

Furthermore, Roach (2000) emphasized that stress characteristics may be studied on two levels: Perception and production. The prominence of syllables results from increased muscular energy and a higher sub-glottal pressure from the lungs, i.e., respiratory activity, pitch movement, and tone. Moreover, stress placement is different from one word to another; this makes it very difficult for foreign learners. Roach (2000) confirmed that "Many writers had said that English word stress is so difficult to predict that it is best to treat stress placement as a property of the individual word, to be learned when the word itself we learned" (p. 88). Hence, stress is limited to each word; learners need to learn about it as they learn a word individually. However, Clark and Yalop (1995) made stress syllable placement easier by observing the word and highlighting that stress falls on the first syllable in nouns and the second syllable in verbs; it can also sometimes fall on the root regarding prefixes and suffixes.

EFL learners may lack knowledge of the importance of stress in English pronunciation and need to know that the position of stress changes based on the meaning of the word.

1.4.2.2. Pitch: Tone and Intonation. Another aspect of suprasegmental pronunciation features is pitch which refers to the speakers 'level of loudness when speaking. All speakers have different pitch levels while producing different sounds; it may be high or low. Rogers (2000) says

that it is pretty similar to how we play different notes on the piano; each note has a quality, and it is the same pitch while speaking; humans control the production of utterances by modifying the vibration rate of their vocal cords. Therefore, speakers maintain the pitch level while speaking by holding the tension of the vocal cords. Tensed vocal cords and more significant air pressure produce a higher pitch, while less tightened vocal cords and little air pressure lead to a lower pitch when producing different utterances.

Additionally, Dobrovolsky and Katamba (2011) emphasized that "A language is said to have a tone or be a tone language when differences in word meaning are signaled by differences in pitch" (p.43). However, English does not work like that. English is a non-tone language. When a word is uttered with a different pitch, it does not change its meaning. Tone languages exist throughout North and South America, sub-Saharan Africa, and East Asia. This may affect EFL learners 'pronunciation because they bring that tone into the foreign language, making them sound accented (Jonah, 2013).

Underhill (2005) defines intonation as the speaker's choice of intonation contours that can convey attitudes, establish grammatical boundaries, emphasize the relationship between utterances, and show the common ground that the speaker and hearer have assumed. Therefore, intonation is essential in delineating the speaker's intentions, conveying his feelings and emotions, and describing his attitudes. EFL learners should be conscious that intonation can impart intelligibility in their speech while transmitting specific knowledge to a particular speaker.

1.4.2.3. Rhythm. Music also has a rhythm and language, but it is not intonation. Unlike intonation, rhythm can change the meaning. Music has a regular rhythm, a mixture of long and

short notes. Similarly, human speech follows a different rhythm while producing various utterances. Roach (1991) said, "It has always been claimed that English speech is rhythmical and that the rhythm is detectable in the regular occurrence of stressed syllables" (p.120). Theories confirmed that English has a stress-timed rhythm or is a stress-based language meaning that syllables have different rhythms controlled by whether the syllable is stressed. Stressed syllables are more extended than unstressed ones. Accordingly, Carr (2019) stated that "the rhythmic beats fall only on stressed syllables" (p.163). You can imagine the way the rhythm of the speaker moves from a long-period syllable to a short-period syllable according to the speaker's intentions and context. In other words, stress categorizes the level of rhythm within the patterns of the language. Despite this, language learners must understand that rhythm is essential to English. They must imitate and master the rhythm of this language to sound accurate and impact the listener.

Gilbert (2008) has another view on teaching suprasegmentals; he emphasizes the importance of teaching prosodic features to EFL learners to help them sound like native speakers. He asserted that prosody combines rhythm and melody, allowing the listener to interpret the speaker's intention. Thus, EFL learners must study the skills of the language in parallel and should not focus on one skill and neglect another.

1.5. Phonological Processes

EFL learners often feel uncomfortable while communicating with native speakers of English. They think that their spoken language has a fast rate, and it is almost impossible for them to understand every word uttered, and they all sound the same. Native speakers of any language, not just English, can speak quickly and naturally in a natural flow where their words are blended.

Underhill (2005) stated, "The degree of simplification of sounds depends largely on the speed and context of the utterance, as well as on the characteristics of the speaker" (p.59). The simplicity and speed of speaking are called connected speech. Words spoken in a particular context are different from words spoken in isolation, as defined by Hieke (1987), "the changes which conventional word forms undergo due to the temporal and articulatory constraints upon spontaneous, casual speech" (p. 41) (cited in Reed and Levis, 2015, p. 160).

Celce-Murcia et al. (2010) emphasized the importance of these processes since they carry the overall meaning. They explained that when two learners do not have the same native language or someone's native language is English, and the other speaker speaks another language; some mispronunciations may occur while communicating. These errors are mainly related to the mispronunciation of some segmental sounds. Therefore, EFL learners must be aware of the different features of connected speech and prosody characteristics to communicate successfully and even sound like native speakers. These aspects of connected speech include linking, which is the only process of connected speech processes that do not involve changes in segments of the words.

Speakers link two words and make them sound like one word without changing their segmental identity. Reed and Levis (2015) define it as "Linking can result in re-syllabification of the segments without them or lengthening of the linked segments in cases where both segments are identical" (p.162). Linking is about joining a word's ending sound to the next word's initial

sound. Further, some accents in English are said to be "Rhotic," which means that the letter "r" is pronounced when it is at the end of the word (Roach, 2000; Kelly, 2001).

Another aspect of connected speech is assimilation which is about the articulation of sounds in an efficient way. Kelly (2001) defines it as "Assimilation describes how nouns modify each other when they meet, usually across word boundaries, but within words too" (p. 109). Thus, assimilation is when a sound takes the characteristics of the proceeding sound.

Similar to assimilation, elision is an aspect of connected speech that occurs in casual, rapid speech. Roach (2001) defines it as "[…] in certain circumstances a phoneme may be realized as zero, or have zero realization or be deleted" (p. 127). In other words, elision is when sounds disappear or are not pronounced; it is also known as deletion or omission. Roach (2001) emphasized that EFL learners do not need to learn to use this, but they need to be conscious that native speakers may delete some sounds from their speech that they may expect to hear.

To this end, EFL learners should be aware of the sounds of English and how they are articulated, the different features of suprasegmental mentioned before, and the various aspects of connected speech. It is the teachers 'role to expose learners to the different aspects of connected speech for them to communicate.

1.6. Pronunciation Instruction

Correct pronunciation has always been a crucial starting point when learning a foreign language. It is required for successful communication to make sense to others while producing different utterances. Pronunciation research has increased throughout the years; Pennington and RogersonRevell (2018) dubbed that:

> A growing body of research demonstrates that pronunciation is an aspect of language and communication that demands attention in educational and workplace contexts where speakers who have different mother tongues seek to communicate in a common language, which in the world today is often English (p.2).

In light of what has been said, pronunciation is an essential aspect of language that must be treated and dealt with equally with the other skills. But how was pronunciation taught before? For decades, teachers taught pronunciation, but it was given little attention compared to the different aspects. Even when there was time to deal with this aspect, it took merely drilling activities or serious tedious tasks that did not address learners 'problems. Kelly (1969) refers to it as the "Cinderella area" of foreign language teaching and learning, and teachers concentrated on grammar and vocabulary more. Therefore, the following section will be dedicated to the approaches and techniques used to teach pronunciation in the past. It will consider comparing the effectiveness of each method or technique used and what is best.

1.6.1. History of Pronunciation Instruction

Researchers and specialists in pronunciation state that pronunciation instruction witnessed rises and falls throughout the history of foreign language teaching and learning.

Pronunciation teachers had different opinions on the importance of teaching it and how to teach it. Some methodologies and techniques had more significant impacts than others when it came to improving learners 'pronunciation skills.

1.6.1.1. History and Scope of Pronunciation Instruction Before the Nineteenth Century. Pronunciation instruction throughout history has been influenced by different trends that resulted in shifts in teaching it. Pronunciation instruction in the nineteenth century has seen teachers implement two main approaches. Before applying the analytic-linguistic principles to teaching pronunciation, the first approach, known as the "intuitive-imitative" method, appeared as "listen and repeat." It depended on learners 'ability to listen to the different sounds through phonograph records, tape records, and language labs, inviting learners to try to imitate them. Moreover, the second approach was recognized before the nineteenth century and known as the "analytic-linguistic approach," also referred to as "analyze and understand," which is all about using phonetics alphabets, articulatory descriptions, and charts of sounds that support listening imitation, and reproduction. It was implemented explicitly to enrich the learners 'knowledge of the different sounds and rhythms of the language. This approach was only created to accompany the intuitive-imitative approach, allowing learners to practice the acquired information (Celce-Murcia et al., 2010).

Similarly, Kelly (1969) asserted that the approaches to teaching pronunciation before the nineteenth fell into two groups: an intuitive and an analytical approach. He claimed that "The first group depends on unaided imitation of models; the second reinforces this natural ability by

explaining to the pupil the phonetic basis of what he is to do" (p. 61). These two approaches signaled a successful era in pronunciation instruction history. From many historical perspectives, they led to a new period in pronunciation teaching that led to the emergence of new approaches until the 20th century. Therefore, the next section will be about the different methods that appeared from the nineteenth century until the 20th century and were used to expand the realm of pronunciation instruction approaches and methodologies.

1.6.1.2. Pronunciation Instruction Towards the 20th Century. From the 1840s to the 1940s, there were two dominant approaches in language teaching: Grammar Translation and Reading-based approaches. Language teachers applied these two approaches to enhance learners' reading and writing abilities which also helped the emphasis on improving grammar and vocabulary. However, pronunciation was neglected and was considered irrelevant according to these two approaches. (Celce-Murcia et al., 2010). Following the late nineteenth century, phonetics was given much more attention due to the Reform movement in language teaching that many phoneticians influenced. Thornbury (2006) defined this movement as "a pan-European initiative aimed at a radical reform of existing language teaching practices in schools" (p. 95). Thus, this movement was necessary since it focused on phonetics and phonology. Phoneticians like Henry Sweet, Wilhelm Viëtor, and Paul Sassy, who developed the International Phonetic Alphabet (IPA), gathered and worked together to establish a new era of language teaching. They rejected the methods that did not focus on teaching phonetics and replaced them with new ones. Furthermore, their work together included significant changes where they argued over the primacy of the spoken

language compared to the written language and the urge to teach it first and provide solid phonetic training to language learners (Celce-Murcia et al., 2010).

Before the Reform Movement and its development, the Direct Method appeared as an alternative to the Grammar-Translation method. This method was similar to the intuitive-imitative approach because it emphasized teaching the spoken aspect of the language, and it was the first oral-based method that prioritized pronunciation instruction. Nevertheless, some different ways and methods emphasize teaching pronunciation. There were still questions about the efficiency of listening and imitating sounds and whether it was the best method for learners to improve their pronunciation. Therefore, more approaches came to life due to the reform movement, which played a significant role in developing the audio-lingual method in the US, and the Oral Approach in Britain, which relied on drilling and repetition (Celce-Murcia et al., ibid).

These primary historical events changed the concept of pronunciation instruction. Hence, EFL teachers and learners need to look at how phonetics and pronunciation were taught before and how the methods employed were effective. Pennington and Rogerson-Revell (2018) shed light on the different methods and approaches used to teach pronunciation before the twentieth century. Along with the reform movement, teachers 'views on language changed. They started to look into new ways to teach the language in isolation and develop learners 'grammatical and communicative competence. In this favor, the Communicative Language Teaching approach evolved in connection with Second Language Acquisition (SLA). This approach's primary purpose is communication,

and it mainly focuses on enabling students to communicate in real contexts for implicit acquisition where communication is central inside the classroom.

Thus, the rise of pronunciation instruction in the twentieth century has witnessed an era of evolving approaches and techniques that changed researchers 'and language teachers 'views on pronunciation instruction. Some researchers focused on strengthening other areas of the language rather than teaching pronunciation, while other specialists looked into ways to teach it (Pennington and Rogerson-Revell, 2018).

1.6.1.3. Pronunciation Instruction in the Twentieth Century. The Reform Movement had given birth to a new modern era that exhibited the importance of pronunciation instruction by developing an analytic-linguistic perspective on how to teach it. As Pica (1984) suggested, the 2000s reestablished pronunciation instruction with a central purpose: effective communication. Teaching a second or a foreign language was merely a job in this period. Pennington and Rogerson-Revell (2018) elucidated some of the methodologies developed during that period. English for Specific Purposes appeared based on needs analysis and for developing certain skills, as well as Task-based Language Teaching focused on enhancing specific skills. Teachers adapted both methodologies to improve learners 'communicative performance during pronunciation instruction. The mid-20th century was the era that emphasized pronunciation teaching and showed great efforts to incorporate communication as a central purpose for teaching it. Recently, pronunciation witnessed a new era of computer-assisted learning technologies that were elaborated to teach the language with a focus on pronunciation. However, teachers still use the previous approaches and

methodologies mentioned before up to this day, for example, listening and repeating related tasks adapted from the Direct Method. Due to the reform movement, teachers followed techniques of phonetic training to teach the different sounds appropriately (Celce-Murcia et al., 2010).

The previous section covered the traditional methodologies and approaches used in the language curriculum throughout the earlier decades that had significantly impacted and shaped the future of language teaching. The language taught was merely based on the type of methodology; some methodologies favored teaching pronunciation for communication, while others were restricted to teaching the whole aspects of the language equally. Hence, this aspect was essential to mention in this research because researchers and scientists still refer to and use these traditional ways of teaching pronunciation. They are still used until the present day. However, it is inevitably undeniable that the internet changed this whole aspect of language teaching.

1.6.2. *The Status of Pronunciation Instruction*

Pronunciation has historically evolved throughout the years and has witnessed many changes. Its instruction has sometimes been marginalized or put as a central goal inside the language classroom, which led to its status continually changing. And its importance has been questioned by many.

Therefore, this section highlights the importance and goals of pronunciation instruction. It will also spotlight the factors that may influence the process of pronunciation instruction and hinder it and how technology may affect pronunciation instruction.

1.6.2.1. Reasons behind Pronunciation Instruction. Pronunciation instruction takes the form of repetition drills that help learners improve their segmental and suprasegmental aspects of the language. However, their pronunciation needs to be stronger or accented. The importance of teaching pronunciation has always been the main big question regarding foreign or second language teaching. It has been argued that little attention is given to pronunciation instruction inside the classroom, and it is neglected for many reasons (Harmer, 2010).

However, regarding improvement in the different areas of the language, some teachers thought that improving learners 'grammar and vocabulary seemed more important than teaching pronunciation. In contrast, other teachers argued that having good grammar and vocabulary is not very important if learners cannot be understood when they speak. Hence, improvements in learners 'knowledge and practice of the different segmental and suprasegmental features of the language are not enough for them to attain a native-like pronunciation. For some teachers, it seems almost impossible after a certain age. Therefore, teachers may neglect this aspect because learners aim at reaching a native-like accent which is not a realistic goal, so teachers try to set an ideal goal of achieving an acceptable pronunciation while learning it.

In this sense, teachers often sacrifice and skip pronunciation-related tasks and activities in favor of the other language areas since they believe learners will develop their pronunciation naturally after practicing the other areas. Ergo, Yates, and Zielinski (2009) addressed this problem by stating that learners need to be proficient in pronunciation and master all the different rhythms and sounds and how to use them in connected speech. They mentioned that "[…] those with

unintelligible pronunciation will remain unintelligible, even if they have expressed themselves using an extensive vocabulary and perfect grammar" (p. 11). Since pronunciation is essential in demonstrating mastery of the language, some listeners may interpret poor pronunciation as incompetency or stupidity. Furthermore, not dealing with pronunciation inside the classroom may be because teachers lack enough resources or shortage of materials and tools to motivate their learners and teach them pronunciation appropriately. Thus, Frazer (2000) argued that teachers should be provided with design courses and materials that may help the process of pronunciation instruction so it will create an appropriate atmosphere for a proper learning process.

So, pronunciation is essential since learners need it to be understood while speaking. Before, learners needed to improve their reading and writing skills to master the language, but speaking also is an important aspect to focus on. Eventually, teachers will realize that their learners are learning the language to use it in real-life situations. Accordingly, even if learners master grammar and vocabulary, if their pronunciation is weak, their pronunciation will not be understood, which may lead their communication to fail.

1.6.2.2. Goals of Teaching Pronunciation. The previous section dealt with the importance of teaching pronunciation and why learners need to acquire it and work on improving it. For decades, many studies have approved that pronunciation has always been a neglected or ignored area in teaching English. Pronunciation is the most challenging area to acquire due to many reasons. Findings from the literature on second language learners 'ability to recognize and produce non-native speech indicate that factors other than age also have an impact. While it may seem appropriate to introduce foreign languages as early as possible, successful L2 pronunciation also

depends on a wide range of factors. For instance, gender, L1 background, the amount of L1 and L2 use, the length of time spent in an environment where L2 is spoken, and aptitude for and motivation for language learning (Flege in Press, cited in Aliaga-García, 2007).

Pronunciation may be a problematic aspect of the language to master since many factors interfere with teaching and learning it. It mainly relates to the differences between L1 and the second or foreign language. In addition, pronunciation is also considered "The Cinderella of language teaching" in many countries, with a low emphasis on teaching and incorporating it into the curriculum (Dalton, 2002). Moreover, Maniruzzaman (2008) disclosed, based on his experience as a learner and a teacher-researcher of English as a foreign language, that pronunciation teaching is excluded from the curriculum/syllabus by the designers who questioned the significance of teaching pronunciation. The following section is dedicated to identifying specialist thoughts and beliefs on learners 'goals regarding learning pronunciation and teachers ' role in making it possible.

1.6.2.2.1. Learners 'Goals. When learning a foreign language, learners strive to have good pronunciation for many other reasons, but the question is: "How good is the learners ' pronunciation aim to be?" EFL learners aim to achieve native-like comprehensible pronunciation, but it is a challenging goal (Munro & Derwing, 1995). However, intelligibility is deemed an optimal goal for EFL learners, to speak and be understood by the listener. Frazer (2000) believes that EFL learners can achieve proficiency and even mastery in grammar and vocabulary. Still, never pronunciation is because learners develop articulatory habits that are mainly related to their

L1 and may hinder the learning process of the new phonetic patterns of L2. Although EFL learners with a native-like pronunciation exist, they are scarce, especially when acquiring the new language occurs in the classroom exclusively, i.e., learners do not practice their language elsewhere outside of the classroom and take the instruction only from their teacher. Accordingly, learning a foreign language at any age will leave one with a slight but noticeable non-native accent. Therefore, a more achievable, reasonable, and realistic goal for EFL learners is to be comfortably intelligible (Szpyra-Kozłowska, 2014).

Furthermore, Kenworthy (1990) related comfortably intelligible to the notion of hearing and listening, two different concepts. Hearing requires the presence of the listener and the ears while listening requires mere work. Sometimes when speaking, we ask someone to listen carefully and accuse them of not listening to what is being said. It is because the listener works harder to interpret the speaker's utterances. The speaker may find difficulties in understanding the speaker if their pronunciation is not clear or if they pronounce in a way in which the listener has to always ask for repetition. "In setting goals for our learners, we must consider the effect of mispronunciation on the listener and the degree of tolerance listeners will have for this" (Kenworthy, 1990, p.4). In this way, intelligibility means different things to different people, and it depends heavily on the listener's attitude and point of view. EFL learners must be aware of setting more realistic goals regarding their pronunciation and should focus on being understood by the listener rather than setting higher aims that may be impossible to achieve.

1.6.2.2.2. Teachers'Role. What was highlighted in the previous section revealed how EFL learners aim to achieve good pronunciation. Although a good accent was hard to define whether

they should master a native-like or a near-native pronunciation, experts suggested that learners seek intelligibility, converse, and be understood while communicating. However, it is not just the learner's responsibility; the teacher also sets goals for his learners and what he aims to achieve and is ready to accomplish his learners 'needs and desires. Pennington and Rogerson-Revell (2018) declared that students must set precise models and reachable and realistic goals, but it can be challenging for teachers to decide what these should be. In other words, the teacher here faces more significant obstacles since he will be dealing with different types of learners with different personalities and goals, which may create challenges for him.

The teacher is the leader and controls everything that happens inside the classroom. When speaking about teaching, some questions may arise, like What is teaching? And how should a teacher be? Cambridge International Dictionary of English defined it as "Teaching is to give knowledge to someone." At the same time, the Longman Dictionary of Contemporary English describes it as showing someone how to do something or even changing someone's ideas. In addition, Gibran (1991) said, "If the teacher is indeed wise, he does not bode you enter the house of his wisdom, but rather leads you to the threshold of your mind" (p. 76). Therefore, the teacher is the primary source of knowledge in the classroom, and he should transmit that knowledge in a way that leads learners toward their goals and shapes their minds and thinking.

Furthermore, teachers may not feel comfortable dealing with pronunciation teaching as Pennington and Rogerson-Revell (2018) said, "There may be other reasons for the limited focus on pronunciation within language teaching as well, including, for example, lack of teacher confidence, skills, or knowledge regarding pronunciation teaching; lack of appropriate teaching

materials; or lack of time or space in the curriculum" (p. 180). Since pronunciation is neglected and given little attention inside the classroom, learners may go through other resources to try and achieve their goals. Moreover, the purpose of pronunciation has shifted from trying to achieve a native accent to a more realistic goal of developing intelligibility (Morley 1991).

Learning a new language is considered a complex process in which learners face different problems and obstacles and vary in their success. Hewings (2004) pointed out many accents and variations within the English language in different regions of the same country. Since English is now widely used in other countries as a foreign language, this has led to the emergence of new varieties. (Volín & Skarnitzl, 2010). The different variations of English are discussed in the following section.

1.6.2.3. Varieties of English: What Model to Teach. English is not a language that is only spoken by English people but also by many people from different countries. English language development and its increased usage by other speakers led to different variations. It varies from country to country based on the people who use it and for what purpose. Studying varieties of English is closely related to language study. Therefore, this section focuses on revealing English's different varieties and variations.

Before speaking about varieties, it is essential to shed light on the notion of the terms accent and dialect. Yule (2014) mentioned that we say someone speaks with an accent, "the description of aspects of pronunciation that identify where an individual speaker is from regionally or socially"

(p.243). Regional and social belonging is what makes someone speak with an accent. A dialect is a language variety distinguished from other varieties by differences in grammar, vocabulary, and accent (Watt, Trudgill, Hughes, 2013; Yule, 2014). Although these two terms differ, many people confuse their meanings. A dialect is about the characteristics of the language, but an accent is about how people pronounce it. A variety is different from another variety in terms of accent, the vocabulary used, the characteristics of its pronunciation, and its various sounds, which are essential for EFL learners to distinguish between these varieties and know the differences between them based on how they are used (Wells & Wells, 1982).

1.6.2.3.1. Standard English. In other languages, the term Standard, sometimes referred to as selected, is about the dialect chosen based on the official decision. However, Standard English was developed in the Royal Court in London. Pre-Standard English was then selected as the predominant language because it was associated with people with mixed geographical origins. Later, over many centuries, the dialect, i.e., Standard English, became the preferred language governing classes chose to write in. Hannah and Trudgill (2017) spoke furthermore of this dialect "[…] the dialect was from the very beginning an upper social class dialect which was not associated with the common workers and peasants" (p.1). Because this dialect was linked to the upper class, it was imposed on society.

1.6.2.3.2. British and American English. Over the decades, English has changed into Englishes. World Englishes is a term that refers to the use of English in many parts and regions of the world, which led to the appearance of different varieties. First of all, English is spoken in nation-states as a native language where people have English as their mother tongue, and they use

it to communicate daily. Although there are wide varieties of English worldwide, there are two native varieties that are typically widespread and used among native language states American English and British English (Hannah and Trudgill, 2017).

Kang, Thomson, and Murphy (2017) described standard variations as the phonological patterns that powerful social groups use. In England, the traditional standard pronunciation is "Received Pronunciation," abbreviated to RP, the most prestigious variety spoken by influential, wealthy natives. The term Received goes back to the nineteenth century when it was perceived as the accepted variety used by most polite people. Many scholars spoke of variety as "[…] the accent of those at the upper reaches of social scale as measured by education, income, and profession or title" (p. 3). It has traditionally been the accent of educated people and the variety used in private schools (Watt, Trudgill, and Hughes, 2013).

American English is the variety spoken in the United States of America. This variety has roots in the early seventeenth-century settlers on the Eastern Coast. Later on, this variety evolved and was influenced by immigrants from the Southern Irish Isles who arrived due to economic reasons. Although it has witnessed more immigrants from Germany, France, and Italy, it did not influence its emerging profile (Hickey, 2014). Additionally, the majority of native speakers of English use American English pronunciation, vocabulary, and grammar due to its high status and dominance in media, business, and science, which made it the variety that is primarily and widely used worldwide. However, many scholars still consider British English as the predominant variety. Hence, Totti (2002), a Swedish-born Professor of English Linguistics at the University of Zurich,

explained that the majority of books on the English language consider British English as the main variety because it is older and a prestige variety.

Overall, it is fair to say that Standard English is the variety used to teach EFL learners grammar and vocabulary or the written form of the language. In contrast, RP is the variety used to teach pronunciation.

1.6.2.4. Factors Affecting Pronunciation Instruction. Pronunciation is crucial to foreign language learning since it can affect learners 'communicative competence. One of the significant points when learning a foreign language is to secure an understandable pronunciation that allows you to communicate successfully with others. However, EFL learners cannot achieve that due to some factors that interfere with learning the different features of the language. Wong (1987) highlighted that teaching pronunciation is not merely a linguistic matter, considering that some factors may hinder that process. Therefore, this section is devoted to shedding light on the factors that may prevent learners from reaching intelligibility: learners 'native language and other biological factors.

1.6.2.4.1. Native Language: Exposure to the Target Language. One factor that interferes with learning a foreign language is mainly related to the person's origins and native language, which affects the learning process in many ways. Kenworthy (1988) asserted that a learner's native language plays a significant role due to the differences between the sound systems of the two languages. Still, she argued that it is not an impossible goal since " […] people from many

different language backgrounds can and do acquire a near-native native pronunciation in English"
(p.4). Further, Celce-Mercia, Brinton, and Goodwin (1996) also argued that when two languages
contrast, different points in sounds of the native and foreign language may affect how the learner
articulates these sounds. Hewings (2004) claimed that phoneme differences between the first and
foreign languages occur when learners are exposed to combining sound rules, i.e., rules limited to
each language and each language's specific rules; here, learners may fail to produce these sounds
correctly. Learners face difficulties in applying stress patterns and intonation to the foreign
language, which results in failing to pronounce the foreign language appropriately (Rosyid, 2016).

Furthermore, exposure to L2 has the utmost importance in achieving intelligibility. Many
studies about L1 acquisition revealed that children acquire their mother tongue language without
effort due to its unlimited exposure, allowing them to acquire it naturally and in ordinary settings.
On the other hand, adult learners need to set goals to achieve proficiency, as explained by Jarosz
(2019) "Adult learners, on the other hand, have to make a conscious effort, and so they seldom
manage to acquire native-like accent and proficiency" (p.54). Therefore, reaching proficiency in
foreign language pronunciation is not an impossible goal to achieve it requires learners 'desire and
efforts and teachers 'encouragement to motivate learners to use and practice the language outside
the classroom setting to improve their pronunciation (Gilakjani, 2012).

1.6.2.4.2. Biological Factors: Age, Aptitude, Attitude, Motivation. Other internal factors
are to be considered as they directly relate to the learners. Zhang (2009) reviewed these biological
factors, or individual differences among learners, which may affect their language proficiency.
Hence, age is a factor many consider the main reason behind proficiency in learning a foreign

language, especially in learning pronunciation. The claim is that younger learners can do better than adults when learning a foreign language "[…] adults who learn to speak a second language fluently, but still maintain a foreign accent, even when they have lived in the host country for many years" (Kenworthy, 1988, p.4). Thus, adults may fail to achieve a near-native-like pronunciation even when exposed to a foreign language, while children can accomplish a native-like accent easily. Confirming that claim, Lenneberg (1967) suggested the "Critical Period Hypothesis," a biological life period in which language can be acquired easily; it ends at 12. After this period, it becomes complicated for the learners to master a foreign language, and they will probably speak it with a foreign accent. However, this theory was unaccepted by some scholars and language teachers who viewed that when learners get older, their perception of the different sounds of the target language increases, and the distinction between the two sound systems of the two languages becomes more stable (Flege, 1987 & Patkowski, 1990). In light of what has been said, adult learners are mature in recognizing the differences between the two languages and have cognitive knowledge to help them achieve native-like pronunciation. However, it is agreed upon that learning a foreign language and being exposed to it at a younger age is more efficient and will leave the learners with a slight and not very noticeable accent.

Another biological factor is the learner's attitude and aptitude toward learning the target language. Zhang (2009), Celce-Mercia, Brinton, and Goodwin (1996), and Caroll (1962) argued that language aptitude constitutes four main characteristics, (1) Phonetic Ability, which is related to learners 'capacity to learn and discriminating between the different sounds of their native language and the target language. (2) Grammatical Sensitivity is about learners 'understanding and awareness of the different language rules. (3) Learning Ability on how learners acquire the

language according to the exposure to the target language. And (4) Learners 'Memory in recalling and learning new things. However, language aptitude is not a factor of success or failure. It does not mean that people with higher language aptitude tend to be more successful than others since it is all a matter of individual differences that change from one person to another. Furthermore, Gilakjani (2012) and Riyahifar (2016) emphasized that in addition to aptitude, learners 'attitude toward learning a new language plays an important role. Their attitude can increase or obstruct their attitude toward learning the target language, as mentioned by Kenworthy (1988) "[…] learners who show positive feelings towards the speakers of the new language tend to develop more accurate, native-like pronunciation" (p. 8). Thus, if the learners have a negative attitude towards the target language, it will influence their language learning in a wrong way. Motivation is also influential in learners 'attitudes toward the target language and can affect their learning. Motivation is an inner drive and desire that urges someone to act in a particular way. According to Gardner & Lambert (1972), and Brown (2002), there are two main types of motivation: integrative and instrumental. The former relates to the person's positive attitudes toward the target language and the desire to belong to the target language community. At the same time, the latter is considered a matter of immediate desire to gain certain things and attain specific goals. Hence, if the learners are motivated to achieve good pronunciation and are eager to learn and use the target language, they will learn it and promote their pronunciation.

Ultimately, many factors interfere with mastering the pronunciation of the target language. On the one hand, some factors are related to learners 'native language due to the differences in the

combination and sound rules between the two languages that may hinder learners 'ability to speak without an accent. On the other hand, other biological factors concern the learners 'characteristics and personalities. Factors like learners 'motivation, attitude, and exposure to the target language may positively or negatively affect learners 'pronunciation, promote it, and enable them to have intelligible pronunciation.

1.7. Pronunciation Assessment

Pronunciation has recently witnessed a lot of interest from researchers "[…] the history of L2 pronunciation has been compared to a pendulum swinging back and forth between times when it has been completely ignored, and times when it has been of primary importance" (Kang and Ginther, 2017, p.1). Thus, researchers recently led research projects on foreign language pronunciation and assessing speaking proficiency (Baker, 1982, Derwing and Munro, 2009). However, Isaacs and Trofimovich (2017) spoke of this matter negatively, as it has taken decades for researchers to realize that it is high time to start investigating how to assess pronunciation. "Interest in L2 pronunciation from within the language assessment community specifically, including researchers and practitioners, has taken much longer to ignite" (p.4). On the other hand, all books on pronunciation are only restricted to aspects of pronunciation and sound articulation. Still, none are dedicated to how to assess second or foreign-language pronunciation.

Assessing pronunciation in a second or foreign language creates a huge issue because of the native-speaking model (Traditionally, the accent was set following a native-speaker model). However, this model was challenged when English became World Englishes and a lingua franca

that impacted pronunciation instruction and assessment (Bøhn & Hansen, 2017). Kang and Kermad (2018) in King, Thompson, and Murphy (2018) confirmed that "[…] given that defining a norm is problematic in the era of English as a lingua franca, assessing L2 pronunciation faces more challenges now than ever before" (p.511). These changes have made assessing the language less native speaker focused, setting language learners 'proficiency, and defining objective rating criteria. Before, pronunciation assessment concentrated on the accuracy of segments and the mastery of suprasegmental. However, intelligibility, comprehensibility, and accentedness have recently been prioritized (Kang, Thomson, Murphy, 2018; Kang and Ginther, 2017).

Therefore, this section investigates how pronunciation was historically assessed and measured. Due to the development of technology, pronunciation is reckoned through new technologies that make it easier to measure it.

1.7.1. Pronunciation Assessment Development

Before, teaching pronunciation has led learners to set an unrealistic goal of achieving a native-like pronunciation and erasing the traces of their foreign accent. However, teachers have recently sought to help their learners reach intelligibility (Spezzini, 2010). Teachers should have the skills to make learners intelligible to conduct pronunciation assessments. Since it is difficult to achieve, they should follow a set of techniques and methods to guide their learners to self-assess their pronunciation and progress. As Pennington and Rogerson-Revell (2018) said, "Assessing a complex, multi-dimensional human ability or skill set such as spoken language competence and

pronunciation specifically is a matter of sampling the domain of interest" (p.289). Thus, how was pronunciation assessed?

In the early years, L2 speakers read passages that included all target language phonemes, which made reading difficult, mainly to try to sound like a native speaker. The errors were later assessed as pronunciation errors, although reading and speaking differ. This early approach focused primarily on evaluating accented speech patterns as Spezzini, Baratt, and Carter (2018) highlighted, "[…] accented speech, and non-native speech patterns were examined from a contrastive approach, which focused on the specific difficulties (mostly segmental) experienced by speakers from a given L1 who were learning an L2" (p.2). This approach's focal point was the approximation of imitating segmental features since imitation was the main feature in the Direct method that began in the late 1800s. Throughout the early years, pronunciation assessment mainly focused on the accuracy of imitation to achieve a native-like speech until the rise of the Cognitive Approach in the 1900s that led to a significant change in the notion of a native-like pronunciation which is unattainable. The assessment shifted from assessing segmental features (vowels and consonants) to including suprasegmental features (stress, rhythm, intonation). After that, the Communicative Approach reformed the assessment of pronunciation internally and led it to primarily focus on developing successful communicative skills. Thus, the focus was more on suprasegmental feature mastery rather than segmental components. Munro and Derwing (1999) emphasized, "Prosodic features appear to be a more potent force in the loss of intelligibility than phonetic errors" (p. 286). Moreover, achieving a native-like pronunciation shifted towards a more realistic goal: to help learners become intelligible. Hence, mastering suprasegmental features has

had high priority in assessing learners 'pronunciation and improving segmental features mastery since they ensure successful communication and consider the notion of World Englishes (Spezzini, Baratt, and Carter, 2018).

1.7.2. How to Assess Pronunciation?

Pronunciation assessment creates challenges in scoring learners 'performances and how the teacher can measure learners 'speaking proficiency and progress. Harding (2018, as cited in Kang & Ginther, 2018) raised some questions based on this issue that is central to success in assessing the language "Are administration and scoring procedures accurate and consistent? Does the task used in the assessment elicit relevant target features? Does the task yield a score that is fit for decision-making purposes? Is the assessment fair?" (p.2). Thus, pronunciation assessment has many critical issues; teachers find it challenging to plan, score, and measure based on validity or scales. As a result, Munro and Derwing (1995a, 1995b) in Kang and Ginther (2018) conducted extensive research on assessing and evaluating second language pronunciation. They entailed three main criteria, (1) accent that can be considered based on the listener's perspective, and it was assessed on a 9-point scale (1= no foreign accent, 9 = very strong accent). (2) intelligibility is about how comprehensive the speaker's message is at the word and utterance level. Another criterion is (3) comprehensibility which is defined by Munro and Derwing (1995a, 1995b) as "listener's perceptions of difficulty in understanding particular utterances" (p.291) which is also measured based on a 9-point scale (1 = extremely easy to understand, 9 = impossible to understand).

Moreover, Kang and Kermad (2018) argued that it might be possible to obtain reliability in numbers, but measurement in pronunciation may not be valid or reliable. Pronunciation scale measurements are inconsistent and unclear since the speaker's speech may not be intelligible and understandable by a group of listeners but may be perceived differently by another group of listeners and in another context. On the other hand, Spezzini, Baratt, and Carter (2018) provided some pedagogical implications to assess learners 'pronunciation. The assessment process involves diagnosing learners 'output by reading, telling a story, or responding to questions. Still, before that, they are exposed to native speakers 'input and allowed to rehearse and repeat the recording. Afterward, instructors analyze those readings "Instructors may assess these recordings analytically by conducting a linguistic analysis of segmental and suprasegmental features" (p.4). To assess learners 'intelligibility, learners receive feedback to help them self-assess and identify their challenges and difficulties.

1.7.3. Technology and Pronunciation Assessment

With the development of technology, computers have eased the way for instructors to teach and assess pronunciation to non-native learners. Nowadays, technologies known as Computer Assisted Language Learning or CALL facilitate language learning and allow learners to master all the aspects of the language. In addition, Computer Assisted Pronunciation Training CAPT has significantly affected teaching and learning pronunciation. These technologies provide different resources (audio, print, video) and the possibility to study anywhere and anytime. Thus, these technologies allow the analysis of speaking proficiency since they provide automated feedback.

Pennington and Rogerson-Revell (2018) said, "One of the main benefits of CAPT technology is the opportunity to provide automated feedback, and the use of speech technologies can be particularly beneficial for giving feedback on pronunciation" (p.236). Thus, computers have high precision and consistency of measurements when analyzing speech.

Moreover, Automatic Speech Recognition ASR is one of the most promising automated pronunciation evaluation systems to assess learners 'pronunciation. This software asks learners to read specific phrases and utterances, and then the computer identifies the incorrect phonemes and sounds. It also enables learners to keep trying until reaching the target pronunciation. They (i.e., software) are considered a great source of feedback since teachers cannot give their learners the same amount of attention and feedback (Kang and Ginther, 2018). However, it is crucial to recognize the model of the software used and the functions it entails to assess non-native speech and prioritize reaching intelligibility.

Conclusion

Teaching pronunciation may be challenging because it requires effort and training. Pronunciation is considered to be a difficult area to deal with, yet it is a crucial aspect of the language that needs special attention. Teachers need to be trained in teaching it to use the most effective ways to teach and assess it. In addition to, the lack of teachers 'training, many reasons are behind the fact that pronunciation is not being taught appropriately, among them the lack of classroom equipment and technological means. Moreover, learners lack the enthusiasm to learn pronunciation, there is no authentic input, the absence of a target language environment, and the appropriate materials to teach and measure pronunciation. However, technological advances may have led to new technologies that could ease this process and help learners improve their pronunciation.

Chapter Two

Automatic Speech Recognition Technology

Chapter Two: Automatic Speech Recognition Technology

Chapter Two: Automatic Speech Recognition Technology

Introduction

Over the past decades, Automatic Speech Recognition, known as ASR, has been an active research area that has made human-human and human-machine interaction an effortless task. This technology allowed human communication with machines and led to the development of artificial intelligence. The latter enabled users to figure out how to communicate and exchange information with technological devices without using traditional known input/output tools such as keyboards and mouses but through voice-based input and output. Recently this technology has seen some technical developments and authentic progress in inventions that even allow language learners to improve their speaking skills in the target language. It has dramatically changed human-machine interaction. Speech has the potential to enable achieving specific tasks where natural language communication is valid, and the keyboard and mouse are no longer helpful. Therefore, this chapter aims to provide information on this technology, how it works, its main mechanisms, and how it can be applied in education as an effective tool in enhancing the learning process, especially in teaching and evaluating pronunciation.

2.1. Automatic Speech Recognition Definition

CALL is a word that surfaced and impacted the world of education decades ago before recognizing ASR systems. This acronym refers to Computer Assisted Language Learning, which uses computers in teaching and learning. Beaty (2013) defined it as "any process in which a learner uses a computer and, as a result, improves his or her language" (p.7). Thus, CALL is a young branch of applied linguistics that has been used for years to teach different aspects and areas of

language. However, pronunciation was the area of language that was neglected because of the difficulty when teaching and assessing it. This has led researchers to develop CALL systems specialized in pronunciation, known as CAPT, which refers to Computer Assisted Pronunciation Training. It is about using computer technology to teach and learn the different segmental and suprasegmental features to improve language pronunciation (Abuseileek, 2017). One of the most developed advanced systems under the realm of CALL and CAPT is known as ASR technology which stands for Automatic Speech Recognition.

As mentioned above, CALL systems allow automatic language learning and assessment and enable access to various activities without time constraints. "CALL is an amorphous or unstructured discipline constantly evolving both in terms of pedagogy and technological advances in hardware and software" (Beaty, 2013, p.8). However, these CALL systems have minimal access to speech recognition. They are limited to vocabulary and grammar drill exercises, which are essential and not sophisticated enough to teach writing and speaking. CAPT systems are powered through the use of ASR technology in order not only to teach pronunciation features but also to assess improvements and provide feedback. These systems are highly acknowledged because they provide automatic, spontaneous, and individual feedback for pronunciation learners. These software and programs on computers give learners more freedom and independence to learn and select the learning material at any place and time with positive and instant feedback that enables the learners to detect and correct their errors (Neri et al., 2002).

To highlight what has been said before, ASR is a technological tool that can recognize and analyze speech. Levis and Suvorov (2014) define it as "an independent, machine-based process of decoding and transcribing oral speech. A typical ASR system receives acoustic input from the speaker through a microphone, analyzes it using some pattern, model, or algorithm, and produces an output, usually in the form of a text" (p. 1). McCrocklin (2014) spoke of ASR as a tool that enables learners to practice the language at their own speed by getting feedback that promotes their autonomy, develops their skills in learning on their own, and provides feedback.

Further, M. Carrier (2017) spoke of this technology as "Automated Speech Recognition ASR which converts audio streams of speech into writers 'text" (p. 46). This technology allows teachers to address their learners 'problems and enables them to work and practice independently. Teachers are now open to using ASR-based CAPT systems. Since they offer significant advantages to teaching and learning a language, mastering its components, providing opportunities for autonomous learning, and endless rehearsing without time constraints or negative feedback. It gives advantages like:

- o   Task-based speaking activities.

- o   ASR-based technologies can reduce learners 'anxiety in the language classroom since many learners feel embarrassed about making errors in front of their classmates. Still, this technology allows student-computer interaction only, and feedback is given privately through headphones.

- o   This technology allows learners to develop the language independently on their own which makes learning more realistic, rewarding, and fun. (Purushotma, 2005).

Furthermore, research into these ASR technological systems has grown since the studies have revealed that learners enjoy training and improving their pronunciation through ASR-based programs (e.g., Chou, 2005, Mich, Neri, and Giuliani, 2005). Many studies conducted in the light of research in the area of ASR technology have investigated the reliability of assessing and scoring algorithms of this technology (e.g., Gerosa & Giuliani, 2004). However, no data has been collected on the effectiveness of using ASR technology. Research mainly focused on applying these technologies rather than their effect on the learners 'needs. These technologies are applied to different devices people use in various domains and daily. However, no one knows their educational impact (Neri et al., 2002).

Learners 'native language and mother-tongue sounds may create a more complicated challenge when applying these ASR technologies. Thus, Neri et al. (2002) pointed out that "it is even more acute in the case of ASR-based CAPT systems, recognizing and evaluating non-native speech with current ASR technology still implies the risk of errors" (p.395). In addition, Katyal, Kaur, and Gill (2014) emphasized that significant problems can be faced while working with ASR technology. This problem concerns the fact that this technology has sensitive recognizers that may respond to other surrounding sounds since specific environments may include other sound sources that limit the technology's performance and only take place in areas with no noise in the background.

2.2. History of Speech Recognition

This section gives a brief historical background on ASR technologies development and researchers, developers, and engineers significant contributions to speech recognition. Below are

bits of history on how ASR flourished. This section includes a brief overview of ASR development, its contributions and adjustments throughout the years, and expectations for this technology in the future.

2.2.1. Early Efforts: Speech Synthesis

The desire to develop a synthetic speech started decades ago by Professor Christian Kratzenstein in 1779, who constructed acoustic resonators that resembled the human vocal tract. Then in 1791, Wolfgang and Von Kempelen in Vienna developed an Acoustic-Mechanical speech machine that could produce different sounds. His device has led to discoveries regarding human vocals. Later on, in the mid-1800s, Charles Wheatstone developed a new version of the Kempelen speech machine that could produce vowels, consonants, and even words (Flanagan, 1972). Below is an illustration of Wheatstone's version of Von Kempelen's speaking machine.

Figure 5

Wheatstone's Version of Von Kempelen's Speaking Machine (Flanagan, 1972, p.206)
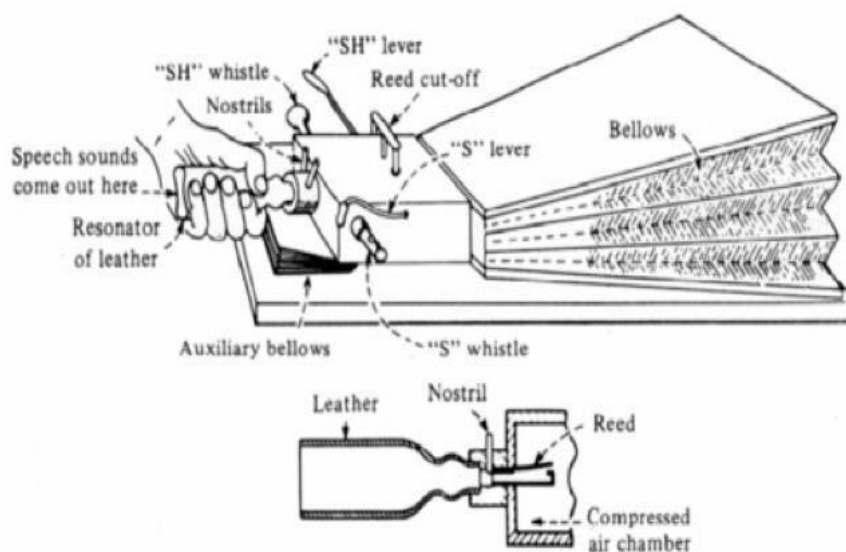
The device shown in Figure 5 is operated by resting the arm in the prominent bellows and blowing the air through a vibrating reed to produce voiced sounds, and the right-hand fingers control the air passages. Vowel sounds require the closure of all air passages and the activation of the reed, while unvoiced sounds are produced with the reed off and a turbulent flow through the passage (Flanagan, 1972).

This discovery, i.e., Kempelen's speaking machine, inspired Alexander Graham Bell to construct a similar device. Bell was inspired when he saw Kempelen's replica of the machine. With the help of his father, Alexander Melville Bell, and his brother Melville's assistance, he created his own speaking automaton (Swamy and Ramakrishnan, 2013, Flanagan, 1972). This discovery opened the door to new advances in the field of speech recognition.

2.2.2. The 1930s to 1940s: Speech Recognition

In 1928, the engineer Homer Duddley conducted the first voice encoding experiments. He developed the first electronic speech synthesizer machine, Voder/Vocoder, inspired by a voice encoder. This machine was later developed with the help of his fellow engineers Riesz and Watkins in 1936. This speech synthesizer was used for secure radio communications in which the voice was digitized, encrypted, and transmitted. Dudley's Vocoder was considered a significant milestone in speech recognition. It was also used and funded by universities and the US government for encrypting communications for Allies during the Second World War (Swamy and Ramakrishnan, 2013).

In 1939, skilled, trained operators displayed an electrical speech mechanism known as the Voder (Taken from VOice DEmonstratoR) at the New York and San Francisco World's Fairs. It was an excellent device since it could talk, and people could get trained to play it (Dudley and Tarnoczy, 1950, Gold, Morgan, and Ellis, 2011).

Figure 6

Sketch of the Voder (Dudley and Tarnoczy, 1950, p. 166)
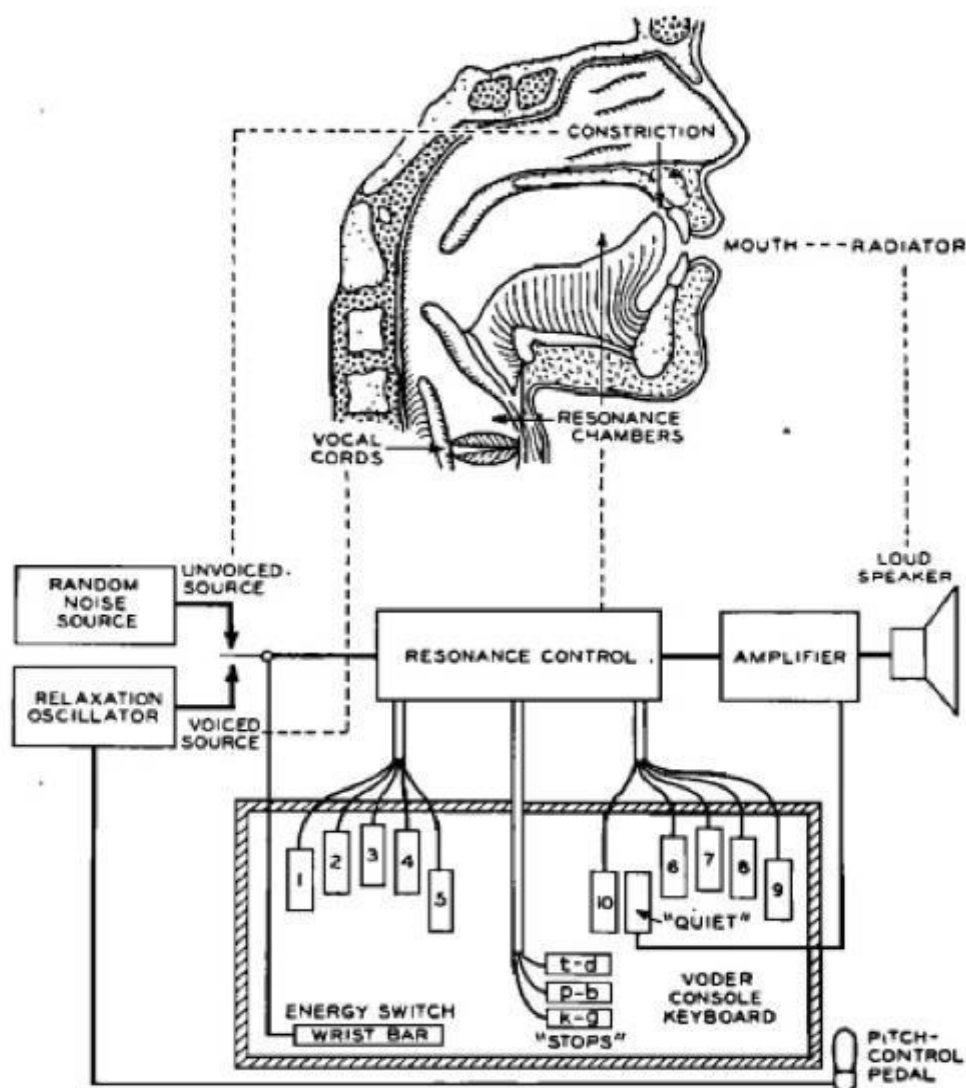
Figure 6 is a schematic of the voder that shows the electrical circuit for producing speech with manual controls.

The following figure displays a device known as the vocoder (Taken from VOice CODER) which has a similar electrical speech synthesizer to the voder.

Figure 7

Vocoder Controls (Gold, Morgan, and Ellis, 2011, p.12)
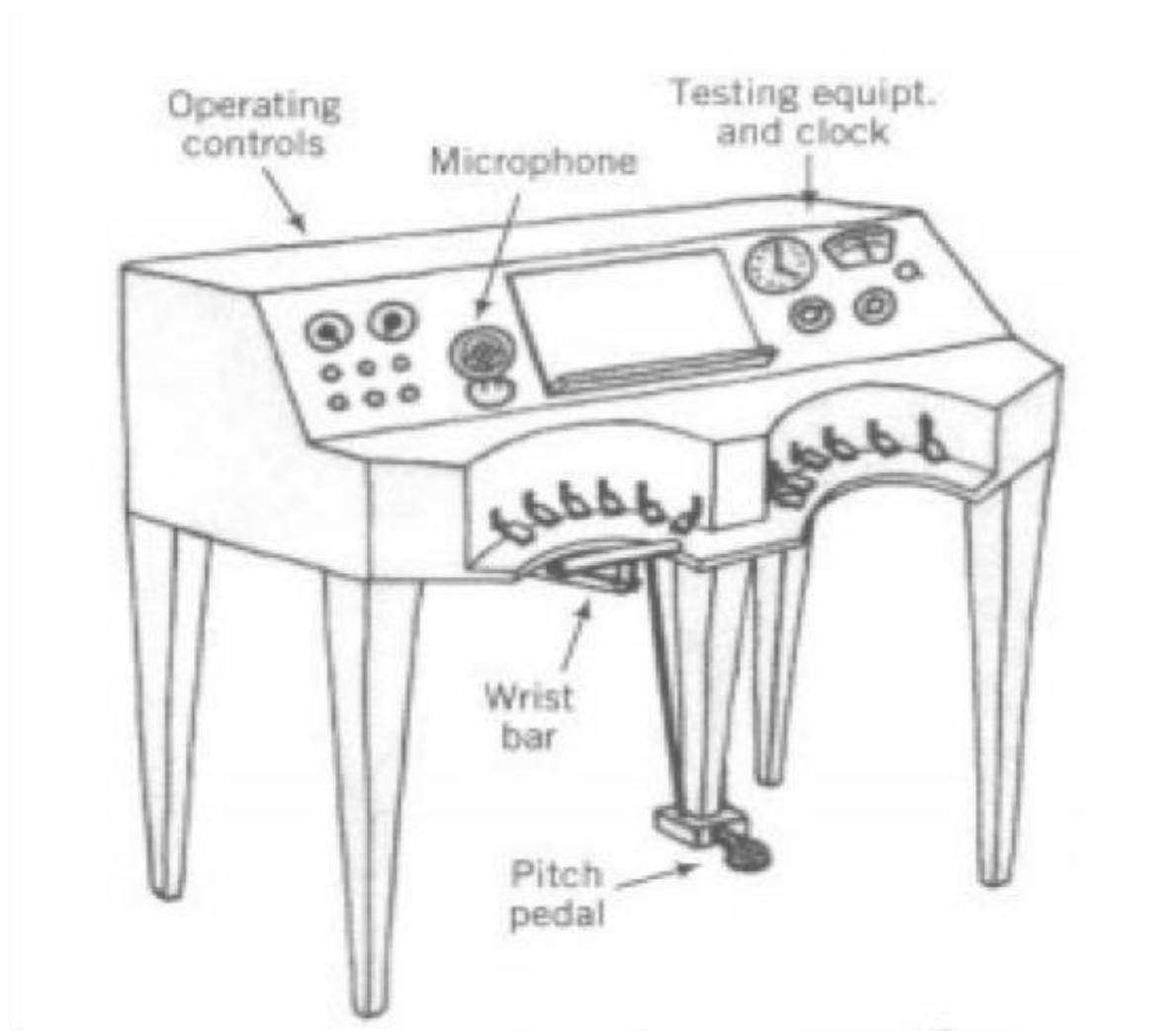
The vocoder is similar to the voder, but rather than using manual controls, it uses control currents from electrically processed speech to operate the synthesizer (Dudley and Tarnoczy, 1950).

2.2.3. The 1950s to 1960s: Automatic Speech Recognition System

In 1952, the very first ASR system was developed by Davis, Biddulph, and Balashek at Bell Telephone Laboratories. This system could recognize isolated digits of numbers from 0 to 9, and it is because early inventors and engineers who developed speech recognition systems only focused on numbers since human language is very complex (Davis, Biddulph, and Balashek, 1952).

Figure 8

Schematic for Bell Labs digit recognizer (Davis, Biddulph, and Balashek, 1952, p.638)
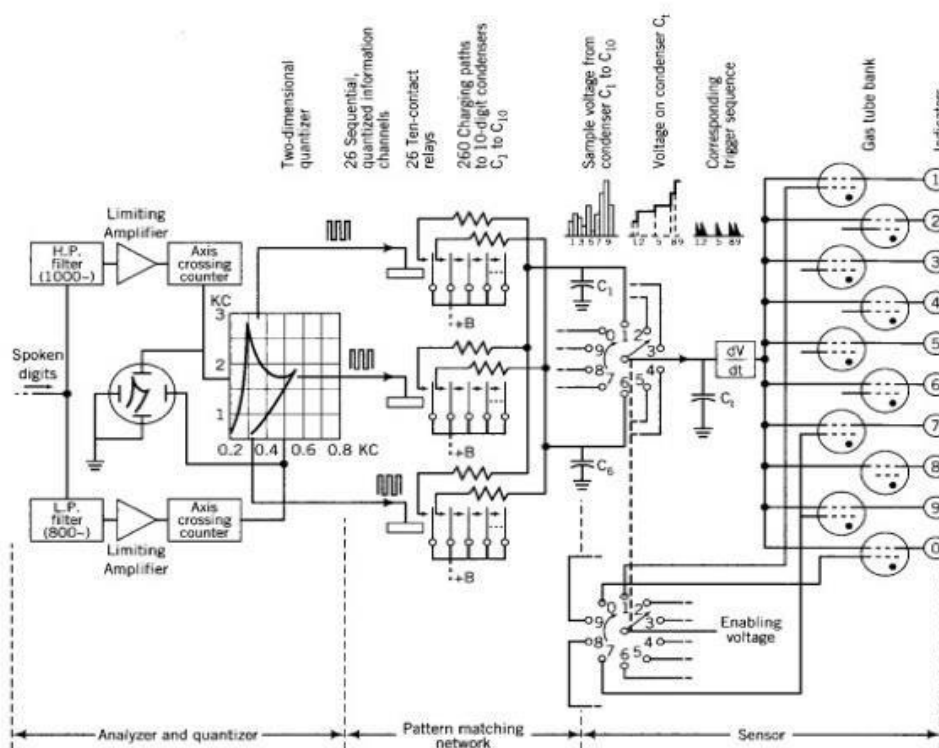
Figure 8 is a schematic for Bell Lab's digit recognizer that demonstrates how speech is produced. Three distinct sequential processes are involved in the recognition stage. First, speech is analyzed to create a formant one vs. formant two. Then, a pattern-matching network determines which pattern in 10 reference distributions most closely resembles an unknown signal distribution. Finally, this information is displayed visually by an indicator of the necessary associated circuits.

A few years later, in 1956, a phonetic typewriter device could recognize a discrete number of syllables, but it was speaker-dependent and required immense training. "These early ASR systems used template-based recognition based on pattern matching that compared the speaker's input with pre-stored acoustic templates or patterns" (Levis and Suvorov, 2013, p.1). Thus, these early-developed systems could operate only at the word level and could only recognize a small set of phonetic items and a small vocabulary. However, these systems 'performance could not identify more extensive vocabulary or complicated speech units. Therefore, other researchers attempted to build new systems that were much more complicated and could perform better. For instance, Forgie and Forgie were the first to use a computer in 1959 to construct speaker-independent recognizers.

Later on, researchers like Martin, Nelson, and Zadell (1964) and Vintsyk (1968) worked on time-normalization techniques to detect when speech starts and ends, while Reddy tried to establish a system that could recognize continuous speech in 1966 (Levis and Suvorov, 2013).

2.2.4. The 1970s to 2000s: The Hidden Markov Model

Lenny Baum from Princeton University invented the Hidden Markov Modeling approach in the 1970s, a statistical model that matches patterns and outputs a set of symbols. This model profoundly impacts ASR development "HMM can adequately analyze both the temporal and spectral variations of speech signals and can recognize and efficiently decode continuous speech input" (Levis and Suvorov, 2013, p.2). Then, in the 1980s due to the discoveries of new approaches, speech recognition systems were upgraded from only recognizing a few hundred words to several thousand words, had an expanded vocabulary, and were used in commercial applications for business and specialized industries. For instance, doctors Jin and Janet Baker founded Dragon Systems in 1982 and Speech Works in 1984, which led to over-the-phone automated speech recognition.

Moreover, due to the arrival of computers with fast processors, ordinary people finally had access to speech recognition software in the 1990s. The first consumer speech recognition product, "Dragon Dictate," was launched in 1990. Years after, Dragon NaturallySpeaking was introduced to the world, and this application could recognize continuous speech. The 2000s witnessed more development in ASR technology and its algorithms and modeling techniques. Many speech recognition technologies were installed in computers and built-in Windows Vista and Mac OS X.

These technologies allow users to give commands and help perform different tasks (Levis and Suvorov, 2013, Gold, Morgan, and Ellis, 2011, Swamy and Ramakrishnan, 2013).

2.2.5. Automatic Speech Recognition: From 2010 to 2013

This era witnessed the arrival of new technologies that paved the way for significant development and discoveries in ASR technology. Hence, cell phones and mobile phones possessed speech recognition applications that replaced virtual keyboards. For example, Google Voice Search enabled users to use their voice and search for whatever they wanted without the need to type on the keyboard. Later, in 2010, Google provided personalized recognition to the voice search, allowing users to record themselves for more accurate speech models. The arrival of Siri also marked this era. These technologies are in continuous improvements; what started from only recognizing a set of digits is now capable of hearing, identifying, and understanding your speech but also replying to it with personality as if it was a natural person (Levis and Suvorov, 2013, Gold, Morgan, and Ellis, 2011, Swamy and Ramakrishnan, 2013).

2.2.6. The Future of Speech Recognition

Progress made in speech recognition technologies in the past decade is impeccable. Nowadays, devices like computers, mobiles, and tablets can hear and understand human speech even in noisy environments, which could only mean that this technology has the power to outgrow even more. As Svendsen believed in 2003, speech recognition systems could become speech understanding where machines will recognize speech and understand what every word means. "We still lack some of the fundamental understanding of speech and language communication that can produce truly speech understanding systems, generation of natural sounding (emotional) synthetic speech from a concept, and conversational machines" (Svendsen, 2003, p.17). This is the case nowadays; different developed systems and applications, as mentioned in the previous sections (Siri, Google Voice Search, Windows Voice Recognition), can understand and respond to human

speech. In addition, Batliner et al. (2011) revealed new trends known as emotion recognition systems that can recognize human emotions from speech through facial expressions, voice tone, and gestures. Speech recognition systems have developed from their primitive origins and become fundamental aspects of humans 'daily lives. They keep every person connected with their device and can perform every command at any time and place. Speech recognition systems could play music, turn the light on or off, and perform endless tasks.

2.3. Speech Recognition Mechanisms: How Does it Work?

Even though ASR technology surfaced decades ago and has been researched several times, it remains superior to human capabilities. Therefore, this section is devoted to understanding how this technology works and its main functions and mechanisms that allow it to recognize speech and communicate with humans.

2.3.1. Speech Production

To better understand the architecture of ASR technology and its different mechanisms, it is essential to understand some fundamentals of speech production and perception, how humans produce speech, and how other sounds are articulated and characterized. First, speech starts with a chemical activity in the brain through the transmission of messages. When this neurological stage ends, brain instructions are sent to various muscles in the vocal tract (Ball and Rahilly, 2014). Then, the air pressure radiates from the mouth of the speaker. The main features of human speech production are the lungs, the glottis, and the vocal tract. Air escapes from the lungs and excites the air cavity of the mouth, and at the glottis, the vocal cords constrict the path from the lungs to the

vocal tract. To produce different sounds, the speaker uses other articulators and different techniques. Some sounds are voiced, so the air pressure from the lungs builds up behind the vocal cords, which are closed until releasing a burst of air, while other sounds are unvoiced, meaning they require the whole vocal cords to open. In addition, different articulators take part in this process, like the tongue, the jaw, and the lips changing the size and form of the oral cavity (Fletcher, 1989, Svendsen, 2003).

The following diagram displays the message's transmission through a speech production chain.

Figure 9

A Theoretical Communication View of the Speech Communication Chain (Juang and Furui, 2000, p.1144)
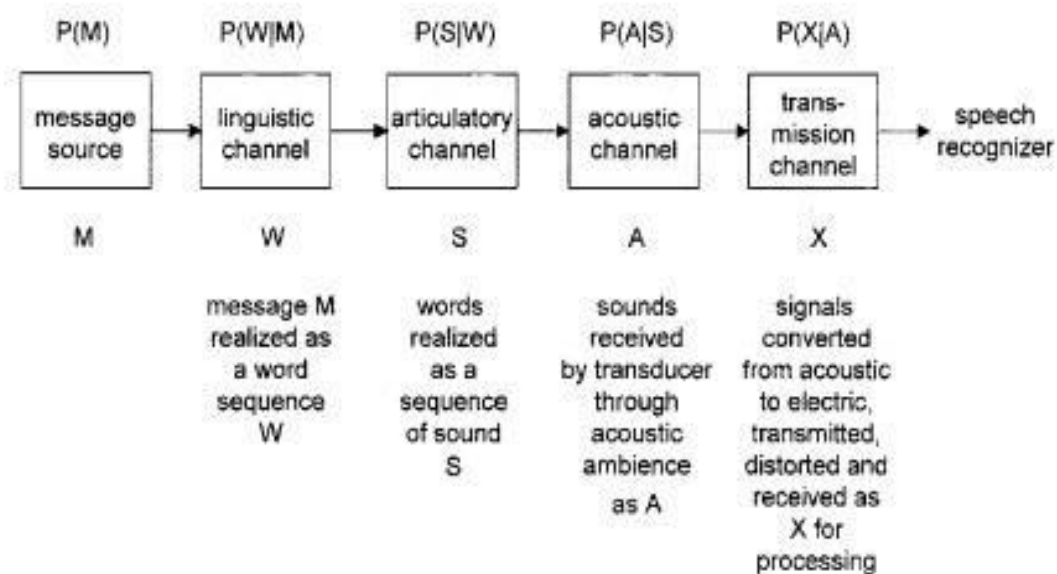
The decision to transmit an intended message M - realized as a word sequence W via a linguistic channel is made by a message source and is determined by a probability measure P. A message source decides to convey an intended message M, realized as a word sequence W through a linguistic channel, specified by a probability measure P. The sound sequence S is then released from the speaker's mouth and travels around the space as acoustic waves (Juang and Furui, 2000).

There is a considerable difference between printed text and speech because humans produce phonemes and words in a flowing, continuous speech. Characteristics of speech sounds are pretty different from printed text, and when people read something is not the same as how they speak it. These differences are due to (a) physiological differences since speech may be produced differently from one person to the other because of the different dimensions of the vocal tract, which is related to the person's head size, which may affect the pitch, (b) behavioral which is mainly related to people's speaking rate. People's accents and the way they speak and use different words differ from one person to the other according to their regional or social belonging, (c) environmental conditions intervene as well because the background noise while recording differs from one to the other according to their environment and the noises around them, and (d) phonetic context is about to the articulation of a sound which is based on the preceding and the following sound in an utterance which changes the way some sounds are articulated (Samudravijaya, n.d).

Further, Errattahi, El Hannani, and Ouahmane (2015) explained that ASR performance depends on different factors. Speaker variabilities concern how a person's voice might alter over time due to aging, illness, emotions, fatigue, and other factors. Hence, these factors may adversely

impact the ASR system's effectiveness. Another factor is the spoken language variabilities that revolve around the fact that speech may be accented due to different variations and dialects, which may be critical to the ASR system. Hence, the extensive vocabulary makes finding enough data to train the language models challenging. The main challenge for an ASR system is mismatch factors since differences in the background noise, the transmission channel, and the recording devices can generate variabilities that could reduce the accuracy of the ASR system.

The factors mentioned above may influence how speech sounds are perceived and how human speech requires precise phonetic units for recognition.

2.3.2. Automatic Speech Recognition Architecture

There needs to be an explanation to understand how this technology works and how it perceives human speech. This section provides different perspectives and descriptions of the various components and mechanisms of an ASR system and how it works since the primary goal of an ASR system is to recognize and understand spoken and acoustic information. Karpagavalli and Chandra (2016) discussed the main features of a typical speech recognition system architecture. They emphasized that it is developed with major components that allow converting the speech signal into appropriate features.

The process of converting speech utterances into words or phonemes is explained and demonstrated in Figure 10.

Figure 10

Automatic Speech Recognition Architecture (Karpagavalli and Chandra, 2016, p. 395)
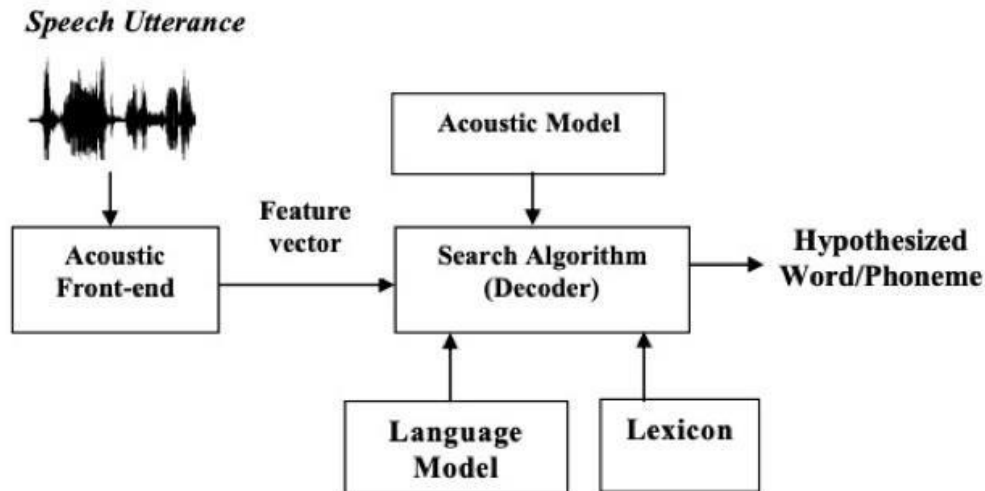
The automatic speech recognition system collects several speech parameters from the acoustic speech signal for each word or sub-word unit. The speech parameters create a pattern that identifies the word or sub-word. The operator then will read all the vocabulary words throughout the training phase, and when a word is recognized, its pattern is compared to the stored patterns (Karpagavalli and Chandra, 2016).

Further, Neri, Cucchiarini, and Strik (2002) described an ideal ASR-based CAPT system as a system that is made of a sequence of five different phases and identified these phases as follows:

    a.   Speech Recognition Phase: The ASR system consists of an engine that translates speech signals into a sequence of words based on internal phonetic and syntactic models.

b. Scoring Phase: This phase gives a global evaluation of pronunciation quality in the form of a score. The scoring happens by analyzing and comparing the spoken utterances and the references available in the system. It checks whether the utterance provided by the learner is close to the native model.

c. Error Detection Phase: The ASR system can detect the errors made by the learner when comparing the learner's produced utterance and the stored native utterance. If the learner produces a sound incorrectly, the system will identify the sound and help the learner work on the problematic sound with more practice.

d. Error Diagnosis Phase: During this phase, the ASR system identifies the error made by the learner and suggests how to improve it.

e. Feedback Phase: The last phase is specialized in giving the obtained information throughout the previous stages, presenting the overall score calculated by the ASR system algorithm.

Accordingly, Katyal, Kaur, and Gill (2014) clarified that a text message transcription of uttered words from speech signals should be the aim of an ASR system. They introduced the main components of an ASR system which are displayed in Figure 12.

Figure 11

Main Components of an Automatic Speech Recognition System (Katyal, Kaur, and Gill, 2014)
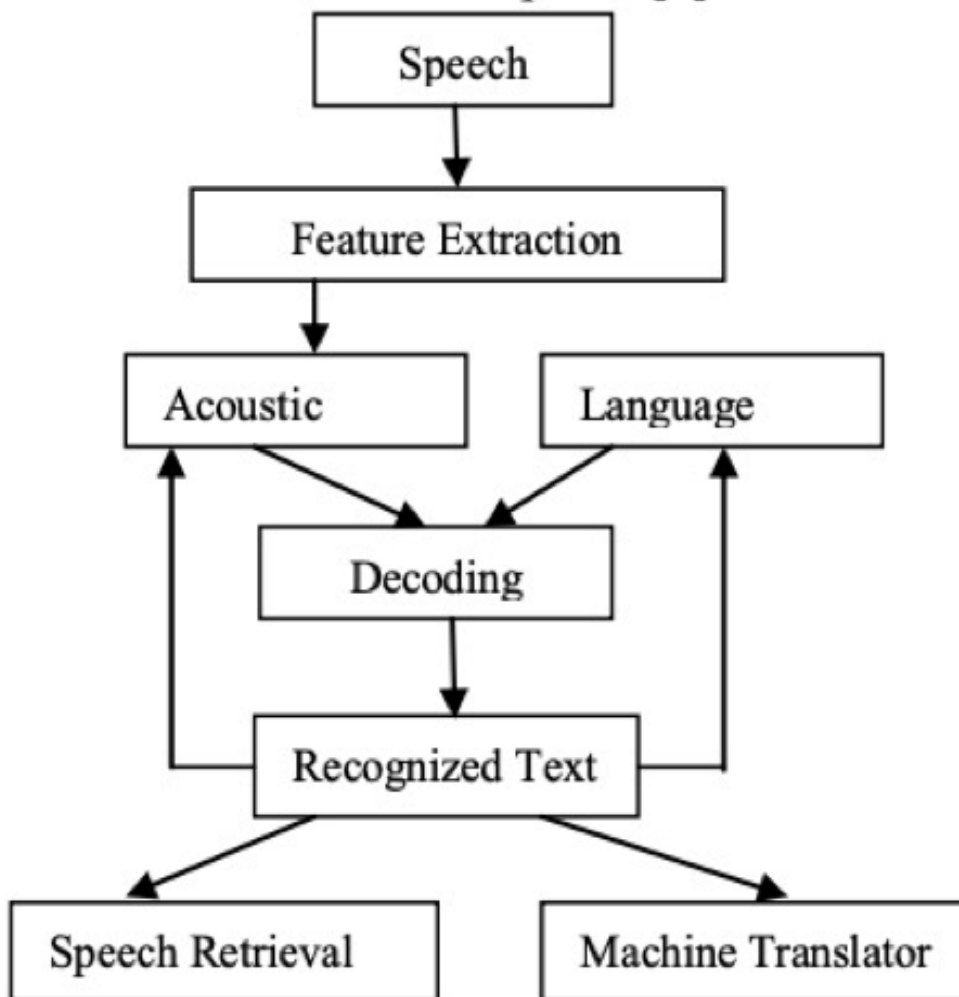
The process of turning speech into a text message is depicted in the above figure. The qualities that distinguish phonemes from one another are first separated from the speech signals. After that, the probabilities for various phonemes at various time instants are provided by the acoustic modeling. The type of phonemes and sequences that can be used in the target language are determined by language modeling. Finally, the acoustic and language models search for the

best recognition hypothesis that matches the models. This is the coding phase (Katyal, Kaur, and Gill, 2014).

Yu and Deng (2014) described in their book entitled "*Automatic Speech Recognition: A Deep Learning Approach*" that the ASR system consists of four main components that allow speech recognition and they are:

a. Feature Processing and Feature Extraction: This feature takes the input audio and then modifies it by enhancing the speech by removing the background noises and converting them into frequencies.

b. Acoustic Model: This model works as described by the authors "[…] takes an input the features generated from the feature extraction component, and generates an AM score for the variable-length feature sequence" (Yu and Deng, ibid, p.4).

c. The Language Model: This model estimates the probability of the word sequence by studying the correlation between the words produced.

d. The Hypothesis Search: This component combines the scores from the previous components AM (Acoustic Model) and LM (Language Model) and outputs the sequence of recognized words that gave higher scores.

These components and the way they work together are demonstrated in Figure 13.

Figure 12

Automatic Speech Recognition Architecture (Yu and Deng, 2014)
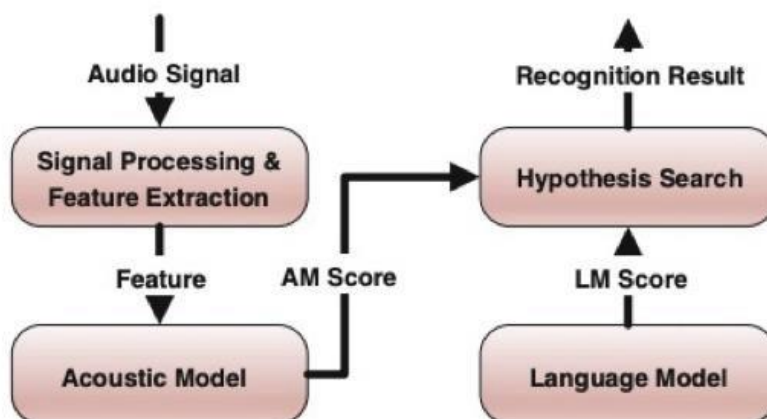
Additionally, Neri et al. (2002) argued, "[…] we specifically want to recognize non-native speech, which is characterized by deviations from native speakers. We also want to identify possible L2 pronunciation errors in that speech and to make this information immediately available and accessible to the learner" (p.121). Thus, to accomplish these requirements, ASR technology needs to be applied through a sequence of four main stages.

First, the chosen ASR-based CAPT system must be compatible with speech recognition, considering that native speech differs from non-native speech. After the recognition stage, the system must segment the spoken utterance into a series of phonemes to gather information about the characteristics of the phonemes being said by the learner. Then it must correctly align the spoken utterance with the recognized utterance to provide feedback. The learner's and native speaker's speech sounds are then compared during the analysis of the pronunciation quality stage.

The chance that the learners' realization of a sound matches the native realization of that sound is typically used to measure the distance between the two types of sounds in this analysis. The ASR system gives the learner an evaluation of the sounds produced and determines whether the input and output are sufficiently similar. The final stage is for providing feedback. In the earlier phases, scoring was done using numbers. Thus, this step is essential because the learner will only benefit from the data collected by the ASR if it is presented in a way that is both pedagogically and meaningfully relevant. It must also include intuitive feedback to help them pronounce words more clearly (Neri et al., 2002).

It is complicated to develop an ASR-based CAPT system that can help learners promote their pronunciation. It requires much work from software developers, speech technologists, and educators. Therefore, teachers and learners must be aware of its different mechanisms and components before trying to incorporate an ASR system into the classroom or use it for educational purposes. Further, they need to check whether the scoring algorithm and the feedback provided will benefit them and help them improve their pronunciation.

2.4. Automatic Speech Recognition Technologies

The previous section illuminated our knowledge of how an ASR system can recognize human speech and gave an overview of scholars 'and researchers 'different views on a typical ASR system. Thus, ASR is an evolving technology used in various domains and other purposes. This technology has many types in terms of how they treat human speech and how they are used. Hence, this part discusses the different ASR technologies and examples of available ASR applications widely used.

2.4.1. Automatic Speech Recognition Types

ASR technology has many types. Katyal, Kaur, and Gill (2014) highlighted the different types of speech recognition systems. Text-to-speech (TTS) is a type of ASR system specialized in transforming a set of words into an audio clip. This kind of system is considered to be useful for blind people who use computers and cannot type using the keyboard. However, others can use it to gain time and effort for a better computer experience. Another ASR system is known as Simple Voice Control or Command, which is a system that is designed to recognize a loss of short commands that will lead it to perform a specific action. For instance, the word "Firefox" will trigger the system to launch a new browser window. Additionally, Full Dictation systems enable the computer or phone user to read sentences or paragraphs and translate them into text. Many people find these systems or applications useful while sending emails as they also help gain time.

These systems and applications significantly benefit human life, making them more accessible and well put together. They are available at any time and place, help in different domains, have searchable text-ability and are accessible to deaf and blind people.

2.4.2. Available Automatic Speech Recognition Applications

Gold, Morgan, and Ellis (2011) spoke of other ASR systems that are available and useful in different aspects. Telephone Applications are speech recognition systems that replaced the touchtone button press to check voice mails, dial a number, or ask to call a specific contact. ASR technology allows hands-free operation and enables users to give commands without holding the phone. Further, ASR-based applications are helpful for people with disabilities since they help those who cannot use their arms and hands and people who are blind or have sight limitations.

Moreover, highly developed ASR technologies allow the dictation into applications and programs and the translation from one language to another. These different types of ASR systems offer other options for their users and make accomplishing complex tasks efficient.

There are various available tools and applications that people can download or install on their devices and use for different speech recognition usage.

- o Praat: This free software can run on various OS devices and is designed to record and analyze human speech.

- o Audacity: is free open software that runs on a wide range of OS platforms, and it can be used to record audio and edit it, cut, copy, and mix sounds.

- o Hidden Markov Model Toolkit (HTK): This application is used for building and manipulating Hidden Markov Models. It allows sophisticated services for speech analysis, HMM training, testing, and result-in analysis. Nguyen et al. used HTK to develop an ASR system in 2004.

- o Sphinx 4: This is a toolkit that provides a series of speech-recognizing tools that are written in Java programming language.

- o SCARF: It is a software toolkit that can recognize speech with the help of segmental conditional random fields.

- o VOXFORGE: It is a free speech corpus designed to transcribe speech for open-source speech recognition engines.

- o Dragon NaturallySpeaking: This software is recognized as the market leader for speech recognition software. It enables dictation (to transcribe speech into written text), text-to-

speech (to synthesize a document as an audio stream), and command input (to issue commands into the software). (Gulzar et al., 2014; Arora & Singh, 2012)

2.4.3. Most Used Automatic Speech Recognition Technologies

In addition to all the technologies mentioned above, leading ASR technologies are compatible with pronunciation practice and available for free on different devices. These technologies include Siri, Google Voice Search, and Windows Speech Recognition.

2.4.3.1. Siri. Apple developed Siri; it stands for "Intelligent Personal Assistant" and is available on both iPhone mobile phones and MacBook. This technology has many potentials to allow users to dictate messages and give commands. This program responds to humans, can answer a wide range of questions, and even hold a conversation as it can understand and speak a wide range of languages. To use Siri on the iPhone, the phone users need to press and hold the home button (the button at the bottom of the screen), and the user can hear two quick beeps. A more straightforward way is just to say, "Hey Siri," which triggers Siri to be ready to listen, understand, and perform any command or answer any type of question (Apple Support, 2020; McCrocklin, 2015).

Picture 1

Siri (Apple Virtual Assistant) (Apple Support, 2020)
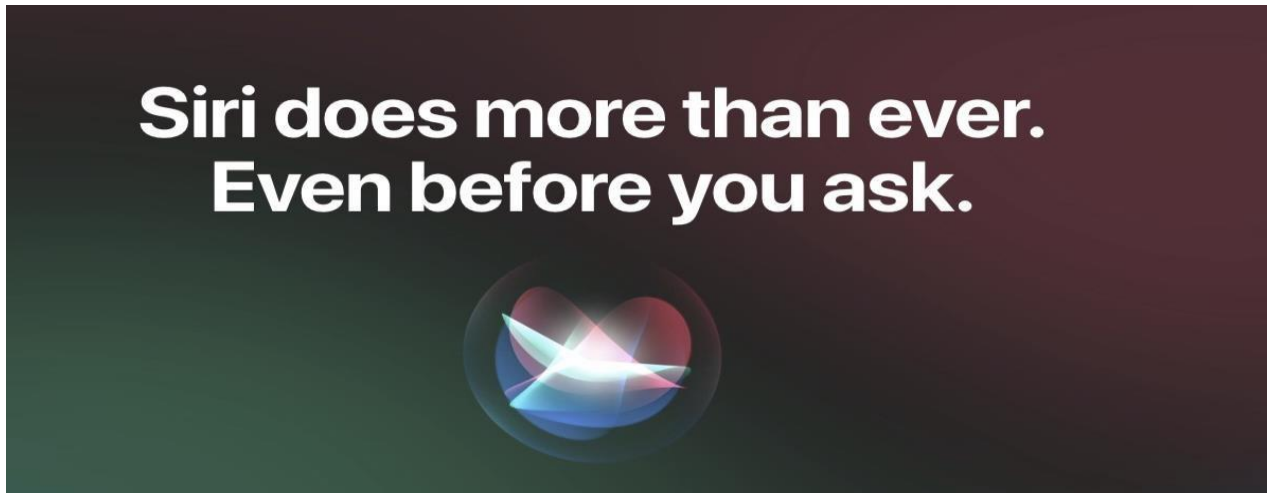
Picture 1 displays the screen that appears when the user speaks to Siri.

2.4.3.2. Google Voice Search. Google Voice Search technology is considered the Siri version for Android devices which is available on different mobile devices and PCs and works as an assistant that can perform any command and answer any question. It can understand and speak a variety of languages. This technology has a feature that allows users to modify this technology set to detect and dictate in five different languages at a time. This assistant can be downloaded on Android-powered mobile phones at the Play Store, or to use Google Voice Search without the application, go to **www.google.com**. Users can click on the microphone icon and enable the microphone to get started with voice searches (Google Help, n.d.; McCrocklin, 2015).

Picture 2

Google Voice Search (Google Voice, n.d.)

Google Voice screen is shown in picture 2.

2.4.3.3. Windows Speech Recognition. Windows Speech Recognition is a technology available on PCs that support Windows operating systems. It is available in many languages (English, French, Spanish, German, Japanese, Simplified Chinese, and Traditional Chinese). This technology allows users to control their computers with their voice since it can respond to everything a user says and perform any command. In addition, it is helpful for dictation into other programs installed on the PC (e.g., dictating into Microsoft Word). To use Speech Recognition, users need to set up their computers for Windows Speech Recognition by setting up their microphones, learning how to talk to a computer, and training the computer to understand its user's speech. This will help the user gain time and effort and stay connected (Microsoft Support, n.d.).

Picture 3

Windows Speech Recognition tab (McCrocklin, 2015)

Picture 3 is a screenshot of a PC screen with an open Windows Speech Recognition tab.

Further, McCrocklin (2015) provided a detailed description of the technologies mentioned above (Siri, Google Voice, Windows Speech Recognition), their benefit, drawbacks, and the level of recognition of each technology are presented in Table 1.

Table 1

Benefits, Drawbacks, and Recognition Levels of ASR (McCrocklin, 2015)

ASR Technology	Benefit	Drawback	Recognition Level
Siri	Unlike other software and programs, talking to Siri feels more natural, like conversing.	Siri cannot help submitting homework but allows dictating emails and other stuff.	This technology consists of a high level of recognition.
Google Voice Search	This technology works best with short utterances and helps work with minimal pairs.	GVS allows learners to copy search results individually into other documents.	It has a reasonably high level of recognition.
Windows Speech Recognition	This technology lets learners dictate into a Word document and save the file.	This technology cannot recognize words uttered in isolation, but sentences put in context.	Offers lower rates of recognition.

The benefits, drawbacks, and level of recognition of each ASR technology are listed in the table above. Given its highly developed level of voice recognition and the ability to give users the impression that they are conversing naturally with a human being, Siri stands out among all the

other ASR tools as the best one. However, both Google Voice Search (GVS) and Windows Speech Recognition (WSR) offer many options and can facilitate completing different tasks and commands.

This section shed light on the available ASR software, systems, and applications available online and offline for free for users worldwide. Although some developed ASR systems are accessible, many ASR technologies offer plenty of options but are very expensive; this is not appropriate for learners who do not have a stable income.

2.5. Implementing Automatic Speech Recognition Technology in Education

The previous sections revealed the importance of ASR technologies and their benefits on human life, such as facilitating different tasks and performing various commands. Many studies have investigated the use of ASR technology and have employed web-based, software, and phone applications in EFL settings. The outcomes of the studies revealed that ASR technology helps assist pronunciation learning. (Wallace, 2015, McCrocklin, 2016, Sidgi & Shaari, 2017). However, educators may encounter challenges if they try to apply ASR technology as an educational tool to improve pedagogical outcomes. Neri (2002) argued that many challenges arise when implementing such a system whose utility has not yet been determined. We need to consider the difficulties of finding participants, particularly adult participants who are prepared to train regularly.

Using a technological system and making it available for learners is considered a challenge, let alone the knowledge and research available on these systems 'accuracy, the algorithms of scoring, and assessing learners 'performance which is still ambiguous and little has been investigated. " […] if it is true that ASR technology is essential to provide automatic feedback, state-of-the-art ASR is also known to suffer from limitations which can result in the occasional provision of erroneous feedback to the learner, possibly compromising the learning process and outcome" (Neri et al., 2008, p.79). In addition, no data has been collected on the effectiveness of the feedback provided by these ASR systems.

2.6. Aspects and Principles of Pronunciation Teaching

This section is about the different aspects of pronunciation that are considered problematic and may create difficulties and challenges for teachers and learners. Hence, Teachers should know some principles and aspects before any pronunciation training. At first, learners must be able to produce many sentences independently. Training the language inside the classroom encourages learners to have meaningful conversations later. Further, learners must receive a corrective type of feedback. Classroom feedback is given from the teacher to the learners by correcting errors made while speaking. Teachers intervene soon enough to stop the mistakes from occurring and being repeated several times. However, their intervention rate is low and does not happen too often, not discouraging learners from speaking and harming their self-confidence.

In addition, peers and classmates may sometimes intervene and try to correct each other or simply point out their errors, which is something they cannot do when working with ASR

technology (Kenworthy, 1987). Another principle that should be considered is exposing learners to different native inputs (Celce-Murcia and Goodwin, 1991).

Furthermore, teachers should emphasize the importance of learning the prosodic aspects of the language. When teaching pronunciation, teachers focus on teaching the segmental features (all the different sounds and how they are articulated) and neglect suprasegmental, which are vital to successful pronunciation. Therefore, teachers must encourage learners to use these prosodic features when speaking (intonation, pitch, duration, stress) since they enhance meaning and add style and emotion to learners 'speech (Chun, 1988). Laroy (2013) also emphasized the importance of making learners feel at ease when learning a language. They may feel less comfortable experimenting with new sounds if required to produce sounds that do not exist in their mother tongue. Teachers should, therefore, only correct students 'errors when necessary and refrain from providing negative feedback.

Moreover, Witt and Young (2000) pointed out these aspects of pronunciation:
o Pronunciation teaching requires the sole attention of the teacher to only one student at a time. Teachers should pay attention to learners 'errors and intervene at the right time, but overcrowded classrooms create a challenge for teachers.
o Learning pronunciation involves a lot of repetition, which may sometimes be tiring for the teacher and requires a lot of patience, effort, and time. o Testing learners ' achievements and the oral outcome could also create a challenge for teachers and is a time-consuming and subjective task.

In light of what has been said, ASR systems can be a great complement and an additional aid to the teacher to help save time and effort.

Building an ASR system requires knowledge from many linguistics, computer science, signal processing, statistics, physiology, and psychology disciplines. Further, every system should be tested before presenting it to the world. Therefore, this section aims to answer questions related to how ASR technology can be applied in the language classroom to teach and assess pronunciation and the challenges that may be encountered by both teachers and learners during that process.

2.7. Automatic Speech Recognition and Pronunciation

Even when using ASR technology to train pronunciation, the teacher is the primary source of knowledge in the classroom. They are responsible for teaching the different aspects of the language (segmental and suprasegmental features). Then the ASR system may work as an additional aid to enable the learners to practice the language and apply what they have learned (Eskenazi, 1999).

This part is about different scholars 'and researchers 'perspectives on the effectiveness of using ASR technology in pronunciation training. Thus, ASR technology may be used since it can allow learners to recognize accented or mispronounced speech and provide meaningful feedback on their pronunciation (Neri, Cuchiarini, Strik, 2003).

Further, Mccrocklin (2015) proved that incorporating ASR technologies into the classroom in teaching segmental features allows the learners to receive feedback on the spelling of different

sounds. In addition, the ASR system can be used as a follow-up to classroom work and can be used at home as well. ASR technology also encourages autonomous pronunciation learning. Moreover, ASR technology offers many advantages that help learners store their progress and document each student's problem in a log file. Further, ASR technology is a great tool to increase students 'self-confidence and encourage them to speak without embarrassment or fear of being judged by their classmates (Elimat and Abuseileek, 2014).

ASR systems may be a great tool to incorporate into the language classroom because they offer endless task-based speaking activities that can be used as practical interaction tools in games or role-plays. Also, these kinds of activities make the learning experience more realistic, fun, and rewarding because of their scoring and feedback options. ASR technologies make learners more autonomous to practice pronunciation independently at any time and place (Purushotma, 2005; Wachowicz and Scott, 1999). They also help learners enhance their sound production and promote their pronunciation because of the instant feedback provided by the system. However, teachers should not become passive participants in the classroom when using an ASR system. Instead, they should remain active in getting learners to speak and help them get exposed to more native input. They must work as a guide throughout the process because even if these systems can assess learners 'progress, the teacher plays a crucial role in the language classroom (Alsuhaim, 2017). Additionally, ASR technologies can facilitate learning phonology and help learners work on their accents since these technologies provide formative assessment and accurate feedback. Besides, it enables learners to communicate in the classroom, practice the target language, and then analyze their pronunciation and comprehensibility level (Carrier, 2017).

Overall, using ASR technology is an excellent tool for English learners to help them learn more effectively and autonomously and improve their pronunciation. Additionally, Li et al. (2017) spoke of how pronunciation is an area of language that requires one-to-one interaction between the teacher and the learner, which is sometimes impossible, especially in overcrowded classrooms. Therefore, ASR devices and systems allow one-to-one interaction as they perform as personal teachers who correct learners 'mistakes and give visible feedback.

2.8. Automatic Speech Recognition Accuracy in Pronunciation Assessment

ASR technologies significantly impact human life and can benefit the educational system. Since the invention of computers decades ago, CALL systems have taken over many aspects of teaching, learning, and assessing language. Later, CAPT systems were used to teach the neglected area of the language, i.e., pronunciation, and provided many tools to train learners 'pronunciation. Thus, how can ASR systems be applied to assess and improve learners 'pronunciation?

ASR technologies can process various voices, compare learners 'produced sounds with the model, assess their pronunciation, and provide feedback. To analyze, give a score, and evaluate pronunciation, the ASR system has four main components, as described by Kang and Ginther (2017):

o The acoustic model is the modeled version of the sound. Words uttered by the learner are analyzed by the system and sliced into a sequence of phonemes. The model then identifies the matching words from several options.

o Language model also known as the vocabulary and grammar model, represents the sequence of words that the speaker might utter. The ASR system anticipates the words to improve recognition accuracy and speed.

o The pronunciation dictionary decides the pronunciation of words. It lists the different pronunciations of the words in the language model. Many words in English may have other ways to be pronounced, and one word may be pronounced in more than one way, which is decided by the system. (Eg. In the word Schedule, the "ch" may be pronounced as /k/ or /sh/) Here the pronunciation of the word depends on the intended use of the system.

o The scoring model is responsible for giving an overall score of learners 'pronunciations "This score-generation algorithm "learns" how to score pronunciation like expert human judges by examining expert ratings and maximizing predictions of those ratings by interactive combinations of features and weights." (Kang and Ginther, 2017, p. 139) This scoring system works based on two main features. The system measures the sound quality and everything related to prosody features, stress, and intonation. In addition, there is a statistical model that handles more complex data and predicts each learner's score.

Hence, ASR technology compares learners 'produced speech and how each sound is uttered compared to the reference available on the system, which is of a proficient language speaker. However, learners may possess an acceptable pronunciation that the system may consider weak and still needs practice. So, it is all based on whether this reference is reliable for judging learners ' pronunciation ability and level of comprehensibility.

2.9. Challenges of Applying Automatic Speech Recognition in the Language Classroom

It is a difficult task to integrate ASR technology into the language classroom successfully, and teachers willing to incorporate these kinds of technologies must be aware that there are challenges that come along with this process. Arora and Singh (2012) highlighted some difficulties that may be encountered when using ASR technology:

o   Human comprehension of speech creates a considerable challenge when considering how an ASR system can predict and comprehend human speech.

o   Spoken language is not equal to written language, the former is two-way communication, whereas the latter is one-way. People tend to make errors while speaking, which include tongue slips and repetitions, but written language is straightforward.

o   Noise interferes when using ASR systems, other background noises can be heard and detected by the system, such as the clock's ticking sound and speakers 'voices in the background which create challenges for the system only to catch the speaker's speech.

o   ASR technology is a system that only recognizes human speech and not physical things such as body gestures, waving hands, or moving eyes. Thus, this problem must be addressed since body language plays a crucial role in communication, and working on it can improve human-machine communication.

o   Speakers differ based on their voices and other variations that make one speaker different. Speakers might have different speaking styles and unique ways of pronouncing words. Also, the speaker's gender determines the speaker's pitch since females have shorter vocal tracts than males. In addition, speakers might belong to different regional and social

backgrounds that define their dialect. Thus, these differences need to be considered before using any ASR technology.

ASR technology can be applied in the language classroom to train and promote learners ' pronunciation. Hincks (2003) spoke of some problems that may hinder the use of ASR in the language classroom. At first, ASR systems have a mathematical nature since they provide learners with numerical scores, which are given as feedback that measures learners 'accuracy of pronunciation of phonemes in comparison to the target phonemes available on the system as a model. However, some believe this kind of scoring is not accurate enough to determine the nature of learners 'pronunciation. Hincks (2013) argued that it is impossible to determine how the signals have diverged from the model with current technology. Therefore, feedback is not corrective or constructive but only signal evaluation. In addition, the speaker's age and gender create an issue that must be considered. The speaker's age is crucial in determining how the ASR system recognizes speech of high frequencies for only adult learners and not children (Probst, Ke, and Eskenazi, 2002, Hincks, 2003).

Moreover, Mccrocklin (2015) pointed out challenges that could stand in the way of learners and teachers when using ASR technology in the classroom. Teachers need to ensure that they can allow the learners to have access to ASR technology through school resources. Using ASR technology requires computers only available in the school computer lab. Working in the school computer lab may make students uncomfortable because others also work there. ASR is a sensitive

technology; when used in a crowded place, it can catch background noise which may affect the overall score and the feedback provided. This can overwhelm learners and make them doubt their improvements and pronunciation abilities. "While it is useful to have a program show students where they are making pronunciation errors, the great amount of feedback can be overwhelming, practically in the first practice with the program" (McCrocklin, 2015, p.131).

ASR technologies can be a powerful tool to promote learners 'pronunciation and help them practice the spoken form of the language.

Conclusion

Several significant advances in automatic speech technologies have occurred over time. These advances make it possible for users to translate speech into a set of words and vice-versa. ASR technologies provide one-on-one interactions and can give corrective feedback. Hence, they may be employed in education to teach particular skills and assess learners' improvements. Further, it is an excellent tool that can also be used to teach and assess pronunciation However, it is crucial first to know what kind of technology to adopt and its features.

Chapter Three

Research Methodology

Chapter Three: Research Methodology

Chapter Three: Research Methodology

Introduction

The previous chapters reviewed the literature related to the study and how pronunciation is taught and assessed. Further, it considered ways to incorporate new technological tools into the classroom to teach pronunciation and provide corrective feedback. Research is a systematic process guided by hypotheses and a purposive design; it starts with a curious mind that asks questions, locates problems, and ends with possible solutions. Hence, this chapter discusses how the data was collected and analyzed. This chapter introduces and justifies the methodology adapted to test the effectiveness of an ASR-based CAPT technological tool in enhancing EFL learners ' overall pronunciation. Based on the main aim of this study, it provides an overview of the research method, the research design, the population from which the sample has been selected, the Automatic Speech Recognition ASR technology used, and the data collection techniques as well as the analysis procedure employed.

3.1. Choice of the Method

Practicing pronunciation requires considerable time, constant supervision, and unlimited feedback. The latter should be provided systematically and consistently in a way that should be perceived as clearly corrective. Besides, it should be followed by time and opportunities for students to repair and correct their pronunciation (El Tatawy, 2002). This process is considered exhausting and time-consuming for teachers, so they have become more interested in Computer Assisted Pronunciation Training (CAPT) technologies. These technologies can teach pronunciation aspects, assess learners 'progress, and provide feedback in a private and stress-free

environment. The most advanced CAPT systems include ASR technology which recognizes and analyses human speech (Neri, 2002).

Further, several studies revealed that learners find using technology motivating and fun (Blake, 2008; Carrier, 2017; Chuang, 2017). Thus, the main aim of this study is to point out that incorporating advanced technological tools into the language classroom may lead to significant outcomes. Moreover, using ASR technology to teach pronunciation has been proven in many studies to be motivating and efficient (Gilakjani & Rahimy, 2019; McCrocklin, 2019; Tsai, 2019). However, teachers and learners must set realistic goals before bringing any additional training into the classroom. ASR technologies may help learners improve their pronunciation but not rationally change it. Accordingly, the main focus of this study is to train learners 'pronunciation to make them have a "comfortably intelligible" pronunciation rather than an accent-free one (Celce-Mercia, Brinton, and Goodwin, 1996; Kenworthy, 1987; Morley, 1991; Munro & Derwing, 1995).

The present study is an investigation that revolves around testing automatic speech recognition [ASR] tools to promote learners 'pronunciation. Cohen, Manion & Morrison (2017) argued that a research method is used in educational research to gather data for inference and interpretation, explanation, and prediction. It means that the method is used to discuss and examine the problem. "As researchers, we have to devise for ourselves a research process that serves our purpose best, one that helps us more than any other to answer our research question" (Crotty, 1998, p. 216). Hence, the nature of this research made it necessary to conduct a quasi-experimental design to explore the effectiveness of the chosen tool. O'Leary, Z. (2017) states that the quantitative

approach aims to test specific hypotheses following the scientific method. The latter allows using the power of numbers, objectivity, and logic to present accuracy.

In contrast, qualitative research mainly explores aspects of human nature and behaviors that cannot be translated into numbers. Hence, choosing the mixed method enabled me to gather qualitative and quantitative data since "mixed methods research provides more comprehensive evidence for studying a research problem than quantitative or qualitative research alone" (Creswell & Plano Clark, 2007, p. 9). Further, Denscombe (2017) confirmed that blending the two approaches can "increase the accuracy of data; provide a complete picture of the phenomenon under study than would be yielded by a single approach, thereby overcoming the weaknesses and biases of single approaches" (p. 272). Thus, I opted for a mixed method to collect and analyze data since each data collection method addresses a different type of informant in the research. Accordingly, the quantitative data were adapted to evaluate the usefulness of automatic speech recognition [ASR] technology regarding pronunciation improvement of EFL learners and the accuracy of the Application Programming Interface [API] for providing assessment and corrective feedback. The qualitative data were used to give an insight into students' and teachers 'attitudes and thoughts on ASR usage in teaching and learning pronunciation.

Further, because of quarantine and the precautions Algerian universities took to contain the virus, the present study adopted some characteristics of internet-based research. The latter was defined by Buchanan & Zimmer (2012) as the approach that enables the researcher to utilize online tools to gather data using the internet. In this regard, I had several face-to-face meetings with the participants. However, the necessary data were collected via online tools such as websites, surveys,

and social media groups. Hence, this study reflects the effectiveness of introducing new technological tools into the language classroom and integrating e-learning to improve learners ' listening and speaking skills.

The following part includes a thorough description and explanation of the experiment and the online tools used in this study. Additionally, it consists of an overview of the overall population and sample, the study schedule and courses, and then the data collection procedures.

3.2. Content of the Experiment

This study was conducted at Mohamed Khider University of Biskra, specifically in the English Department's Faculty of Letters and Foreign Languages. It offers a degree of Bachelor of Arts, Master's, and Doctorate courses within the framework of the LMD system in English Language and Literature. Further, English pronunciation is taught in two different modules, Oral Expression module, and the Phonetics module. The former focuses on improving learners ' speaking and communication skills as it allows them to speak and participate in engaging activities such as role-plays, debates, and presentations. In the latter, pronunciation is taught through learning and practicing the target language sounds and participating in phonetic transcription activities. Hence, the module of Oral Expression with EFL first-year LMD students was targeted for selecting the sample because it allowed me to take an overview of learners 'pronunciation based on their speech and communicative skills. Since I opted for the quasi-experimental design, the department administration already formed groups that were selected for the observation. Then, the research was conducted with randomly created groups for the treatment and training sessions. Two

groups were included in this experiment, an experimental group and a control group. Quantitative data were collected from a pre-test and a post-test which are discussed in detail in upcoming parts. The use of a control group in this study is to give more evidence that pronunciation improvements made in the experimental group are due to the treatment. This research went through a set of phases to collect and analyze data. To select the sample, I attended Oral expression sessions of the target population (14 sessions since there are 14 groups in total). After selecting the sample, the researcher pre-tested the participants who agreed to participate. Then, a training period was launched to enable the participants to practice their pronunciation using an ASR-based technological website. A questionnaire was administered to teachers, and a post-test was conducted after the intervention; the experimental group participants took part in an interview to answer questions related to their opinions on the training. All of the data were collected, analyzed, and interpreted.

3.3. Sampling and Population

First-year learners at Mohamed Khider Biskra University who study English as a foreign language at the university's English department present the study's population. The reason behind this selection of population was because of many reasons. Their mother tongue is Algerian Arabic, and they studied English for seven years throughout their education journey (four years in middle school and three years in secondary school). However, at the university level, they are exposed to new modules and aspects of the language that they have never seen before. Accordingly, these students are millennials and belong to the e-generation. They have been accustomed to using technology daily in every aspect of their lives. However, teachers of the targeted university rely

only on conventional traditional learning, and technological tools are rarely used. As a result, these learners may find this teaching method tedious and tiring and feel they need to be more encouraged and motivated to participate in the learning process.

Moreover, pronunciation is taught conventionally, the learners have no exposure to native input, and teachers 'feedback is limited. This may hinder them from improving their speaking skills and practicing the spoken language. In addition, these learners are expected to have good communication skills and unintelligible pronunciation since they are language learners. However, they only practice their spoken language in the Oral Expression module, which is insufficient.

3.3.1. *Sample and Sampling Techniques*

The researcher attended fourteen Oral Expression sessions with fourteen different groups. The study sample consisted of twenty (24) first-year students in the second semester of the academic year 2021/2022. Sampling is the process of selecting a population group for a research study. The more randomly assigned the participants are, the more representative the sample would be and the more generalizable the findings would be (O'Leary, Z., 2017). The study was conducted during a pandemic. Therefore, the researcher could not take the entire population as a sample or randomly select a sample and assign its members to an experimental and a control group. Hence, only participants who wanted to volunteer and participate in the training were selected. Selecting volunteers helps avoid withdrawals during the training. Learners were expected to communicate with their classmates and teachers in their first year using English only. Thus, the researcher aimed to allow them to get comfortable with their first year as language learners and give them a

background in the basics of English phonetics. So, they had their regular classes in the first semester. In the second semester, they participated in the training period to help them improve their pronunciation.

Further, this study aimed to enable students to explore the benefits of technology and its ability to improve their skills. The experiment was an additional instruction to practice English's different sounds and utterances and improve their overall speaking skills and fluency. To select the sample, the researcher attended Oral expression sessions with the targeted population. After each lecture, the purpose of the study was explained to the learners. Then, the learners were asked if they wanted to participate in the study. Hence, a new non-probability sampling design called "Voluntary Sampling." Was implemented. The latter is a sampling technique that allows selecting from the whole population of potential participants who are willing to participate in the study and qualified to participate. After choosing the population (first-year LMD students), the researcher explained the entire study process and the participants 'roles. Only participants who are committed, have serious intentions, and are willing to participate in the whole training period without dropping out were chosen. Elder (2009) spoke of probability and non-probability sampling: "The main difference between a pure volunteer sample and a probability sample of volunteers is that, in the former case, volunteers make all the effort; no sampling frame is used." Hence, the learners interested in our training course can volunteer and are motivated enough to accept participating in the study. Afterward, the participants confirmed their participation and were given an information card to fill in their contact details (See Appendix B).

3.3.2. Study Groups

The experimental and control groups consisted of twelve students each which were randomly divided. The experimental group had to confirm participation by signing an informed consent outlining the training procedure. Both groups received an explanation about their involvement in this study and that it would include two tests, one before and one after the training period, in which they would practice their pronunciation using ASR technology. Further, the researcher explained to the students that they must use their smartphones or PCs during the training. Then, the groups were randomly assigned into control and experimental groups after receiving their agreement to participate in the tests and the training.

Before starting the pronunciation tests and the training period, the researcher collected information about the participants that was needed for the study. After each observation session, information cards were distributed and learners were asked to fill them in upon their approval to volunteer and participate. The information displayed in Table 2 was collected from the volunteers. Each participant identified themselves: gender, age, years studying English, and their English mark in the BAC exam. Then, they were asked whether English was their first choice as a university major to understand their motivation behind learning English.

Further, Table 2 includes information that investigates the aspect of the language that participants are more eager to improve and checks whether pronunciation is that aspect. Since the participants are going to be asked to train using an online technological website on their own, we

wanted to discover participants 'ICT competencies (i.e., mastery of the usage of ICT tools, connectivity, and access to technological means).

Table 2

Participants ʹInformation

Participant	Gender	Age	Years Studying English	English Mark in the BAC Exam	English as a first choice	The aspect of language desired to improve	Technical Skills
Mila	F	18	8	18	No	Speaking	Average
Alexander	M	20	6	15	Yes	Writing	Good
Ella	F	18	8	15.5	Yes	Speaking	Good
Scarlett	F	18	8	16.5	Yes	Speaking	Average
Camilla	F	18	6	13	No	Speaking	Good
Penelope	F	19	7	16	Yes	Speaking/ Listening	Good
Aria	F	18	7	19	Yes	Speaking/Listening	Good
Owen	M	18	7	13	Yes	Speaking	Good
Levi	F	18	5	15	Yes	Speaking/Writing	Very Good
Luna	F	19	8	17	Yes	Speaking	Good
Olivia	F	18	6	17	Yes	Speaking	Good
Amelia	F	19	7	16	Yes	Speaking/ Writing	Good
Isabella	F	20	8	19	Yes	Speaking	Very Good
Noah	M	19	7	13	No	Speaking	Average
Ava	F	20	8	15	No	Speaking/ Writing	Average
Liam	F	20	8	14	No	Speaking	Good
Sophia	F	18	6	16	Yes	Speaking	Very Good
Madison	F	18	6	17	Yes	Speaking/ Listening	Good
Lucas	F	18	6	16.5	Yes	Speaking	Very Good
Mia	F	18	6	18.5	Yes	Speaking/ Writing	Good
Alice	F	23	11	14.5	No	Speaking	Average
Lorrie	M	18	7	16	No	Speaking	Good
Grace	F	18	9	15	Yes	Speaking/Listening	Very Good
Ezra	M	18	6	12	Yes	Writing	Excellent

Table 2 involves each participant's identifier, gender, age, years studying English, their English mark in the BAC exam, if English was their first choice as a university major, the aspect of the language they want to improve, and their technical skills.

3.4. Outline of the Study

The study was conducted during the second semester of 2021/2022. The study adopted an internet-based mixed-methods quasi-experimental design that comprised four phases: a sample selection phase, a pre-experimental phase, an experimental phase, and a post-experimental phase. Each stage included a set of steps to gather the necessary data. The pre-experimental stage included a questionnaire that was addressed to Oral expression and Phonetics teachers to elicit their perceptions and opinions concerning using ASR as an effective tool to promote learners' pronunciation. Then, the sample selection phase allowed me to attend classes and collect information about the population, explain the whole procedure to the population, and then receive approval. Before the experiment, the selected sample was randomly assigned to experimental and control groups. The division of the groups was done haphazardly and without any intention. All the results were then analyzed and interpreted.

On the one hand, a session was scheduled to meet with the sample and explain the content and use of the ASR website to clear any potential doubts or difficulties the participants may have. Later on, the experimental phase investigated the effectiveness of using an ASR tool to improve the participants' pronunciation and speaking proficiency compared to traditional teaching methods in conventional classrooms. During this phase, their teachers taught the experimental and control

groups using regular instruction in all the modules. However, the students in the experimental group participated in a discrete training (separated from their main courses) designed to teach different aspects of pronunciation to promote learners 'speaking skills. In addition, this training allowed them to learn how to use the automatic speech recognition [ASR] website to practice the spoken language.

On the other hand, the control group did their lectures as usual and did not take part in any course or training. For this purpose, both groups participated in a pre-test. Then a post-test was administered to the experimental and control groups directly after the training program was finished. It was used to check the improvements in the participants 'pronunciation, the effectiveness of the training, and the instructional application used. Moreover, two WhatsApp private groups were created to keep up with the participants, help them if they faced any issues, inform them of the schedule of the tests and the meetings, and enable them to share screenshots and video shots of their training. Last, the post-experimental phase included a post-training interview to gather the experimental group participants 'thoughts on training their pronunciation using the ASR website. All the results were then analyzed and interpreted.

Table 3

Study Outline

Study phases	Steps	Time allocated	Participants
Sample Selection Phase	- Attending classes - Sample selection - Information card and consent form	4 Weeks	The whole population (1st-year LMD students)
Pre-experimental Phase	- Pre-test - Teachers' Questionnaire	2 Weeks	Control group Experimental group Oral expression and Phonetics teachers
Experimental Phase	- Pronunciation training using an ASR website - WhatsApp Groups	4 Weeks	Experimental Group
Post-experimental Phase	- Post-test - Post-training Interview	3 Weeks	Control group Experimental group

Table 3 summarizes the study outline and demonstrates all of the phases and stages of the study.

3.4.1. Sample Selection Phase

The researcher attended Oral Expression sessions with first-year EFL learners, and the reason behind this was mainly to select the sample. This step allowed me to present the study to the entire population. This procedure took time, but explaining the entire process to the learners to convince them to participate to avoid dropouts during the experiment phase was necessary. This phase included more steps that are presented in the previous section.

3.4.1.1. Information Card and Consent Form. Before the start of the intervention and after completing the observation phase, an introductory session took place. All the students who agreed to participate attended the session. Because of a shortage of time and because students were occupied with their busy university schedules, the session took 15 minutes.

It was enough for me to explain the participants 'role in the study, a presentation of an information card that included necessary information related to the study, and a consent form outlining the procedure. The students showed great interest and were highly motivated to participate. They filled out the information card during the first session and were asked to take the consent form with them to read it carefully. The consent form outlines the entire procedure and shows participants 'roles and what they are expected to do. Participants read the consent form carefully and made sure they understood all of the details that were highlighted. After they signed it, it was collected the following day (See Appendix C).

3.4.2. Pre-Experimental Phase

This phase involves a pre-questionnaire addressed to teachers and a pre-test before introducing the intervention and the training procedure. Accordingly, both the control group and the experimental group undertook a pre-test before the pronunciation training and a post-test after to verify the effectiveness of the ASR application in improving EFL learners 'pronunciation and speaking skills. Then, the results of the pre-test and the post-test are compared.

3.4.2.1. Teachers 'Questionnaire. To get teachers 'insights on using technology to teach and assess pronunciation, the researcher designed a semi-structured questionnaire and teachers of oral expression and phonetics modules were targeted.

3.4.2.1.1. Piloting. To ensure the reliability and validity of the questionnaire, it was pilot—tested by an expert teacher. It was administered to my supervisor who is a teacher of oral expression and phonetics modules at Mohamed Khider Biskra University. She took part in answering the questionnaire first and leaving her reviews and comments. She piloted the questionnaire since she has the experience needed to provide feedback on the questions, and the overall structure of the questionnaire. This step was crucial because her insights helped improve the questionnaire and allowed me to identify and correct all the issues before administering the questionnaire to the teachers. Only questions relevant to the study were left and were divided into sections based on her feedback. The questionnaire was piloted twice, at first, she left her comments and feedback on the questions and the form and then answered the questions after it was updated based on her review. This has increased the chances of receiving meaningful and accurate results.

3.4.2.1.2. Administration. The final version of the questionnaire (Appendix A) was administered to four EFL teachers who are currently teaching Oral Expression and Phonetics modules at Mohamed Khider Biskra University. This instrument was used to discover teachers' attitudes and thoughts toward using technological tools to improve their learners 'pronunciation and speaking skills. Brown (2008) argued that a questionnaire is a set of systematically structured questions that a researcher designs to get information from potential respondents.

3.4.2.1.3. Questionnaire Sections. The semi-structured questionnaire (Appendix A) comprised different questions: Open Ended Questions, Closed Questions, Likert Scale Questions, and Multiple-Choice Questions. It was divided into six main categories, and every category encompasses questions including (1) Teachers 'personal information, (2) Teaching Phonetics and Oral Expression, (3) Teaching Pronunciation, (4) Pronunciation and Technology, and (5) Teachers 'perspectives and opinions towards the use of technology to teach pronunciation. Later, the responses were collected before launching the experimental phase. One of this research aims was to test the effectiveness of ASR technology in improving EFL learners 'pronunciation and speaking skills. Thus, the researcher thought that teachers of oral expression and phonetics were the ones who could contribute to collecting relevant information. Therefore, four oral expression and phonetics teachers were invited to answer the questionnaire questions. This step enabled me to gather data about their views and thoughts on incorporating technology into their language classrooms to teach different pronunciation aspects and improve their learners 'pronunciation and speaking skills.

3.4.2.2. Speaking Tests. During the pre-experimental phase, the participants agreed to participate   in a pre and post-test, and the consent form included an agreement to participate in

these two tests. Boyle (2011) justified that a test is a set of stimuli to elicit an individual response based on an assigned numerical score. Hence, the pre-test and post-tests were used to determine participants 'pronunciation abilities before and after the training. In our case, we tested participants' pronunciation improvements after training using an ASR application. Both groups, the experimental and the control group, were tested before starting the training period. Upon the completion of the training, both the pre-test and post-test were conducted on the training Application Programming Interface (API) official website (speechace.com/Speaking-test_), which is designed to assess users 'speaking skills by giving them a set of questions to answer and then a score. The analysis was also done on the speech analysis application and website, which scored and analyzed participants 'speech based on correct pronunciation, intonation, words per minute, and fluency analysis. All participants participated in the pre-test and the post-test. Individual meetings were scheduled with them and the researcher took into consideration their timetable and their free time.

3.4.2.3. The Context and Material. The pre-test and post-test took place in empty classrooms at the English division building. One participant was invited at a time to avoid any background noise that would negatively affect the recording process. The researcher also considered students' self-confidence, and tried to make them feel at ease when speaking. The material used in the tests belongs to the researcher, 4G LTE internet connection was used to access the website, a MacBook Air (13-inch, 2017), and its internal microphone for a smooth voice recording. The website offers speaking tests that measure users' pronunciation and speaking skills, as it also provides them with a score of their achievement and a fluency score. The test has a wide range of topics that users can choose from (Traveling, Food, Jobs, etc.).

However, the topic "internet" was selected since it is related to the study's theme. Once users access the website, a realistic avatar examiner appears on the screen, and they monitor and guide the users throughout the test. The avatar creates an engaging and fun atmosphere for the participants. The test involves a set of three questions which are open-ended contextual questions that asses users 'ability to speak. Each question has a preparation phase and an answering phase. Useini (2019) argued that this type of pronunciation testing (i.e., reading aloud) requires preparation time to give some time for the students to get familiar with the text they will read. Thus, the preparation phase lasts 30 seconds, allowing the participants to practice their speech before recording their answers.

Since this study aimed to improve learners 'pronunciation, handouts with answers to the questions were prepared and handed to the participants. The main aim was to expose all the participants to the same material and input. The Speechace speaking test is a fully automated speech recognition-based test that evaluates a user's spoken English ability through a 10-minute simulated interview. It uses a virtual avatar-driven interface wherein the avatar provides instructions to the user while taking the test. At the end of the test, the student is given a total score on their pronunciation, vocabulary, grammar, and fluency (Speechace, 2022). Hence, their scores were only based on their pronunciation and fluency to avoid the intervention of other skills like vocabulary and grammar mastery. The users select the test theme, and the website records their answers to the questions. The answering phase lasts for 60 seconds to allow the participants to read their answers, and it automatically records them, saves them, analyses them, and then provides a score. The whole experience is explained and demonstrated in the following pictures.

Once the users access the website, a realistic avatar appears and interacts with the users. Firstly, the avatar invites the users to check their microphone by pressing the record button and saying, "Hello, how are you?". The participants were impressed by this interaction and showed great interest and curiosity toward this Artificial Intelligence (AI) technology. They were shy at first, but they were encouraged to answer the questions and the researcher explained every ambiguous step and guided them through the entire process.

Picture 4

A Realistic Avatar Picture Instructs the User to Check the Microphone (Speechace, 2022)
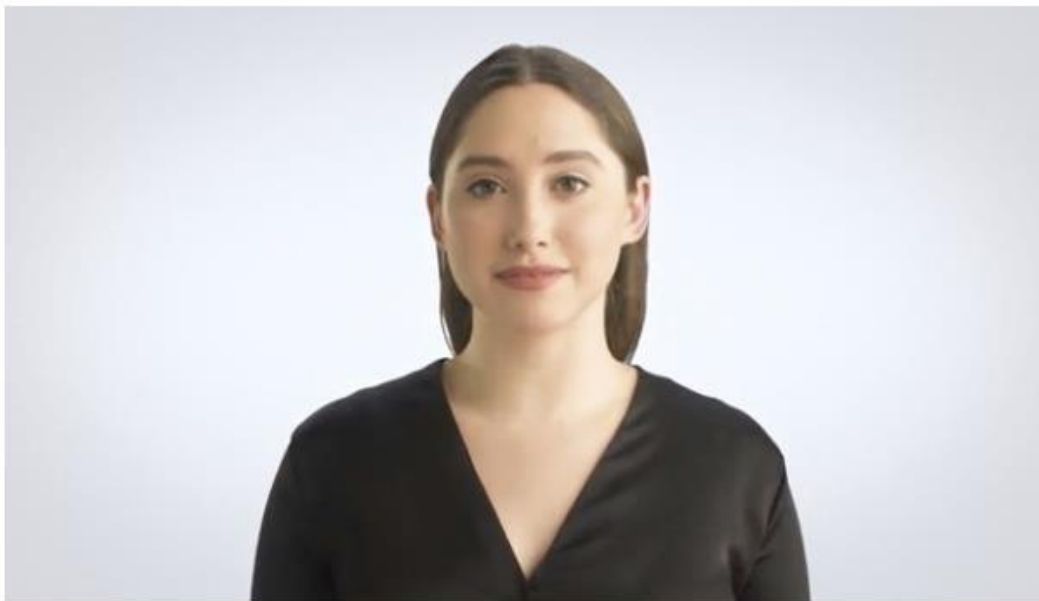
Once the avatar checks the microphone, a green circle on the screen indicates that the user's voice is being understood and recorded. Then, the questions screen appears, and the avatar

asks three questions, each at a time. Since the internet connection (4G LTE) was slow, the website took longer to process and save the answer. The website allows the users to have 30 seconds of preparation time, allowing the participants to read the answers silently and prepare themselves before recording.

Picture 5

The Preparation Time Screen and Timer (Speechace, 2022)
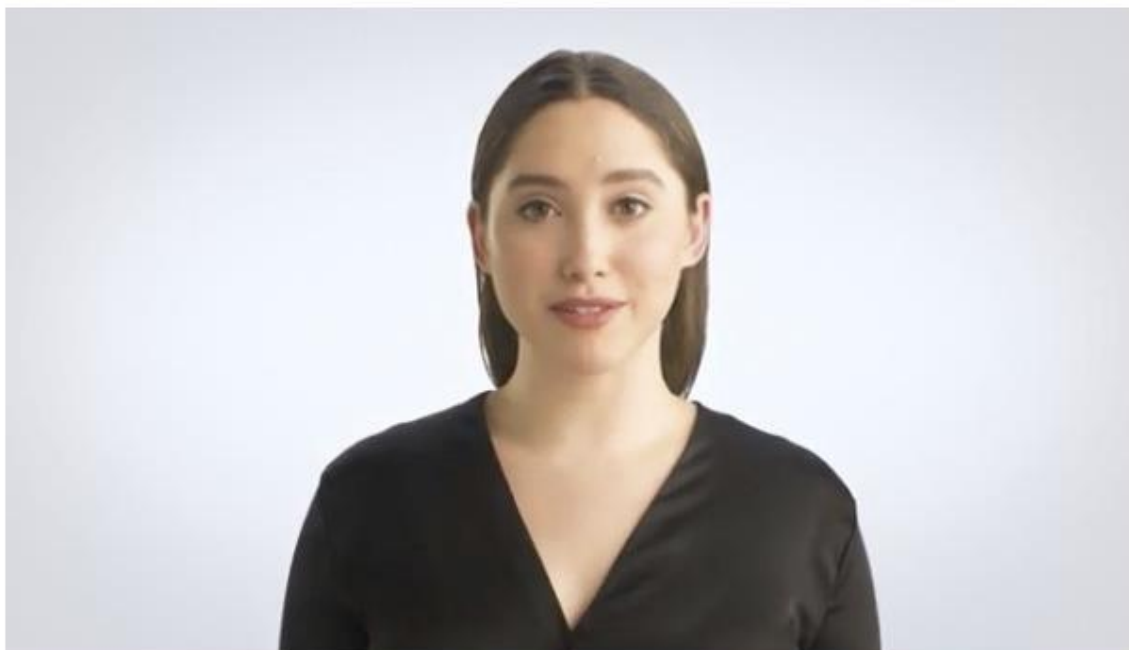
After the preparation time, a recording screen directly starts recording users 'answers to the asked question. In this step, the website gives 60 seconds to allow the users to record their answers. Hence, the participants were requested to read with a loud and clear voice before their time was up.

Picture 6

Answering Phase with a Recording Timer (Speechace, 2022)
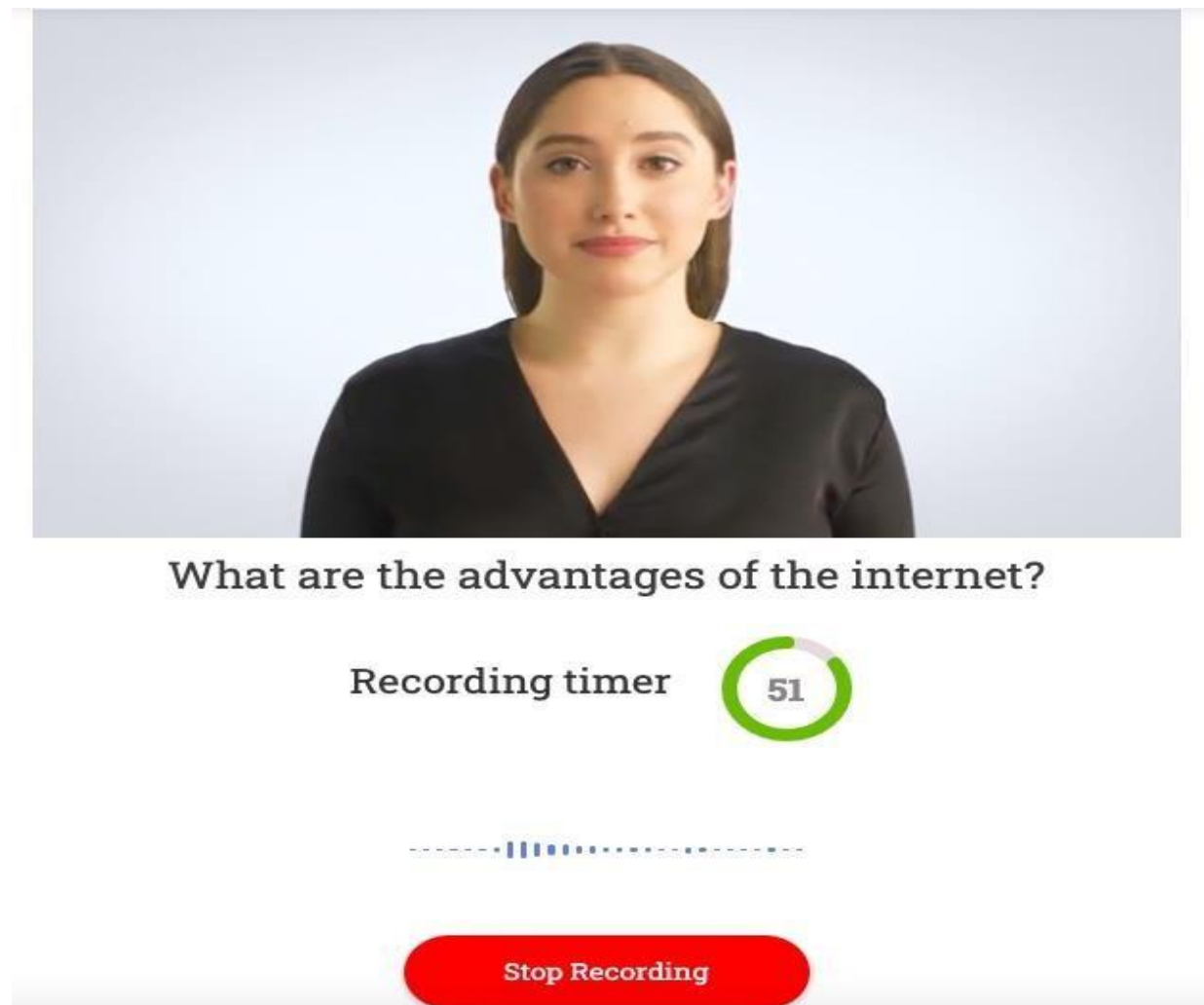
Once the recordings of the questions are over, the website processes the answers for a few minutes and then displays the score. The upcoming picture depicts the scoring screen.

Picture 7

The Scoring Screen (Speechace, 2022)

Summary of scores

Speechace | CEFR | IELTS | PTE | TOEFL | TOEIC

Speechace score 0.0/9

? Scoring guide

Pronunciation 6.5/9
Fluency 5.7/9
Vocabulary 5.3/9
Grammar 5.5/9

Retake test | Detailed report
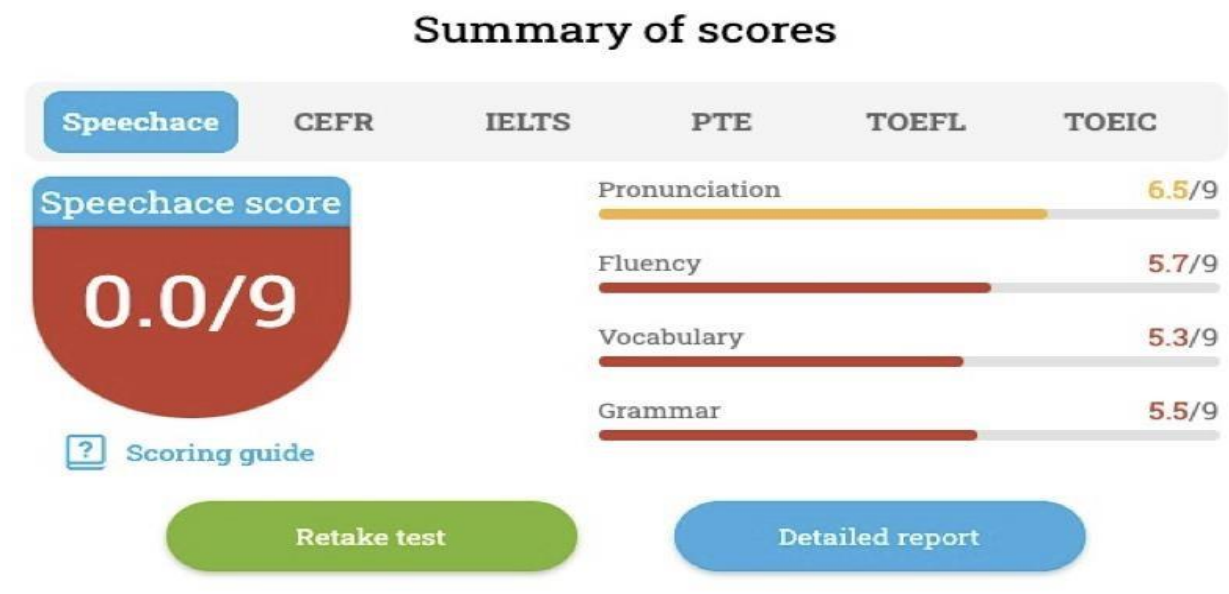
Picture 7 demonstrates the general score for all the questions, a score for each question answered in a detailed report that involves users 'written answers and the recordings, emphasizing the errors committed in pronunciation and fluency.

3.4.3. Experimental Phase

After conducting the pre-test, collecting and saving the participants 'answers, these data were kept for the analysis phase. This phase involves an intervention which is discrete training using an ASR application. All the information related to the ASR website and the application is mentioned in detail in the upcoming parts.

3.4.3.1. Instructional Website: Speechace. In this study, the researcher looked for an ASR interface that has the potential to test, improve, and assess learners 'pronunciation. Kaiser (2017) confirmed that there is no one "best app" that pronunciation teachers will use to meet the needs of their learners. Instead of searching for one perfect app, it is more productive to ask which apps might be helpful to integrate into pronunciation teaching. Moreover, Derwing and Munro (2015) invited teachers to "read reviews and recommendations from authoritative sources and then to screen apps carefully before recommending them to students" (p. 124). After examining many ASR websites, apps, and tools that are available or were used in previous studies that possess the ability to assess pronunciation do not work or do not provide accurate corrective feedback. Even though many free apps and software can be installed and used. We opted for an online website that is ASR-based and possesses a high ASR technology that provides many options that can allow users to improve their pronunciation. Hence, the automatic speech recognition [ASR] website used in the current study is Speechace. It is the first and only speech recognition Application Programming Interface [API] designed to evaluate language learners 'pronunciation and fluency. This API only serves the Education segment and is used by some of the largest worldwide publishers, language learning providers, Universities, and K-12. This Application Programming Interface [API] is used because it provides immediate feedback and correction of learners 'errors in pronunciation by giving them a score (Speechace, 2022).

3.4.3.2. Instructional Website: Configuration. The current COVID-19 circumstances obliged the teachers and learners to adapt to continuous unanticipated changes. For instance, traditional and online classrooms were adjusted to deliver the lessons. Hence, traditional classrooms took place at the university, whereas online learning was administered on online

websites and platforms such as Moodle and Zoom. As a result, the researcher looked for an online website that may be used to conduct the study and launch the training. However, the website was a new interface and the participants needed to gain prior knowledge on how to access it and use it. Hence, the researcher took the time to schedule extra meetings and used other online platforms to help participants know how to access the website and utilize it. The participants were requested to sign up on the website by creating their profiles using their emails and information. The website is straightforward to access, use, and record. It uses native HTML5 recording for Secure sites (HTTPS), which spares users the pain of installing and configuring software since it simply needs an internet connection, a computer, or a smartphone. Participants used their equipment due to Covid-19 constraints their timetables and personal circumstances were taken into consideration. Further, the training phase occurred online, i.e., each participant would access the site independently at any time and place. The training phase was handled individually at a convenient time and place. For this purpose, the researcher had to select a website that offers easy access and usage. The SpeechAce website is supported on laptops and personal computers and works well with all mobile devices. Participants with an Android device can easily access and use the site using the Chrome browser. For participants who own iOS/Apple devices such as iPhones, iPads, or iPods, the SpeechAce website works very well on the Safari browser. Therefore, each participant can easily access the site using their device without the need to attend regular sessions in the computer lab or download any application. This was more flexible and made learners feel more conformable and self-confident while completing the training.

3.4.3.3. Instructional Website: The Training. After conducting the pre-test with both groups, the experimental group received the treatment for two weeks. The online sessions were

devoted to teaching pronunciation skills using the online website. The content of the training

lessons and the targeted pronunciation skills are highlighted in Table 4.

Table 4

The Training Sessions

Session	Targeted skills	Content	Duration
1	**Introduction to the website and how to utilize it**	A PowerPoint presentation that shows how to access and use the website. Video, screen, and voice recordings.	30 minutes
2	**Beginner Level**	Simple Alphabet single words, sentences, word numbers, and number sentences.	20 minutes
3	**Vowel mastery**	Practice of simple short vowels: /ɪ/, /i/,/ɛ/ ,/æ/, /ə/ ,/ʌ/, /ɔ/ , /ɑ/, /ʊ/ and /u/.	10 minutes
4	**Consonant mastery**	Practice the consonants: /b/, /p/ , /d/ , /t/, /g/, /k/, /dʒ/, /tʃ/, /v/ , /f/ , /ð/ , /θ/.	10 minutes
5	**Pronunciation of simple sentences / basic sentences**	Practice a set of sentences that are in different tenses. (Simple present / Present progressive / Past simple / Future tense)	30 minutes
6	**Beginner Sentences**	The practice of different daily topics: • Time • Reported speech • Countable and uncountable nouns • Shopping conversation • Check in to a hotel • Dinner conversation • Giving directions	20 - 30 minutes
7	**Fluency speaking Practice**	Practice fluency by speaking about a set of different topics: Sports / Travel / People	10 - 20 minutes
8	**Poem**	Practice reading a poem.	10 minutes

The first session of the training took about 30 minutes. Unlike all the other online sessions, the first session was conducted face-to-face with the experimental group participants. It aimed to expose the experimental group participants to the ASR website and how to access and use it. Additionally, it aimed at making them recognize the benefits of the training, motivate them to participate, and develop positive expectations about it. Hence, the experience would help them enrich their background knowledge or acquire new knowledge about ASR technology and raise their awareness about the importance of AI in various life domains in general and in language learning in specific. In advance, the participants were given consent forms outlining the whole procedure and their roles in the study. The session included a presentation about the website and its usage.

The other sessions were held online once the participants created their profiles on the website. WhatsApp groups were used to communicate with the participants, launch the training sessions, and guide the participants through the entire training phase. As mentioned in the table above, each session focused on a specific pronunciation aspect—the second session aimed at allowing the participants to practice a set of simple basic English sentences. At first, the website exposes the participants to single-word practice where they have to utter simple Alphabet words (25 words in total). The participants are provided with the correct pronunciation of the word since a correct input is crucial for correct pronunciation, a record button to record their answer, and then a score of their pronunciation based on the correct pronunciation of each letter.

Picture 8

Simple Alphabet Word Practice (Speechace, 2022)
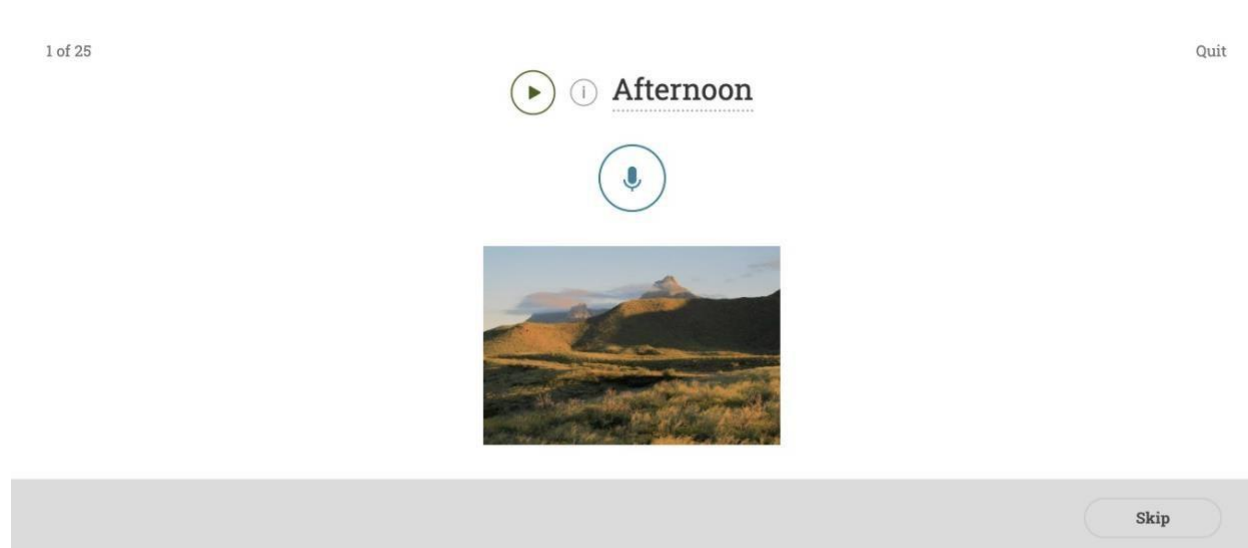
An example of the alphabet word practice is demonstrated in Picture 8.

Then, the third session involves the practice of simple short vowel sounds. The practice contains a set of words with many simple short vowels. English has far more vowels than Arabic, including 20, 12 long, short, and eight diphthongs (Kennworthy, 1987; Roach, 2009 RogersonRevell, 2011). The researcher aimed to help the participants reach an intelligible pronunciation that is accent-free and avoids the interference of their mother-tongue sounds. Hence, the participants were exposed to the simple sounds of English before engaging in complex sentence and fluency training. The vowels practiced during the first session are highlighted in Table 4. Accordingly, the fourth session includes consonant practice.

The English language contains 24 consonants, while Arabic includes 28. Hence, the website allows the participants to learn and practice words containing different consonants. After practicing basic English sounds, the fifth session aims to expose the participants to complete sentences. This enables the participants to apply their previous knowledge to utter more complex sentences. They must record themselves speaking sentences in different tenses (present simple, past simple, past progressive, and future tense). The fifth session was devoted to simple basic sentences pronunciation. Since my study aimed to improve certain pronunciation sounds and speaking fluency, this session aimed to enable the participants to practice a set of sentences that are in different simple tenses. The sixth session was devoted to simple basic sentences pronunciation. Since my study aimed to improve certain pronunciation sounds and speaking fluency, this session aimed to enable the participants to practice simple everyday life dialogues and conversations. The practice contained a wide range of topics, allowing the participants to practice their pronunciation and fluency and enrich their vocabulary. The seventh session included fluency practice on different topics. Byrne 1986 defined fluency as "[…] the ability to express oneself intelligibly […] reasonably accurately and without too much hesitation" (p. 9). Hence, fluency is crucial, and improving this aspect makes participants intelligible and able to speak without facing any difficulties. The website assesses participants 'fluency by counting the words they utter per minute, their repetitions, their self-corrections, hesitations, and pauses. The participants were asked questions and then provided the answers in the form of a long paragraph related to different topics to measure their fluency.

Picture 9

Fluency Speaking Practice (Speechace, 2022)

▶ **Press play to listen to the question and then record your answer**

I agree. The world is big, and we need to make an effort to step out of our comfort zone to meet people. Otherwise we become narrow minded and we stop learning and growing. By getting to know each other, we are less likely to be misunderstood. I want to meet and befriend with people who have different experiences in life. I think by not interacting with each other, we become arrogant and

🎤   Skip

The picture shows an example of fluency speaking practice.

The concluding session encouraged the participants to read a poem. According to Elting and Firkins (2006), poetry is a rich language resource for English learners, while Gilbert (2008) stated that poems break the "rules" of a language because they have no particular order of grammar. However, poems require reading every word individually and must be pronounced correctly by considering syllables and stress. Thus, it can help the students to improve their English pronunciation skills. Additionally, it was a fun practice to conclude the training phase. The training took longer than expected because of the beginning of the tests, and the participants were busy preparing for their tests and other course assignments.

3.4.3.4. Instructional Website: Scoring. As mentioned earlier, the website offers many lessons and topics that allow users to practice and improve their pronunciation. The participants practiced on the website after creating their profiles, and their answers were processed, analyzed, and then saved by the website using an algorithm. Every time the participants perform a task, the feedback is displayed as a score on the screen with a percentage of achievement. The screen turns green, indicating that the user uttered the target words with the correct pronunciation. Accordingly, the user can listen to their recording and re-record it if unsatisfied with the outcome.

Picture 10

Correct Pronunciation Score Percentage (Speechace, 2022)

Picture 10 shows the scoring screen and feedback after achieving a correct pronunciation.

If the user commits an error in pronunciation, it shows the mispronounced phoneme colored red and its correct transcription. This indicates that the sound was mispronounced. The feedback is precise and encouraging since it allows the user to realize their error, how to correct it, and the option to repeat it repeatedly until achieving the target pronunciation. Further, the feedback also

provides users with individual listening options to a recording of themselves and how they pronounced the phoneme. A "play button" also allows users to listen to the target pronunciation and try to imitate it again. This exposes users to continuous practice and enables them to improve their pronunciation. Picture highlights this process.

Picture 11

Incorrect Pronunciation Score Percentage (Speechace, 2022)
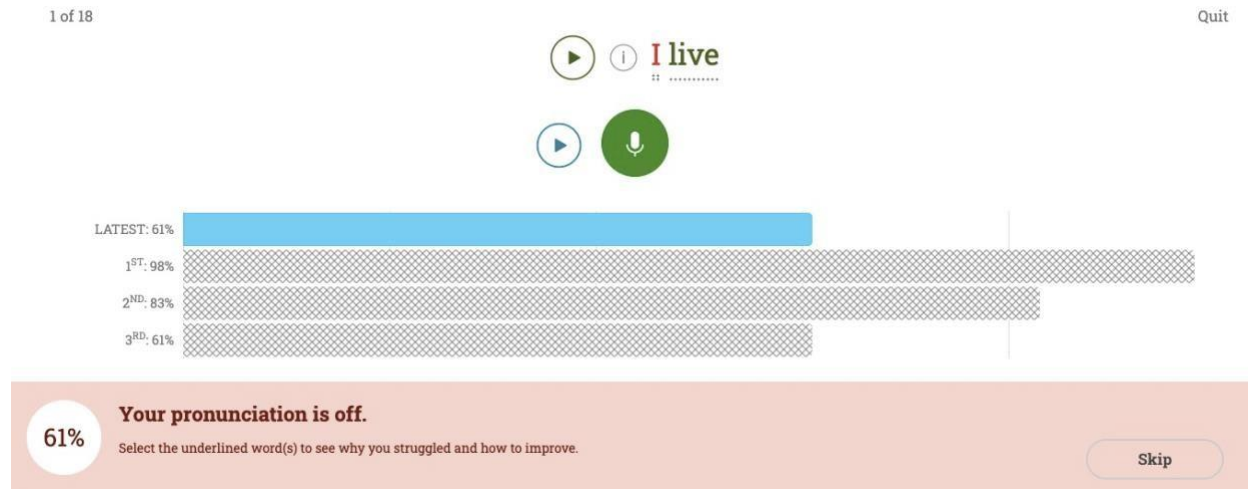
Upon completing each task, the participants were provided an overall score of 100 on every session saved on their profile. Further, the researcher had access to all of the participant's profiles and was allowed to report and review their pronunciation scores on every task as well as to listen to their audio and their achievement by checking their attempts (The website shows the number of attempts the user took to reach a correct pronunciation).

3.4.3.5. WhatsApp Groups. Since the study was conducted during the pandemic and the participants were unavailable daily, the researcher opted for an online social media platform to stay in constant contact with them. Online social platforms offer various benefits for connecting and sharing interests and information while allowing users to maintain physical separation. These internet-based applications permit users to create personal profiles and maintain a list of other users or friends with whom they may share content and participate in social interactions and networking (Boyd & Ellison, 2007; Kaplan & Haenlein, 2010). Hence, after receiving the participants' approval, WhatsApp groups were created. One group included the control group participants to get informed about the schedule of the pre-test and post-test, whereas another group included the experimental group participants. This eased the communication process between the researcher and the participants. The group was used as an online technological tool to keep up with the participants and inform them of any new announcements.

Because of the pandemic, the groups allowed the participants to share their experiences in a more relaxed, closed, and personal space. They could practice daily on the ASR website on their smartphones or PCs and then access the group to share their thoughts and experiences. It was a fun experience for them; each shared screenshots and video shots of their practice progress. Hence, it was a successful tool allowing the participants to share positive thoughts about the experience and the training. They could contact the researcher or their classmates if they had questions or faced obstacles while using the website. The groups enabled us to communicate in a more relaxed, natural environment and capture participants 'attitudes throughout the training. Further, the participants were encouraged to share their thoughts concerning the website. The researcher tried to keep in touch with all the participants throughout the training, directed and encouraged them to

practice using the ASR website on their smartphones or computers, and invited them to share their thoughts, experiences, screenshots, and video shots with the entire group.

3.4.4. Post-Experimental Phase

Post the experiment, the researcher evaluated the effectiveness of the intervention and the training. Hence, the post-test was conducted to measure the participants 'progress and compare their scores before and after the training. Additionally, the participants 'overall satisfaction with the training was investigated as well.

3.4.4.1. Post-Training Interview. Upon the completion of the training, an interview was conducted with the participants of the experimental group. This was done to explore the participants 'attitudes toward the training. After the post-test, meetings with the experimental group participants were scheduled based on their timing.

3.4.4.1.1. Setting. The interviews occurred in empty classrooms in the Department of Letters and Foreign Languages. However, some interviews were conducted online because some participants lived far from the university and could not meet in person. Both in-person and online interviews were recorded using the researcher's mobile phone and laptop (iPhone XR/Macbook Air 2017). There were no time restrictions to the interviews; each lasted based on the participant's answers. The purpose of conducting a semi-structured interview was to investigate participants ' opinions and attitudes towards their training and their thoughts on the ASR technology. A semi-structured interview offers a flexible process; open-ended questions allow the emergence of unanticipated responses and spontaneous issues (Tod, 2006; Berg, 2009).

3.4.4.1.2. Validity and Reliability. To ensure the quality and credibility of the research, the questions of the interview were designed based on the participants' comments and feedback. After each interview, participants were asked additional questions and were asked to leave comments and feedback on the interview questions. This step allowed us to test the interview questions, ensure that they are clear, appropriate, and relevant to the study, and will enable the researcher to get the information needed to achieve the purpose of my research. The interactions were held privately either in person or via WhatsApp messages, mother-tongue was also used to enable the participants to express their ideas freely to elicit further in-depth information. Their feedback was taken into consideration, improvements were made to the interview questions, and their comments were classified with the other answers. The interview entailed only open-ended questions because they allowed the participants to speak freely. The questions inquired about participants 'thoughts on using this technology to improve their speaking skills and attitudes towards the website used during the training period. It also contained questions about their overall experience, difficulties, and whether they noticed any improvements in their speaking skills after the training. It is essential to point out that the participants had no prior experience using this or any other ASR tool to train and improve specific language skills. The participants agreed to take the interview; their answers were written down by the researcher and recorded. All these data were then saved and prepared for further analysis (See Appendix I).

This section provided a thorough explanation of how this study has been conducted. It included the experimental method and design, the sample population, and the different data collection tools and techniques used to gather data. Further, it demonstrated how the ASR tool

helped in the recordings and how the scores were generated. How the data were analyzed is revealed in the next section.

3.5. Data Analysis

Various data-gathering tools were used to collect and analyze data. A questionnaire was administered to teachers, interviews were conducted with the participants, participants' pronunciations were recorded, and scores were generated and compared. Thus, this section is dedicated to introducing the data analysis methods and techniques followed to analyze the data collected throughout the different segments of the current research. According to Powney & Watts (2018), data analysis is the process in which the data can be interpreted in an organized manner. The data analysis phase is about transforming raw collected data into lucid information. It is considered the most crucial phase in research because it should answer the research questions and prove or deny the hypotheses.

3.5.1. Analysis of the Pre-experimental Phase

After the observation phase and recruiting the participants who will take part in the study, the pre-experimental phase was launched. This phase included a questionnaire for Oral expression and Phonetics teachers. Further, during this phase, the participants of both groups, the control and experimental groups, participated in a speaking test. The content of the test and the way it was conducted and analyzed are discussed in the upcoming parts.

3.5.1.1. Analysis of Teachers 'Questionnaire. This instrument enabled the researcher to gather teachers 'insights and opinions on teaching pronunciation. The questionnaire was administered to four (4) teachers who teach oral expression and phonetics at the Department of English at Mohamed Khider Biskra University. The questionnaire was selected as an additional data-gathering instrument to explore teachers 'attitudes and opinions towards pronunciation teaching, types of classroom activities practiced, pronunciation assessment, and technology use inside their classrooms. The questionnaire was divided into five main sections, and every section encompasses different questions. Content analysis was employed to determine the occurring themes and differences between teachers 'answers, and to identify their attitudes toward technology and pronunciation teaching. Their responses were read and then categorized. Then their answers were presented as frequency distribution. The questionnaire questions are displayed in the appendices (See Appendix A), and the teachers 'answers are revealed in the results chapter.

3.5.2. Analysis of the Speaking Test

To investigate the effectiveness of the ASR technology used in the study, the participants took part in an online speaking test that focused on scoring their pronunciation and fluency—the test used in this study aimed at measuring the students 'general English pronunciation proficiency. The pretest and post-test were given to measure the participants 'pronunciation ability and speaking proficiency before and after providing the treatment. The previous section discussed that the ASR website provides scores based on the users 'pronunciation and fluency. Hence, in the speaking test, the website gives a score on a nine-point scale where 0 = incorrect pronunciation and 9 = excellent pronunciation. For the analysis to be reliable, it is vital to understand the

algorithm followed by the ASR website and the basis on which these scores are provided and given. The scoring screen of the speaking test on the online site (speechace.com) analyzes the speech based on specific criteria. The site offers a SpeechAce score, CEFR (Common European Framework of Reference for Language), IELTS (International English Language Testing System), PTE (Pearson Test of English), TOEFL (the Test Of English as a Foreign Language), and TOEIC (Test of English for International Communication). Further, the site provides the user with a final test report that entails the scoring of each question, the possibility to listen to your voice recording, the overall score, the pronunciation score, and the fluency score.

Picture 12

A Snapshot of the Overall Feedback (Speechace, 2022)

Descriptive feedback

Overall (7)
Demonstrates generally good fluency and coherence while speaking but may take occasional pauses. Is proficient in using sophisticated vocabulary and idiomatic structures. Proficient in expressing complex thoughts using a range of grammar structures. Has reasonably good pronunciation with some accent.

Pronunciation (7.6)
Uses a wide range of pronunciation features. Sustains flexible use of features, with only occasional lapses. Is easy to understand throughout; First language accent has minimal impact on intelligibility.

Fluency (6.5)
Capable of willing to speak at length, though may lose coherence at times due to occasional repetition, self-correction or hesitation. Uses a range of connectives and discourse markers but not always appropriately.

Picture 12 shows the overall feedback provided by the website, pronunciation score, and fluency score with comments.

3.5.2.1. The Analysis of Participants 'Scores on the Speaking Tests. The site includes a guide on how this scoring was done and the criteria behind each score from zero to nine.

Table 5

Criteria of the Speaking Tests Scoring

Score	Overall score Criteria	Pronunciation Score Criteria	Fluency Score Criteria
4 and below	The user does not have adequate fluency and coherence and needs remedial training in speaking fluently.	The user uses a limited range of pronunciation features.	The user speaks with long pauses and cannot link simple sentences.
5	The user demonstrates below-average fluency and coherence and may need remedial training in speaking fluently and language construction.	The user produces primary sentence forms with reasonable accuracy in pronunciation.	The user usually maintains a good flow of speech but uses repetition and self-correction.
6	The user demonstrates reasonable fluency and coherence. The user can generally be understood while speaking but mispronounce some words frequently.	The user uses a range of pronunciation features with mixed control.	The user is capable of being willing to speak at length.
7	The user demonstrates fluency and coherence while speaking but may take occasional pauses. The user has reasonably good pronunciation with an accent.	The user confidently uses various pronunciation features and can be easily understood.	The user speaks at length without apparent effort or loss of coherence.
8	The user demonstrates excellent fluency and coherence with occasional pauses. The user has very good pronunciation and with very mild accent.	The user uses a wide range of pronunciation features.	The user speaks fluently with only occasional repetition or self-correction.
9	The user demonstrates excellent fluency and coherence with rare to no pauses. The user has an excellent pronunciation and sounds similar to a native speaker.	The user uses full range of pronunciation features with precision and subtlety.	The user speaks fluently with only rare repetition or self-correction.

Table 5 demonstrates the reason behind the scoring of the overall speaking test, the speaking fluency, and the pronunciation accuracy. Later on, the scores obtained from the pre-test are compared with those of the post-test to check participants 'improvements.

3.5.2.2. The Analysis of Participants 'Pronunciation. The pre-post-test aims to measure the participants 'level of speaking and pronunciation before and after the training phase. The last part inspected the overall scores of the pre-test and the post-test as calculated by the online website. In addition to the score, the researcher aimed to assess and analyze the recordings of the participants and compare them with the post-test to detect any improvements. The participants ' speech was recorded on the website. Before this, the researcher helped each participant create a profile using their email to save each participant's score and recording. Once the recording is over, the website provides an overall score discussed in the previous part. Additionally, a more detailed test result is also displayed below. The detailed report includes the audio recording of the participant's speech, word transcription of the recording, feedback provided in the form of an emphasis on the mispronunciations that are highlighted in red, the words uttered per minute, the pauses made while speaking, and the accuracy of speaking in the form of a percentage.

Picture 13

Question Evaluation and Recording with Feedback (Speechace, 2022)

Picture 13 demonstrates the participant's complete answer, all pronunciation and fluency proficiency information, and a play button to listen to the recording.

Screenshots of participants 'answers, mispronunciation, and accuracy percentage are mentioned in the appendices.

3.5.2.3. The Analysis of Participants 'Fluency. Additionally, the study was interested in improving EFL learners 'overall pronunciation using ASR-based technology. Focusing on other aspects of pronunciation was crucial rather than only mastering a set of sounds. Accentedness is about judging how far a non-native speaker's pronunciation diverges from a native speaker's target; intelligibility is related to how accurately listeners can identify the spoken language; and comprehensibility is about how easy it is for a listener to understand speech (Munro & Derwing,

1995a; Munro, Derwing & Morton, 2006). Hence, the researcher focused on measuring participants 'fluency since it combines all the above-mentioned aspects (Derwing & Munro, 1997; Derwing, Munro & Thomson, 2008; Derwing, Thomson, & Munro, 2006; Isaacs & Thomson, 2013). In applied linguists, language teachers use the term fluency to refer to the fluidity or ease with which the language is spoken (Derwing et al., 2004; Freed, 2000; Isaacs & Thomson, 2013; Kaponen & Riggenbach, 2000). An EFL learner may be described as fluent, even though they have only rudimentary grammatical ability, limited vocabulary knowledge, and poor pronunciation. In this context, describing lower proficiency learners as eloquent means that their language knowledge is easily accessed and that their oral language is produced without hesitation (Segalowitz, 2010). In addition to collecting and comparing pre-test and post-test scores and mispronunciations, measuring participants 'fluency and checking if they improved since it affects their overall pronunciation mastery was crucial.

Derwing et al. (2004) defined fluency as "an automatic procedural skill on the part of the speaker and a perceptual phenomenon in the listener" (p. 656). Thus, listener judgments are used to measure a speaker's fluency. Since the current study included recordings, we looked for other criteria to measure participants 'fluency. Hence, many researchers have followed more objectively quantifiable measures of fluency. Derwing et al. (2004); Derwing, Munro, Thomson, & Rossiter (2009); Kormos (2006); Kormos & Dénes (2004); Towell, Hawkins & Bazergui (1996) argued that the following criteria should be followed to measure a speaker's fluency.

- o Speech Rate: refers to the average number of syllables spoken per second or minute.

- o Phonation Time Ratio: the percentage of time that is devoted to speaking. It refers to the total time taken to produce an utterance.

- Pruned Syllables: the average number of syllables spoken per second or minute after removing self-repetitions and self-corrections.

- Articulation Rate: the average number of fluent syllables per second or minute between pauses of a predetermined length.

- Mean Length of Run: the average number of words or syllables produced between pauses of a length.

- Silent Pause Ratio: the number and time attributed to silent pauses.

Picture 14

Fluency Scores (Speechace, 2022)

The online website measured fluency in speaking. Picture 14 demonstrates the fluency evaluation that the website provides. The fluency score screen shows a score out of 10 (similar to

the pronunciation score and the overall score), the words uttered per minute, the speaking rate, and the silent pauses taken.

Finally, the overall accuracy of all the participants 'pronunciation was calculated. The Pretest and Post-test results were exported into SPSS and compared. The mean accuracy and the standard deviation SD were also identified for the pre-test and post-test. The other data related to fluency were also calculated by counting the mean score and the standard deviation SD. The results are discussed and compared thoroughly in the next chapter.

3.5.3. Analysis of the Experimental Phase

After the pre-experimental phase, which included the speaking tests and saving its data, the training phase was launched. This phase involved discrete training using an online ASR website. The latter lets users practice and improve their English pronunciation and speaking skills. During this study stage, the participants practiced individually at home using their equipment (mobiles, laptops, internet access). There was no observation or direct communication with the participants; only online interactions existed.

3.5.3.1. Analysis of the Participants 'Accounts. Each participant from the experimental group needed private access to the website. The researcher helped and guided each participant on how to use their email to create a private profile on the online website. The emails and passwords were shared with the researcher to check participants 'completion of the required tasks and their scores. The data collected from each participant's profile was coded for good, average, and bad scores. Although the scores are unrelated to the testing and have no significance to the

study, it was crucial to collect them to determine the usefulness of the ASR website and how well the participants were accustomed to using it. The data was collected through snapshots of participants 'profiles and classified according to their achievements.

      3.5.3.2. Analysis of the WhatsApp Group Interactions. This part of the analysis included a qualitative analysis of participants 'WhatsApp messages and comments about their experience and attitudes towards the ASR website and its usefulness. Cohen, Manion, and Morrison (2007) describe this type of analysis as "organizing, accounting for and explaining the data; in short, making sense of data in terms of the participants 'definitions of the situation, noting patterns, themes, categories, and regularities" (p. 461). Hence, the data were collected and classified into a set of themes. The participants 'messages and screenshots were collected, copied, and saved in a Google document for further analysis. All the personal information of the participants was removed. The messages were coded into themes, positive attitudes, and beliefs to analyze the data. It was argued that the content of social data media could provide many facts, opinions, imagination, and people's feelings (Yang & Zhao, 2016). For example, the messages in which the participants expressed positive attitudes toward the ASR website. In some messages, they expressed negative thoughts and messages about the obstacles or issues they faced while using the ASR website. In other messages, they discussed the usefulness of the ASR website and so on. Similar messages were classified and coded in different colors and comments to facilitate analysis. The screenshots were also organized into themes. Screenshots taken by the participants in which they share their high scores, rate of completion, the percentages they got, and screenshots of them inquiring about problems they faced.

3.5.4. Analysis of the Post-experimental Phase

After the training, the experimental group participants were only requested to participate in a follow-up interview that entailed questions concerning their overall experience using the ASR technology.

3.5.4.1. Analysis of the Interview. The interview was conducted with the experimental group after the training period. It was mainly administered to investigate the participants ' perceptions and attitudes toward the ASR technology used in training. The interview was improved and conducted with all of the participants. After that, their responses were gathered and the interview items were classified; some questions were meant to discover participants 'opinions and feelings towards the technology, how they are taught pronunciation at the department, and the aspects of pronunciation that the technology helped them improve. The interview presents and discusses the answers provided by 12 students from the experimental group. All the questions are shown in the appendices (See Appendix H). Content analysis was used to detect each underlying theme, emotion, or attitude. Participants 'answers were coded into sub-themes and were reviewed to understand their feelings toward the entire experience. The purpose of each interview question and the answers are discussed in detail in the results' chapter.

Conclusion

This chapter was devoted to presenting the study's research methodology. It referred to the approach that was used to conduct the study and collect the data. It described the research method and explained and justified the importance of opting for a mixed-method approach. This kind of method allows for providing a complete and thorough understanding of the research. It offered flexibility in selecting the population and the sample. It also explained the data-gathering tools and techniques. It further discussed the qualitative and quantitative data collected throughout the different stages of the current research. Further, it involved a description of the way the data is going to be analyzed to answer the research questions and test the hypotheses. The next chapter presents the findings and discusses the results.

Chapter Four

Analysis and Interpretation of the Results

Chapter Four: Analysis and Interpretation of the Results

Chapter Four: Analysis and Interpretation of the Results

Introduction

This chapter is concerned with the analysis and discussion of the results generated through the teachers 'questionnaire, the participants 'online accounts, the WhatsApp groups, and the semi-structured interview conducted with the experimental group participants. It is divided into three main sections to answer the research questions. The first one presents the results regarding teachers' questionnaires which aimed at knowing their opinions and thoughts on teaching pronunciation and technology. The second one reveals the quantitative effects of the participants' pronunciation and fluency scores generated by the ASR technology. This section compares participants' results during the pre-test and the post-test answering the main research question regarding how ASR technology can improve students' pronunciation and tests the hypotheses. The third one discusses students' perspectives and attitudes toward ASR technology.

4.1. Analysis of Teachers' Questionnaire

The questionnaire, administered to four oral expression and phonetics teachers, investigated teachers' attitudes toward pronunciation teaching. It was about their practices in teaching pronunciation and their thoughts on incorporating technology into the pronunciation classroom. Moreover, it explored teachers' attitudes and opinions, focusing on the types of classroom activities they practiced in phonetics and oral expression modules, the extent to which they follow the curricula, how they teach and assess pronunciation, and their attitudes towards pronunciation and technology. The questionnaire was divided into five main categories. Every category encompasses questions including (1) Teachers' personal information, (2) Teaching

Phonetics and Oral Expression, (3) Teaching Pronunciation, (4) Pronunciation and Technology, and (5) teachers' perspectives and opinions towards the use of technology to teach pronunciation.

The questionnaire answers were then gathered and analyzed. Following the content analysis method, teachers 'responses were listed and added together, and the frequency of occurrence of each particular answer was counted. The responses were then presented in tables and percentage forms. For specific questions, the responses were grouped into common themes and categories.

The findings of the questionnaire are discussed in the upcoming section.

Section 1: Teacher's Personal Information

The first section of the questionnaire included general information about teachers. In this section, teachers were asked two questions about their teaching experience and teaching the first-year level.

Question 01: *How long have you been teaching English?*

Table 6

Teaching Experience

Answer	Less than 5 years	5 to 10 years	More than 10 years
Frequency	1	2	1
Percentage	25%	50%	25%

With time, teachers gain more experience, so it was crucial to ask about their teaching experience. Further, most teachers who participated in this questionnaire have a short teaching experience. This implies that they are young, close to the students 'generation, and are aware of the modern era of incorporating technology in teaching.

Question 02: *Have you already taught the first-year level?*

In this study, first-year students at Biskra University were the targeted population. Hence, teachers were asked if they teach or have taught this level.

Table 7

Teaching the First-Year Level

Answer	Yes	No
Frequency	4	0
Percentage	100%	0%

The answers revealed in the table above show that all teachers are teaching or have already taught first-year level.

Section 2: Teaching Oral Expression/Phonetics

This section sought to identify the aspects teachers focus on when teaching the target language. It enabled the researcher to get insights into the teaching content and pronunciation aspects taught in Oral expression and Phonetics modules.

Question 01: As a phonetics teacher, what is the overall goal for teaching Phonetics to first-year students?

Table 8
Goals of Teaching Phonetics

Answer	Frequency	Percentage
Identify and produce target sounds	0	0
Master prosodic features Rhythm, stress, pitch, and intonation	2	66.67%
The ability to communicate	1	33.33%

Table 8 statistics show that only three teachers answered this question revealing that one of the teachers has never taught the Phonetics module. Regarding this question, two teachers (66.67%) announced that the primary goal of teaching Phonetics is to help their learners master prosodic features. The other teacher (33.33%) said that they focus on enabling their learners to communicate appropriately. They added that they aim to help them produce intelligible and comprehensible pronunciation.

Question 02: As a teacher of oral expression, what is the overall goal for teaching oral expression to first-year students?

Table 9

Goals of Teaching Oral Expression

Answer	Frequency	Percentage
To allow students to speak the target language correctly	3	75%
To enable students to communicate inside the classroom	0	0%
To allow students to use the target language in real-life situations	1	25%

All teachers who took part in answering this question have already taught the Oral Expression module. Most teachers agreed that the purpose of teaching the Oral expression module is to allow their students to master speaking the target language. On the other hand, one teacher specified that they want to enable their students to use the target language in real-life situations.

Question 03: *When planning your lesson, do you decide the teaching content on your own, or do you have other considerations?*

This question was asked to investigate whether teachers decide on the teaching content they bring into their classrooms. Planning a lesson is essential when teachers synthesize the curriculum and select specific teaching content that meets their learners 'needs and levels. All teachers (100%) stated that they consider their students' level and needs before deciding on their teaching content. Moreover, they emphasized the importance of considering time constraints and access to the available teaching materials. Such criteria are critical before planning a lesson and its content, especially at Algerian universities where teaching materials are limited or non-existent.

Question 04: *If your students need help to speak, do you pay attention to their errors and provide immediate feedback?*

Investigating whether immediate feedback helps allow learners to improve their speaking skills is one of the research's main objectives. Therefore, it was important to determine whether teachers give feedback to their students or not. Ellis (2009) stated that corrective feedback has two types, immediate and delayed correction. Hence, most teachers (3) confirmed that immediate feedback should be provided directly or indirectly. One teacher said he prefers to give delayed feedback only on communication errors. Regarding teaching speaking, feedback is a great technique that enables the students to realize their mistakes and avoid them in the future.

Section 3: Teaching Pronunciation

The following questions were asked in the third section of the questionnaire. This section investigated teachers 'observations, practices, and experiences teaching pronunciation to EFL learners.

Question 01: How important is pronunciation teaching?

Regarding this question, teachers were asked to rate the importance of teaching pronunciation from "Not at all important" to "Very Important." The Phonetics module teaches sounds, and the Oral Expression module focuses on enhancing learners 'speaking skills. Thus, it was crucial to get teachers 'insights on the importance of focusing on teaching pronunciation while teaching these two modules.

Table 10

The Importance of Teaching Pronunciation

Answer	Not at all important	Slightly important	Important	Fairly important	Very important
Frequency	0	0	0	2	2
Percentage	0%	0%	0%	50%	50%

As reflected in Table 10, the responses to this question indicate that all teachers confirm the importance of teaching pronunciation.

Question 02: *As far as first-year students are concerned, how would you evaluate their level of pronunciation?*

Since the sample of this study consists of first-year students, the researcher gathered teachers 'evaluations of their students' pronunciation and were asked to judge their level of speaking.

Table 11

First-Year Students 'Pronunciation

Answer	Poor	Fair	Good	Excellent
Frequency	1	3	0	0
Percentage	25%	75%	0%	0%

The data displayed in Table 11 shows that teachers do not believe their learners have good or excellent pronunciation at this level. 25% declared poor pronunciation, while 75% clarified that their pronunciation is fair.

Question 03: *Do you assess your students 'pronunciation?*

Assessing pronunciation can be a challenging task because it requires taking into account various factors and aspects. It is difficult to determine what to assess and how to measure it. So, this question aimed to determine whether teachers evaluate and measure their students' pronunciation.

Table 12

Assessing Pronunciation

Answer	Yes	No
Frequency	4	0
Percentage	100%	0%

The percentages in Table 12 show that all teachers assess their learners 'pronunciation. Their responses were based on specific criteria when asked about what they assessed. Their answers revealed that they focus on assessing their learners 'accuracy, comprehensibility, and intelligibility. Further, they evaluate it by targeting word stress, sentence stress, and silent letters. Teachers listen carefully to their learners while taking part in planned and spontaneous tasks that allow them to speak and enable the teacher to detect and correct their pronunciation errors. This technique is vital for developing learners 'communicative competence and speaking skills.

According to the teachers 'responses, it is clear that each of them evaluates an aspect of pronunciation that, in their opinion, impacts their students' pronunciation. Yet, it is evident from all of their responses that they all agree that comprehensible pronunciation is vital to successful communication.

Question 04: *Is providing authentic input and instant feedback crucial to improving students ' pronunciation?*

Table 13

Providing Authentic Input and Corrective Feedback

Answer	Yes	No
Frequency	4	0
Percentage	100%	0%

As can be seen from the table above, all teachers (100%) indicated that authentic input and corrective feedback are crucial for improving their students 'pronunciation. Teachers explained that adapting authentic materials provides language context and puts the learners in real situations in which they are exposed to native pronunciation. Accordingly, corrective feedback helps learners enhance their oral production and prevent them from repeating the same errors.

Section 4: Technology and Pronunciation

Since this study investigated the usefulness of using technological tools to promote EFL pronunciation, this section is dedication to getting teachers 'insights and opinions on incorporating technology into their language classroom.

Question 01: *Do you think that using technology to teach pronunciation can be effective?*

Table 14

Using Technology to Teach Pronunciation

Answer	Yes	No
Frequency	4	0
Percentage	100%	0%

Table 14 shows that all teachers are keen on using technology and believe it can effectively teach pronunciation.

Question 02: *Have you ever used a technological tool to teach pronunciation?*

The reason behind asking this question was to determine if teachers have ever implemented any technological tool to teach pronunciation.

Table 15

Using Technology in the Classroom

Answer	Yes	No
Frequency	3	1
Percentage	75%	25%

Table 15 demonstrates that one teacher (25%) reported that they have never been employed technology to teach pronunciation, compared to most teachers (75%) who had done so. We also were intrigued by what kind of technology the teachers were using. Teachers said they used

applications and tools like TeacherApp and Google Meet to stay in touch with their students. They also claimed to use audio-visual aids in the classroom to expose their students to native pronunciation. These statistics appear promising, and teachers are eager to look for new technologies and use them in the classroom.

Question 03: *What were your students 'attitudes toward using technology inside the classroom?*

After asking teachers if they incorporate technology into their classes, the researcher was interested in discovering how their students reacted when they were taught using technology. Most teachers (75%) said that their students were motivated and had a positive attitude toward using technology.

Question 04: *Before selecting and incorporating new technology, what options do you want this technology to offer your student to improve their pronunciation?*

Table 16

Technology Options

Answer	Authentic input	Corrective Feedback
Frequency	4	4
Percentage	100%	100%

All teachers agreed that incorporating technology into their classrooms should provide their learners with authentic input and corrective feedback. They added that the technology should be self-paced and allow self-assessment.

Section 5: Teachers 'Opinions on ASR Technology

Question 05: *My research study includes pronunciation training using an ASR website to improve students 'pronunciation. Do you think teachers should incorporate technology inside and outside the classroom/design additional training online?*

Teachers 'answers to this question revealed positive attitudes and opinions. This open question allowed me to get teachers 'insights and comments on the usefulness of using technology to improve specific language skills and whether they believe that my study could have a positive impact on students 'pronunciation. All the teachers believe that any new, innovative, technological tool introduced in the classroom will benefit the students and help them develop their pronunciation and overall communicative skills. They have explained that CAPT tools are now widely spread and teachers are keen on using such kind of technology, but they emphasized the importance of training teachers on how to use these kinds of technologies. Further, teachers declared that technological means would make it easier for teachers to cope with their students' learning difficulties. Accordingly, they have explained that out-of-class activities should be administered to students since in-class activities are insufficient to fulfill learners 'needs and promote their skills.

4.1.1. Interpretations of the Teachers 'Questionnaire Findings

According to the responses, all teachers who participated in the questionnaire have a positive attitude toward using technology to teach and assess pronunciation. Their responses assisted me in gaining a broad overview of specific crucial ideas that are pertinent to our study and are included below:

- The proficiency of second-year students' pronunciation does not meet teachers' expectations since they recognize the importance of teaching pronunciation. Teachers believe that focusing on teaching this aspect will help their learners become intelligible and comprehensible and can communicate in the classroom and real-life situations.

- Teachers of oral expression and phonetics think that exposing their learners to authentic input and providing them with corrective feedback is critical to helping them enhance their pronunciation and develop their communicative skills. o Teachers believe that using technology has a positive impact on learners 'performance and has the potential to help them promote their pronunciation and speaking skills. Further, using technology inside or outside the classroom is motivating. However, they know that incorporating technologies inside the classroom requires training and tremendous work and planning.

- It is insufficient for students to rely solely on their teachers. They must work independently and look for online resources such as online sites, software, programs, and applications that might assist their language learning outside of the classroom and help them develop many skills.

4.2. Analysis of the Experimental Phase Results

After having selected the study sample and receiving their agreement to participate, both groups (i.e., the control group and the experimental group) took part in a pre-test and a post-test. The results of the speaking tests are presented and analyzed in this section. Additionally, the experimental group underwent discrete training using the ASR online program. Hence, more data was gathered from each participant's online accounts and WhatsApp groups.

4.2.1. Participants 'Online Accounts

Before launching the training period, participants from the experimental group were introduced to the ASR program and were requested to create online accounts and share their details with the researcher. Their emails and passcodes were sent to the researcher via WhatsApp private messages. After each training session, the researcher accessed their accounts and checked their progress and level of completion. This was a way to monitor their commitment to completing the requested tasks and check their achievement. The website includes a set of lessons related to different themes, and each lesson targets a particular pronunciation aspect. The participants used the website by logging in and performing the tasks included in the lesson. After completing each task, the participants get a score and completion rate. After the training, the researcher checked each participant's account, the scores they got in each lesson, and the tasks they finished. The following tables display the different lessons that the participants took part in.

The beginner level has nine different tasks; each task targets a specific pronunciation skill. Each task is scored out of 100% based on the participants' pronunciation, and then a general score is provided for this lesson; it also demonstrates the assignments completed (9 of 9 completed).

Table 17

First Training Session "Beginner Level"

Lesson	Participant	Lesson Score (percentages/ace score)	Tasks accomplished
Beginner Level	Camilla	96%	9/9
Beginner Level	Luna	90%	9/9
Beginner Level	Ella	90%	9/9
Beginner Level	Alexander	83%	9/9
Beginner Level	Levi	95%	9/9
Beginner Level	Mila	90%	9/9
Beginner Level	Aria	90%	9/9
Beginner Level	Scarlett	91%	9/9
Beginner Level	Penelope	90%	9/9
Beginner Level	Owen	67%	7/9
Beginner Level	Lorrie	81%	9/9
Beginner Level	Ezra	89%	9/9

Table 17 shows that the participants achieved high 80% and 91% scores. Additionally, it reveals that all participants completed this lesson's tasks except **Owen,** who only did seven out of nine due to unknown reasons.

The second lesson, "Vowel Mastery," includes five different tasks. Each task enables the participants to practice pronouncing a set of vowels.

Table 18

Second Training Session "Vowel Mastery"

Lesson	Participant	Lesson Score (percentages/ace score)	Tasks accomplished
Vowel Mastery	**Camilla**	89%	5/5
Vowel Mastery	**Luna**	94%	5/5
Vowel Mastery	**Ella**	89%	5/5
Vowel Mastery	**Alexander**	89%	4/5
Vowel Mastery	**Levi**	95%	5/5
Vowel Mastery	**Mila**	88%	5/5
Vowel Mastery	**Aria**	95%	5/5
Vowel Mastery	**Scarlett**	91%	5/5
Vowel Mastery	**Penelope**	93%	5/5
Vowel Mastery	**Owen**	95%	5/5
Vowel Mastery	**Lorrie**	92%	5/5
Vowel Mastery	**Ezra**	81%	4/5

Table 18 shows that the scores were higher in this lesson, ranging between 80% and 95%, because this lesson only involves pronouncing single words. Alexander and Ezra skipped one task, unlike all participants who accomplished this lesson's tasks.

The third lesson, "Consonant Mastery," includes six different tasks. In this lesson, participants are expected to pronounce a set of words to practice certain consonants.

Table 19

Third Training Session "Consonant Mastery"

Lesson	Participant	Lesson Score (percentages/ace score)	Tasks accomplished
Consonant Mastery	**Camilla**	92%	6/6
Consonant Mastery	**Luna**	94%	6/6
Consonant Mastery	**Ella**	95%	6/6
Consonant Mastery	**Alexander**	81%	2/6
Consonant Mastery	**Levi**	90%	6/6
Consonant Mastery	**Mila**	93%	6/6
Consonant Mastery	**Aria**	96%	6/6
Consonant Mastery	**Scarlett**	96%	6/6
Consonant Mastery	**Penelope**	95%	6/6
Consonant Mastery	**Owen**	96%	4/6
Consonant Mastery	**Lorrie**	77%	6/6
Consonant Mastery	**Ezra**	93%	5/6

Alexander, **Owen**, and **Ezra** skipped a few tasks in this lesson while all the participants accomplished them. In this lesson, Lorrie scored lower than all the other participants.

Four tasks are included in the fourth lesson, "Simple Sentence.". In this lesson, participants recorded themselves uttering a set of simple sentences. All participants got high scores and completed all tasks. However, **Aria** failed to do the tasks because she could not access an internet connection, so she did not get a score. Sometimes the participants skipped tasks because they did not have time or access to an internet connection.

Table 20

Fourth Training Session "Simple Sentences"

Lesson	Participant	Lesson Score (percentages/ace score)	Tasks accomplished
Simple Sentences	Camilla	92%	4/4
Simple Sentences	Luna	95%	4/4
Simple Sentences	Ella	96%	4/4
Simple Sentences	Alexander	93%	4/4
Simple Sentences	Levi	88%	4/4
Simple Sentences	Mila	93%	4/4
Simple Sentences	Aria	0%	0/4
Simple Sentences	Scarlett	95%	4/4
Simple Sentences	Penelope	95%	4/4
Simple Sentences	Owen	91%	3/4
Simple Sentences	Lorrie	86%	4/4
Simple Sentences	Ezra	92%	1/4

The fifth lesson, "Beginner Sentences," requires completing several tasks that include reading full sentences to get a score.

Table 21

Fifth Training Session "Beginner Sentences"

Lesson	Participant	Lesson Score (percentages/ace score)	Tasks accomplished
Beginner Sentences	**Camilla**	93%	8/8
Beginner Sentences	**Luna**	92%	8/8
Beginner Sentences	**Ella**	88%	8/8
Beginner Sentences	**Alexander**	93%	5/8
Beginner Sentences	**Levi**	86%	8/8
Beginner Sentences	**Mila**	92%	8/8
Beginner Sentences	**Aria**	93%	8/8
Beginner Sentences	**Scarlett**	92%	8/8
Beginner Sentences	**Penelope**	86%	8/8
Beginner Sentences	**Owen**	93%	6/8
Beginner Sentences	**Lorrie**	93%	7/8
Beginner Sentences	**Ezra**	88%	8/8

Table 21 shows that all participants achieved high scores and completed all tasks.

The sixth lesson, "Fluency," evaluates participants 'ability to listen to and answer questions. This lesson includes three tasks that tackle three themes (Sports, Travel, and People). Each participant should listen to the question and then record themselves answering it.

Table 22

Sixth Training Session "Fluency"

Lesson	Participant	Lesson Score (percentages/ace score)	Tasks accomplished
Fluency	Camilla	7.00/10	3/3
Fluency	Luna	6.50/10	3/3
Fluency	Ella	7.10/10	3/3
Fluency	Alexander	8.00/10	3/3
Fluency	Levi	7.00/10	3/3
Fluency	Mila	8.20/10	3/3
Fluency	Aria	7.50/10	3/3
Fluency	Scarlett	7.60/10	3/3
Fluency	Penelope	8.10/10	3/3
Fluency	Owen	6.50/10	3/3
Fluency	Lorrie	7.10/10	3/3
Fluency	Ezra	8.00/10	3/3

The table displays each participant's scores and the number of tasks they completed. The fluency score is not given in percentages but in a score out of 10. The criteria behind this scoring were mentioned in the previous chapter.

In the final lesson, participants recorded themselves reading a poem and getting a score on each verse.

Table 23

Seventh Training Session "Poem"

Lesson	Participant	Lesson Score (percentages/ace score)	Tasks accomplished
Poem	Camilla	95%	1/1
Poem	Luna	96%	1/1
Poem	Ella	95%	1/1
Poem	Alexander	88%	1/1
Poem	Levi	90%	1/1
Poem	Mila	91%	1/1
Poem	Aria	90%	1/1
Poem	Scarlett	95%	1/1
Poem	Penelope	91%	1/1
Poem	Owen	93%	1/1
Poem	Lorrie	88%	1/1
Poem	Ezra	86%	1/1

Table 23 displays each participant's scores and whether they completed the lesson. Most participants enjoyed this lesson and expressed that they enjoyed reading poems.

The researcher was granted access to their private accounts and witnessed their achievement on every task and lesson. All participants achieved high scores of 80% and 96%. The reason behind

these high scores is that the ASR website allows users many tries until reaching the desired correct pronunciation. The website gives the users unlimited tries; it enables them to repeat as many times as they want until reaching the correct articulation. Further, it allows them to listen to their recordings again, select the highest score they achieved, and provide corrective feedback on their pronunciation errors and how to correct them.

Further, in the previous tables, most participants participated in all tasks. However, some participants needed to complete specific tasks, but their overall score remained the same. The picture below demonstrates the tasks of a lesson.

Overall, the purpose of this phase was to monitor participants' development, evaluate their improvements, and check their scores because high scores imply that participants have gotten used to the website and are successfully doing the activities. In addition to these online accounts, a private WhatsApp group was used to keep up with the participants; more details about the WhatsApp group are discussed in the upcoming section.

4.2.2. Speaking Tests Results

To answer the main research question, to what extent can Automatic Speech Recognition ASR technology improve learners' pronunciation? The test scores were generated from the online website based on participants' pre-test and post-test recordings. The scores provided by the website were out of ten (the criteria behind these results were mentioned in the previous chapter) with detailed feedback. Both groups' scores were collected. Then, they were compared to check for a difference between them.

This section shows the results of the pronunciation scores achieved by the participants. The speaking tests done in this study enabled the researcher to evaluate the possible impact of the training on the experimental group. Since the participants' anonymity and confidentiality were respected, the names of the sampled students were changed with others of fictional character names from English movies.

Table 24

Control Group Pronunciation Scores

Group	Participant Pseudo	Pre-Test Score	Post-Test Score
Control Group	**Olivia**	7,2	7,1
Control Group	**Noah**	7,2	7,2
Control Group	**Ava**	7,4	7,0
Control Group	**Liam**	8,0	8,1
Control Group	**Sophia**	7,7	8,1
Control Group	**Amelia**	7,6	7,5
Control Group	**Mia**	7,3	6,9
Control Group	**Isabella**	7,8	7,9
Control Group	**Lucas**	7,0	7,0
Control Group	**Madison**	7,7	7,6
Control Group	**Alice**	7,6	7,6
Control Group	**Grace**	7,6	7,7

Graph 1

Control Group Pronunciation Scores
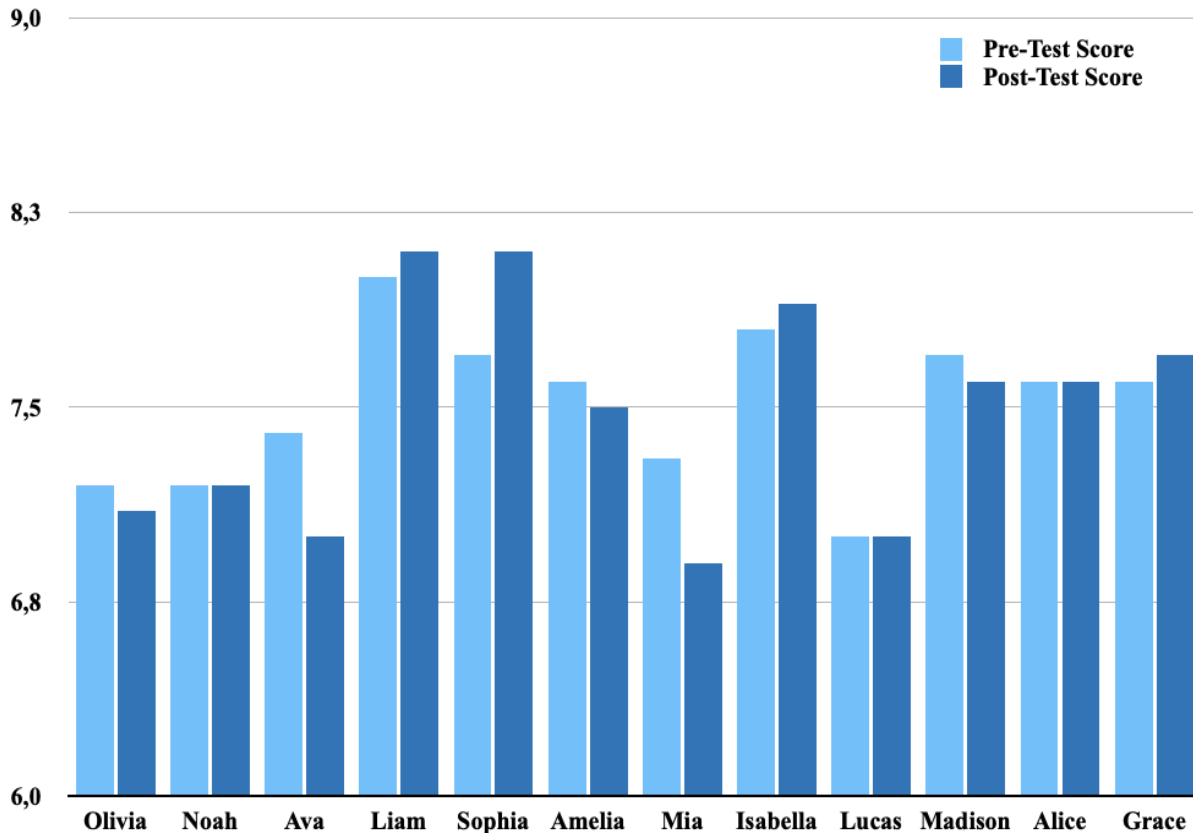
Table 24 and Graph 1 demonstrate the control group participants. Further, they show the pronunciation scores that they obtained in the pre-test and post-test. Their scores range between seven and eight, which indicates that their pronunciation level is good and they only had a few mispronunciations. However, the scores 'differences or improvements are analyzed in the next part.

Table 25

Experimental Group Pronunciation Scores

Group	Participant Pseudo	Pre-Test Score	Post-Test Score
Experimental Group	**Camilla**	8,0	8,4
Experimental Group	**Luna**	8,3	8,5
Experimental Group	**Ella**	7,5	8,2
Experimental Group	**Alexander**	7,8	7,7
Experimental Group	**Levi**	7,6	8,5
Experimental Group	**Mila**	7,8	8,3
Experimental Group	**Aria**	7,7	7,9
Experimental Group	**Scarlett**	8,4	8,5
Experimental Group	**Penelope**	7,8	8,2
Experimental Group	**Owen**	7,7	8,0
Experimental Group	**Lorrie**	7,7	8,1
Experimental Group	**Ezra**	8,3	8,4

Graph 2

Experimental Group Pronunciation Scores
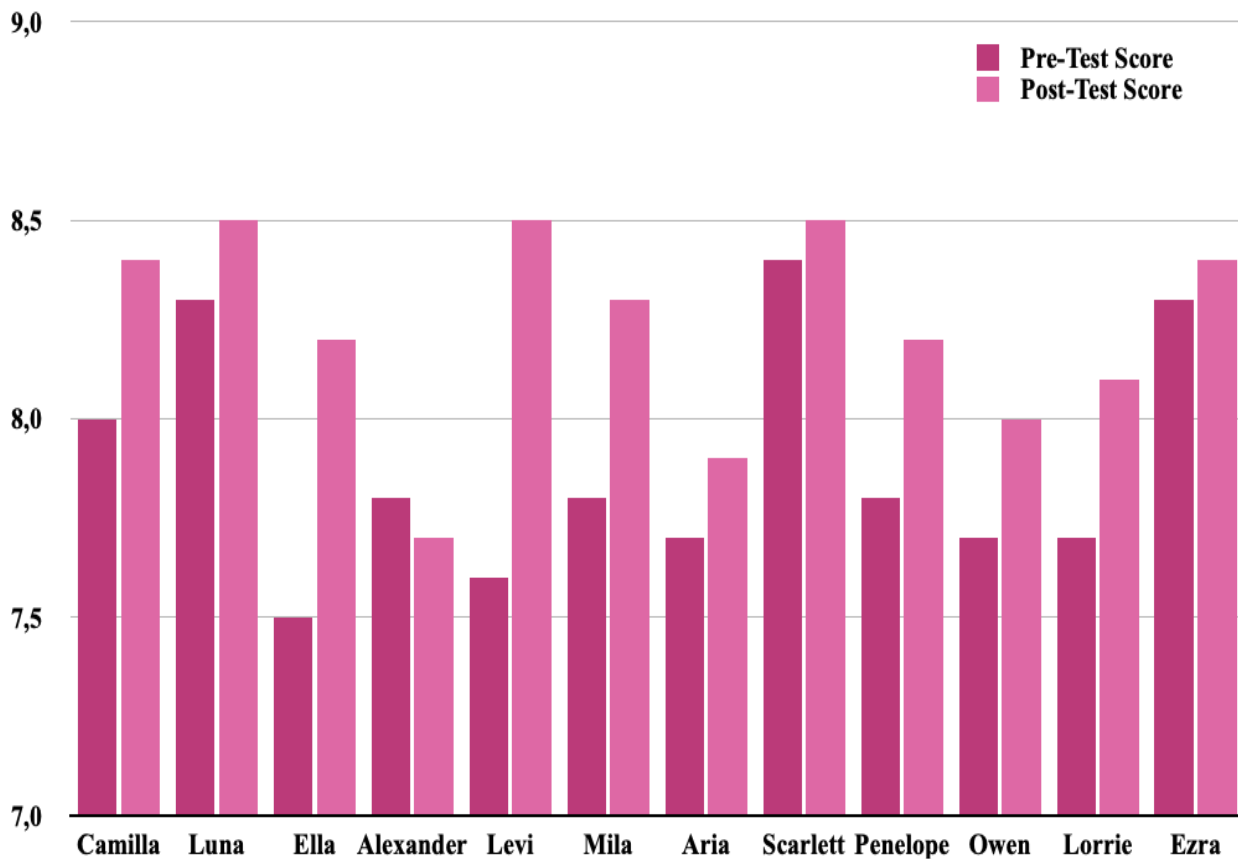
Table 25 and Graph 2 reveal the experimental group participants and the pronunciation scores obtained in the pre-test and post-test. Their scores range between seven and eight as well. However, it is noticeable that they got higher scores than those of the control group.

The tables and graphs indicate that the participants from each group mispronounced plenty of words, and even though they made pronunciation errors, they still got high scores. The website revealed their pronunciation was good and sometimes sounded like native speakers. The words that were mispronounced are shown in the appendices.

The tables and graphs show the pronunciation scores the participants got in the pre-test and the post-test. These scores were provided based on some criteria that are summarized in the statements below:

- Participants with scores ranging from 7 to 7.9 have reasonably good pronunciation and some accent.

- Participants with scores ranging from 8 to 8.9 have very good pronunciation and a mild accent.

- Participants with scores ranging from 9 to 9.9 have excellent pronunciation and sound similar to native English speakers.

The pronunciation scores achieved in the pre-test and post-test reveal that the control group test results indicate no notable difference between the pre-test and the post-test. There was a slight decrease in their scores. Their post-test scores were slightly lower than their pre-test. However, a somewhat more significant improvement was noted regarding the test scores of the experimental group participants. The results confirm that the training enabled the experimental group participants to improve their results during the post-test.

The test scores were copied into an Excel sheet to discover any difference between the pretest and post-test scores of the control and experimental groups. The researcher then compared the differences in the scores of pronunciation of the pre-test and the post-test for both groups. The following table reveals the differences between the pre-test and post-test scores.

Table 26

Pronunciation Score Differences (CG = Control Group , EX = Experimental Group)

Group	Participant	Pre-test Score	Post-test Score	Difference
CG	Olivia	7.20	7.10	-0.10
CG	Noah	7.20	7.20	0.00
CG	Ava	7.40	7.00	-0.40
CG	Liam	8.00	8.10	0.10
CG	Sophia	7.70	8.10	0.40
CG	Amelia	7.60	7.50	-0.10
CG	Mia	7.30	6.90	-0.40
CG	Isabella	7.80	7.90	0.10
CG	Lucas	7.00	7.00	0.00
CG	Madison	7.70	7.60	-0.10
CG	Alice	7.60	7.60	0.00
CG	Grace	7.60	7.70	0.10
EG	Camilla	8.00	8.40	0.40
EG	Luna	8.30	8.50	0.20
EG	Ella	7.50	8.20	0.70
EG	Alexander	7.80	7.70	-0.10
EG	Levi	7.60	8.50	0.90
EG	Mila	7.80	8.30	0.50
EG	Aria	7.70	7.90	0.20
EG	Scarlett	8.40	8.50	0.10
EG	Penelope	7.80	8.20	0.40
EG	Owen	7.70	8.00	0.30
EG	Lorrie	7.70	8.10	0.40
EG	Ezra	8.30	8.40	0.10

Table 26 shows a slight difference between the scores of the tests, either positive or negative. After calculating the differences between the test scores, they were exported into SPSS software to be analyzed statistically and represented graphically.

Before conducting any statistical analysis, it is crucial to determine whether sample data have been drawn from a normally distributed population. The normality assumption requires that the population from which the two samples were taken show a normal distribution (Tsagrisa and Pandis, 2021). Since the researcher selected a small sample size, a Shapiro-Wilk test was performed in the statistical SPSS software to test for normality. The Shapiro–Wilk test was chosen because it is used with small sample sizes (less than 50). The null hypothesis states that when the p-value is more significant than (0.05), it indicates that the data were taken from a normally distributed population.

Table 27

Shapiro-Wilk Test for Normality (Pronunciation)

	Statistic	**Df**	**Sig.**
Pre-test	0,955	24	0,349
Post-test	0,918	24	0,054

Table 27 shows the results obtained after conducting a Shapiro-Walk test on SPSS software. The p-value is more significant than (0.05), allowing me to accept the null hypothesis that the data were drawn from a normally distributed population. Then, more statistical analysis was conducted on the pre-posttest results.

The researcher ran an independent t-test to get both groups' mean and standard deviation of the pre-test and post-test. The t-test is the most effective test to compare different groups' means since it determines whether the two groups have a statistically significant difference.

Table 28

Overall Pronunciation Results (CG = control group, EG = experimental group)

Group	N	Pre- Test		Post- Test	
		Mean	Std. Deviation	Mean	Std. Deviation
CG	12	7.508	0.2906	7.475	0.4309
EX	12	7.883	0.2985	8.225	0.2598

Table 28 demonstrates the interpretation of the results obtained. Histogram 3 shows the difference between the mean and standard deviation scores of the pre-test and post-test of both groups.

Graph 3

Pronunciation Mean and Std.Deviation of the Pre-Posttest (CG = Control Group , EG = Experimental Group)
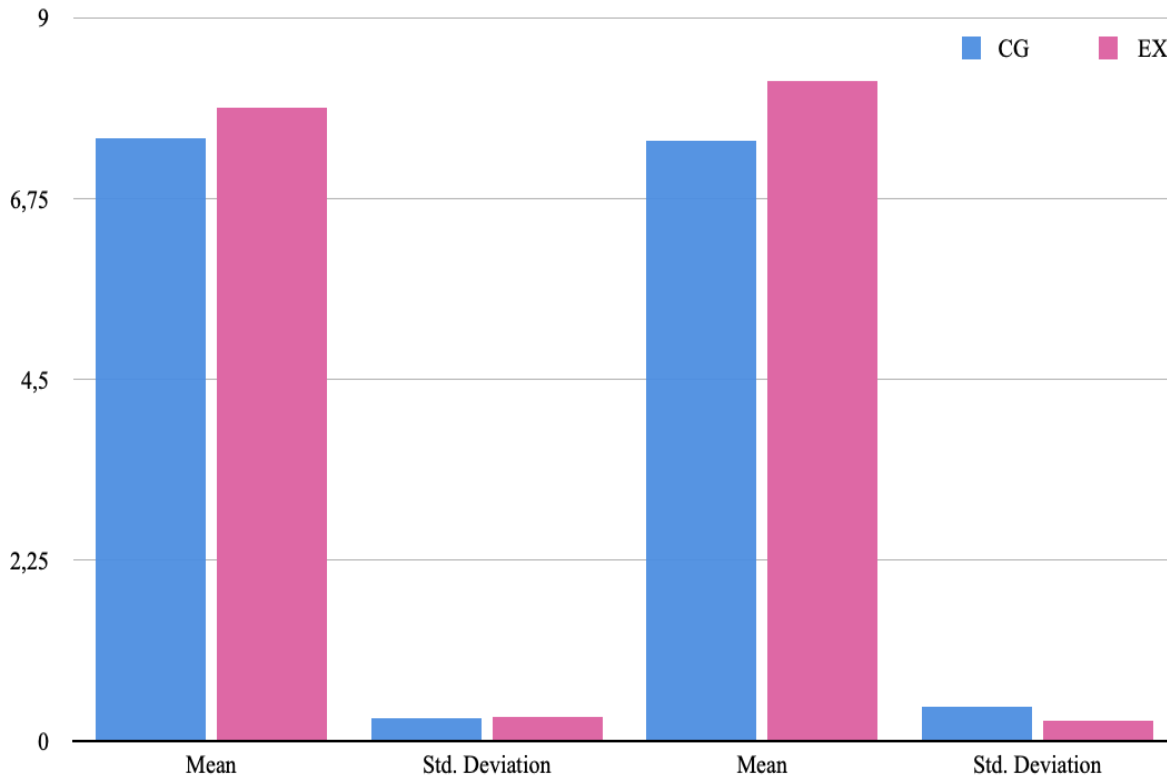
As shown in Table 28 and Graph 3, on the one hand, the treatment group participants (i.e., the experimental group) made slightly better overall pronunciation scores after participating in the ASR training. This was demonstrated in the mean results difference between the pre-test and post-test. On the other hand, the no-treatment group (i.e., the control group) showed no improvements since they did not participate in any training and carried on their regular lectures. The tables and graph show a slight drop in their results between the pre-test and post-test.

To give more evidence and confirm the results mentioned earlier, the researcher ran an independent t-test on SPSS Statistics, an inferential statistical test followed to determine whether

there is a statistically significant difference between the means of the two groups. This test was used to determine if there is a statistically significant difference between the two groups (i.e., the control and the experimental group results in the pre-test and post-test).

Table 29

Analysis of Pronunciation Results

Group	N	Mean	Std. Deviation
CG	12	0.0333	0.21881
EG	12	0.3417	0.27455

Table 29 displays the mean difference between the two groups pronunciation scores on the pre-test and the post-test and the standard deviation, the t-statistic, and the p-value of the pronunciation scores.

Conducting an independent samples t-test allowed me to assume the variances of the two groups and whether they were equal in the population. The researcher performed Levene's Test of Equality of Variances on SPSS Statistics software to do so. Levine's test entails that the variances are insignificant if the p-value is more significant than 0.05. However, if the p-value is less than 0.05, then there is a considerable difference between the variances.

Table 30 demonstrates the results obtained from the homogeneity test of variances, including the F-statistic and the significance value (p-value).

Table 30

Levene's Test for Equality of Variances (Pronunciation)

	Levene's	Test
	f	**Sig.**
Pre-test —- Post-test	0,675	0,420

The significance value is (p = 0.420), which is greater than 0.05 (i.e., $p > .05$), indicating that the group variances are equal. This enabled me to accept the null hypothesis entailing that the groups are not significantly different.

Since the homogeneity of variances was met, I then calculated the independent samples t-test on SPSS Statistics software.

Table 31

T-test of Equality of Means (Pronunciation)

	t	**Df**	**Sig.**
Pre-test —- Post-test	3.700	22	0,001

Table 31 displays the t-statistic (t), the degrees of freedom (df), and the significance value (p-value).

The level of significance observed in the results obtained after the analysis is that the p-value is significant at (p = .001), which is less than the probability value of 0.05 p < .05. This demonstrated a statistically significant difference between the two groups' results. Since there is a statistically significant difference between the two groups, I can preliminarily say that the null hypothesis H0 is unsupported, which means we accept the alternative hypothesis H1. This is an indication that the

participants in the experimental group have better achievement than the participants in the control group. The results in Table 31 show that the pre-test and post-test scores are different. The differences are not caused by chance but are due to the instructional ASR training that the participants from the experimental group underwent during the second semester. More score details are demonstrated in the appendices.

Although the independent t-test showed significant results, several international journals no longer accept statistical significance and favor effect size since it is thought to be more relevant than significance (Olejnik and Algina, 2000; Capraro and Capraro, 2002; Thompson, 2002). Hence, for further confirmation of the results, I performed an effect size differential measure because it is more valuable than statistical significance. An effect size is "simply a way of quantifying the difference between two groups. For example, if one group has had an 'experimental treatment' and the other has not (the 'control'), then the Effect Size is a measure of the effectiveness of the treatment" (Coe 2000:1, Quoted in Cohen et al. 2007: 521). Hence, to discover how significant the effectiveness of ASR training on students, I measured the effect size of the treatment by following these steps:

Coe (2000 in Cohen et al. 2007) suggested using a 'pooled' estimate of standard deviation because it is believed to be more accurate in calculating the effect size of two groups. To count the pooled standard deviation of the experimental and the control group, I used the following formula.

$$SD_{pooled} = \sqrt{\frac{(N_E - 1)SD_E^2 + (N_C - 1)SD_C^2}{N_E + N_C - 2}}$$

(N_E = number in the experimental group, N_C = number in the control group, SD_E = standard deviation of the experimental group, and SD_C = standard deviation of the control group).

Then, the effect size was calculated following the formula that was suggested by Muijs (2004, 136)

$$\frac{\text{(Mean of the experimental group - mean of the control group)}}{\text{Pooled standard deviation}}$$

I use the results obtained from the t-test to count the effect size of the treatment.

The effect size of the treatment is = 1.27

Cohen (1988) suggested that the effect size can range from 0 to 1. If Cohen's d effect size lies between 0 to 0.20, it has a weak effect; between 0.21 and 0.50, it has a modest effect, 0.51 and 1.00, its effect is moderate, whereas >1.00 indicates that it has a strong impact. Hence, I obtained the effect size value (1.27), which shows a powerful effect of the ASR instructional training on participants' pronunciation scores in the post-test. The statistically significant value of the effect size determined that pronunciation score differences between both groups were due to the ASR training that the participants had gone through.

The ASR technology also provided the participants with pronunciation scores and scores on their fluency (The criteria behind fluency scoring is mentioned in the previous chapter). The fluency scores were collected and analyzed.

Table 32

Control Group Fluency Scores

Group	Participant Pseudo	Pre-Test Score	Post-Test Score
Control Group	Olivia	6,1	6,5
Control Group	Noah	6,4	6,5
Control Group	Ava	6,3	6,5
Control Group	Liam	7,8	7,4
Control Group	Sophia	6,6	6,9
Control Group	Amelia	6,5	6,5
Control Group	Mia	6,3	5,9
Control Group	Isabella	6,5	6,6
Control Group	Lucas	6,4	6,5
Control Group	Madison	6,4	6,1
Control Group	Alice	5,5	6,0
Control Group	Grace	7,0	7,0

Graph 4

Control Group Fluency Scores
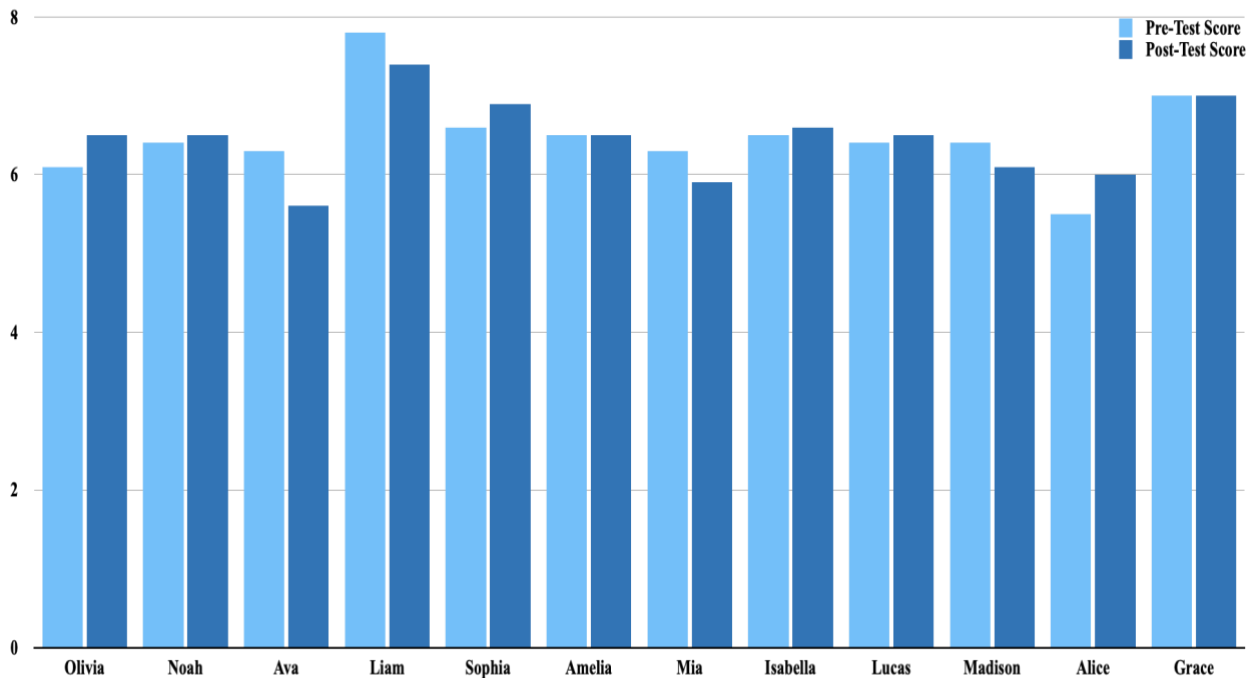
Table 32 and Graph 4 show the control group participants' fluency scores obtained in the pre-test and the post-test. Their fluency scores were lower than their pronunciation scores since the criteria behind assessing fluency is different than pronunciation. Their scores ranged between six and eight.

Table 33

Experimental Group Fluency Scores

Group	Participant Pseudo	Pre-Test Score	Post-Test Score
Experimental Group	**Camilla**	6,0	7,5
Experimental Group	**Luna**	7,5	8,4
Experimental Group	**Ella**	6,2	7,6
Experimental Group	**Alexander**	6,0	6,5
Experimental Group	**Levi**	5,8	8,0
Experimental Group	**Mila**	7,3	8,4
Experimental Group	**Aria**	7,2	7,5
Experimental Group	**Scarlett**	7,7	8,0
Experimental Group	**Penelope**	6,0	7,5
Experimental Group	**Owen**	7,0	7,0
Experimental Group	**Lorrie**	5,5	6,0
Experimental Group	**Ezra**	7,5	7,5

Graph 5

Experimental Group Fluency Scores
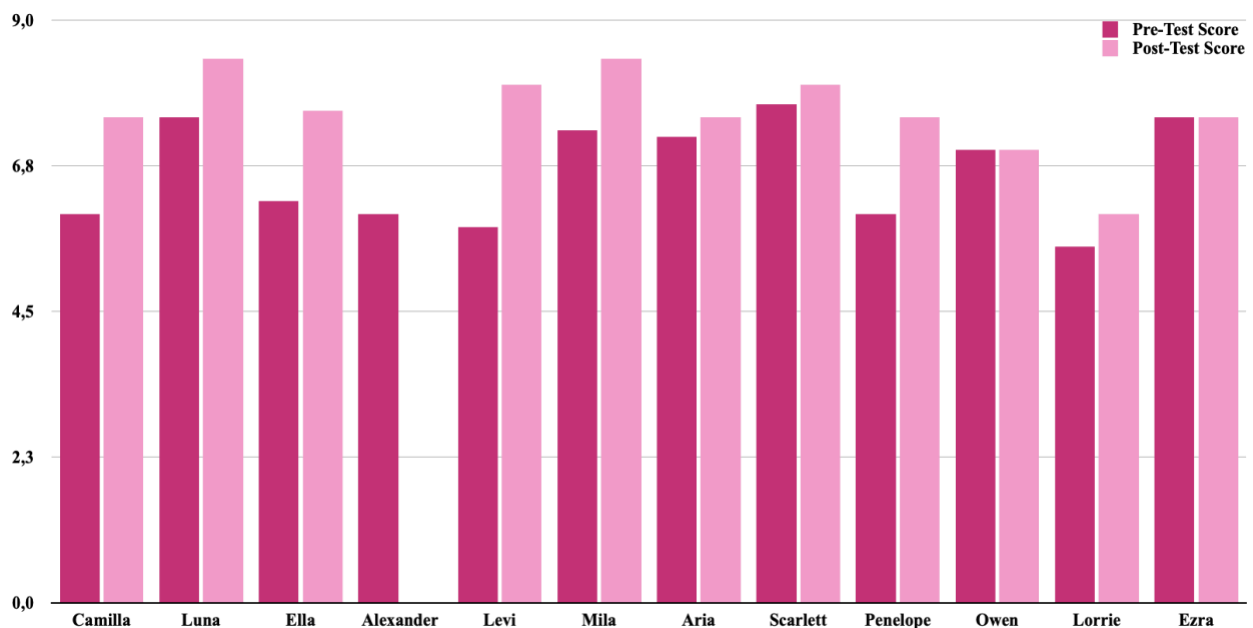
Table 33 and Graph 5 show the experimental group participants' fluency scores obtained in the pre-test and the post-test. Their scores ranged between five and eight. They made slightly lower scores than those of pronunciation. This is because achieving a native-like fluency of speech is difficult and requires extensive practice.

The tables and graphs indicate the participants' pre-test and post-test fluency scores. These scores were provided based on some criteria that are summarized in the statements below:

o Participants with scores ranging from 5 to 5.9 demonstrate below-average fluency and coherence. They need remedial training in speaking fluently and language construction.

o Participants with scores ranging from 6 to 6.9 demonstrate reasonable fluency and coherence. They sometimes hesitate while speaking.

o Participants with scores ranging from 7 to 7.9 generally demonstrate good fluency and coherence while speaking but take occasional pauses.

o Participants with scores ranging from 8 to 8.9 demonstrate excellent fluency and coherence with occasional pauses.

For further analysis of the participants 'scores, their speeches were recorded and saved on the ASR account of each participant. The researcher had access to each participant's voice recording. In addition to the recordings, the data were transcribed by the website. All parts of words and complete words articulated by the participants in a minute, the filled pauses as hesitation and self-correction or silent pauses committed by each participant were included in the transcription. The words spoken per minute are the speed at which the participants talk; it is calculated by adding the number of words in their speech and dividing the total number of words by the number of minutes taken. Moreover, the pauses and self-repetitions were highlighted in triangle pause icons.

Table 34

Fluency Analysis (CG = Control Group, EG = Experimental Group)

		Pre-test		Post-test	
Group	Participant	Words per minute	Number of pauses taken	Words per minute	Number of pauses taken
CG	Olivia	103	9	107	6
CG	Noah	104	6	106	7
CG	Ava	94	8	112	7
CG	Liam	91	14	98	12
CG	Sophia	110	8	110	7
CG	Amelia	94	10	100	10
CG	Mia	91	13	94	8
CG	Isabella	106	5	91	15
CG	Lucas	106	7	71	19
CG	Madison	106	8	109	7
EG	Alice	84	12	120	8
EG	Grace	108	10	110	8
EG	Camilla	95	9	131	5
EG	Luna	110	2	120	4
EG	Ella	106	10	124	8
EG	Alexander	107	5	120	4
EG	Levi	120	7	139	4
EG	Mila	110	2	138	5
EG	Aria	78	12	94	9
EG	Scarlett	120	5	112	6
EG	Penelope	94	10	110	8
EG	Owen	109	7	124	6
EG	Lorrie	76	17	96	10
EG	Ezra	99	8	131	5

Table 34 shows the words the control and experimental group participants articulated per minute and the number of pauses they took while recording their speech in the pre-test and the post-test.

The scores in Table 34 differ from the average uttered words per minute and the filled pauses. The participants received high fluency scores despite frequently pausing while answering the questions and speaking inaccurately. Hieke (1981), MacGregor, Corley & Donaldson (2009) argued that when it comes to fluency, any pause, hesitation, self-repetition, or self-correction indicates cognitive disfluency. However, Davies (2003) Schmidt (1992), and Segalowitz (2010) claimed that speakers might employ self-repetition and self-corrections as a discourse strategy to emphasize or clarify a piece of information for the benefit of the listener. Similar to self-repetition, pausing can also be employed as a discourse strategy.

When the filled pause ratio is calculated, lexical filled pauses, which are frequently considered an indication of dysfluency, should be considered. The breaks are occasionally made to gain more time for planning and producing the following utterances. Fluency is measured by the efficiency and automaticity of processing rather than the speaking rate. As a result, the speech rate and the pauses were ignored, and the fluency scores were used for further analysis.

The fluency scores were then copied into an Excel sheet, and the scores of each participant were compared (i.e., the pre-test score was compared with the post-test score).

Table 35

Fluency Scores Differences (CG = Control Group, EG = Experimental Group)

Group	Participant	Pre-test Score	Post-test Score	Difference
CG	Olivia	6.10	6.50	0.40
CG	Noah	6.40	6.50	0.10
CG	Ava	6.30	5.60	-0.70
CG	Liam	7.80	7.40	-0.40
CG	Sophia	6.60	6.90	0.30
CG	Amelia	6.50	6.50	0.00
CG	Mia	6.30	5.90	-0.40
CG	Isabella	6.50	6.60	0.10
CG	Lucas	6.40	6.50	0.10
CG	Madison	6.40	6.10	-0.30
CG	Alice	5.50	6.00	0.50
CG	Grace	7.00	7.00	0.00
EG	Camilla	6.00	7.50	1.50
EG	Luna	7.50	8.40	0.90
EG	Ella	6.20	7.60	1.40
EG	Alexander	6.00	7.60	1.60
EG	Levi	5.80	8.00	2.20
EG	Mila	7.30	8.40	1.10
EG	Aria	7.20	7.50	0.30
EG	Scarlett	7.70	8.00	0.30
EG	Penelope	6.00	7.50	1.50
EG	Owen	7.00	7.00	0.00
EG	Lorrie	5.50	6.00	0.50
EG	Ezra	7.50	7.50	0.00

The fluency scores differences between the pre-test and post-test of the experimental and control groups are shown in Table 35.

Looking at Table 35, one can preliminarily notice that, unlike pronunciation scores, participants from each group displayed more significant differences in their fluency scores.

Then, the researcher wanted to determine whether sample data had been drawn from a normally distributed population. A Shapiro-Wilk test was performed in the statistical SPSS software to test for normality.

Table 36

Shapiro-Wilk Test for Normality (Fluency)

	Statistic	**Df**	**Sig.**
Pre-test	0,946	24	0,221
Post-test	0,955	24	0,341

Table 36 shows the results obtained after conducting a Shapiro-Walk test on SPSS software using the fluency test results of the pre-test and the post-test. It shows that the p-value is more significant than (0.05) (p = 0.221, p = 0.341). This enabled me to accept the null hypothesis and confirm that the data were drawn from a normally distributed population.

Then, the fluency scores were exported to SPSS Statistics software to be analyzed statistically and represented graphically. At first, I ran an independent t-test to calculate both groups' mean and standard deviation of the pre-test and post-test.

Table 37

Overall Fluency Results (CG = Control Group, EG = Experimental Group)

		Pre-	Test	Post-	Test
Group	N	Mean	Std. Deviation	Mean	Std. Deviation
CG	12	6.843	0.5408	6.458	0.5035
EG	12	6.642	0.7925	7.583	0.6436

The fluency results obtained are interpreted in Table 37. Then, a histogram was created to demonstrate the difference in the means and the standard deviation of the pre-test and post-test of both groups (i.e., the control and experimental groups).

Graph 6

Fluency Mean and Std.Deviation of the Pre-Posttest (CG = Control Group, EG = Experimental Group)
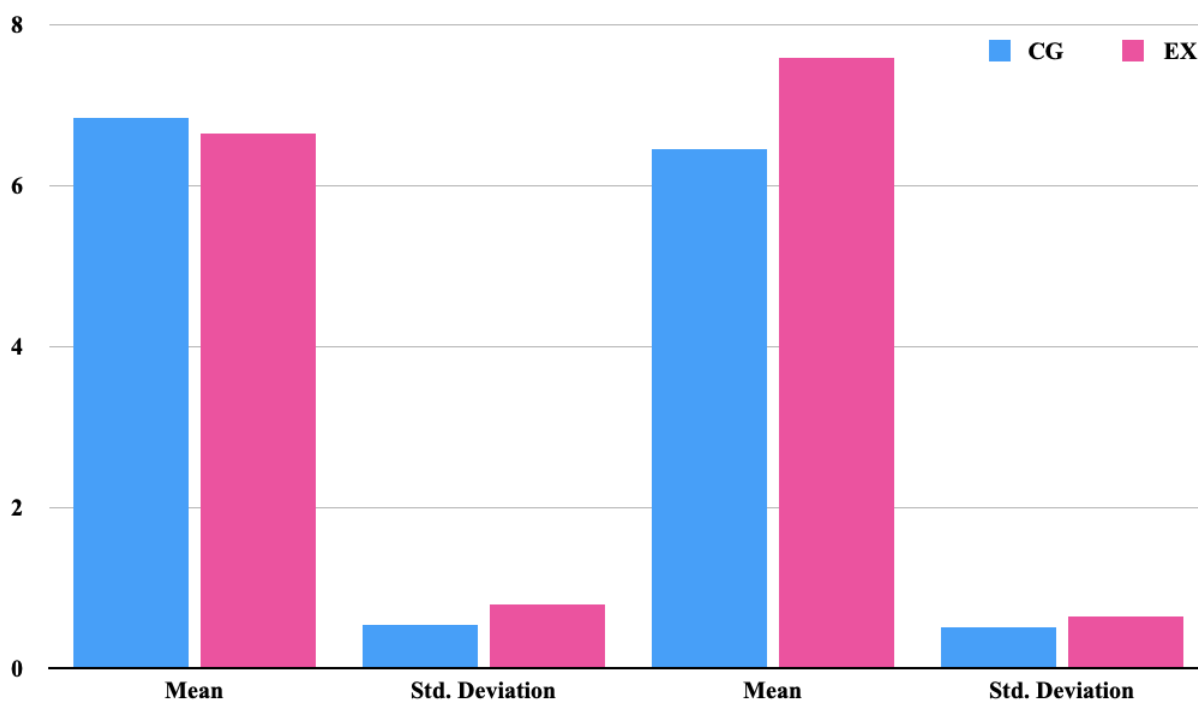
The data in Table 39 and Graph 6 demonstrate the progress in the experimental group participants' pre-test and post-test scores. It is expressed in the difference in the means (6.642 vs. 7.583). Based on these data, I can preliminarily claim that the ASR training that the participants from the experimental group had gone through led to more significant outcomes. However, the participants from the control group did not improve their fluency outcomes (6.843 vs. 6.458). One might claim that the control group underwent no treatment or training besides regular classes.

Table 39

T-test of Equality of Means (Fluency)

	t	Df	Sig.
Pre-test —- Post-test	4.171	22	0,001

Table 39 demonstrates the t-statistic (t), the degrees of freedom (df), and the significance value (p-value). The significance level (p = 0.001) is less than the probability value of 0.05 (p < .05), which indicates that there is a statistically significant difference between the two groups' fluency results. Since there is a statistically significant difference between the two groups, I can preliminarily say that the null hypothesis H0 is unsupported, which means we accept the alternative hypothesis H1. This is an indication that the participants in the experimental group achieved better fluency scores than the participants in the control group. Thus, I can confirm that the instructional ASR training and treatment that participants from the experimental group underwent caused the differences between the two groups.

4.2.3. WhatsApp Private Groups

The researcher created private WhatsApp groups to answer the research question regarding participants' attitudes toward using the ASR website. WhatsApp was chosen since it is a messaging app that has end-to-end secure messaging, is rich with features, and is easy to use. In addition, all participants already have a WhatsApp account on their phones. These groups served as a technological tool that allowed us to track participants' progress and provide them with an online platform that enabled them to communicate with the researcher and each other. Hence, I monitored the participants daily, taught them how to use the website, and helped them with their issues during the training. Further, the groups offered the participants a more private environment to share their ideas, opinions, and thoughts on the ASR training.

After the pre-test, the researcher created the WhatsApp groups and added the participants. Two groups were created, one for the control group and the other for the experimental group (WhatsApp requires having the private number of the user, which was collected through the information card after receiving each participant's agreement). However, only the private WhatsApp group messages of the experimental group were collected and analyzed since the WhatsApp group that belongs to the control group only served as a tool to inform them of the post-test schedule. First, instructions and explanations were given on how to use the ASR website, creating an account, and accessing the activities. The participants' messages were classified into positive or negative attitudes and issues faced. These classifications are discussed below.

4.2.3.1. Positive Messages. The messages collected from the WhatsApp group showed that the participants' attitudes toward the ASR website were positive. The conversations with the participants indicated that most participants enjoyed practicing using the website. Initially, the participants expressed fears about using the website, especially those who used their cell phones. However, they became more interested once they learned how to use it. After every lesson, the participants left positive messages on the group reflecting on how they enjoyed practicing their pronunciation and sharing screenshots and video shots of their training. One participant said: '… I'm enjoying this!' Others said: '… me too'. After each lesson, I would send messages to the WhatsApp group inquiring about participants' progress and ask them to share their thoughts on the website. Hence, participants would reply, ' Yes, I finished all the tasks.' And the others respond: 'I finished the first three tasks.', 'I completed the first lesson.', and 'Four tasks are done.' Since the group provided the participants with a private environment to express themselves freely and without limitations, some participants sent messages using their mother tongue (Arabic). Since most internet users are accustomed to utilizing their mother tongue when using online platforms, I did not impose any restrictions when communicating in the group. Hence, I aimed to allow the participants to feel comfortable and not judged when expressing their thoughts and feelings concerning the training.

Further, in the group messages, I played the role of the motivator. I continued to send messages to the participants to encourage daily practice, check their progress, and get feedback on the website. The statements contained questions like: 'Did you finish the first task?', 'how many tasks did you accomplish?', 'Are you enjoying the website?' and 'Did you find the website helpful?'. Afterward, I would respond to the participants' replies with encouraging messages such as: 'Great

job!', 'Nice!', 'Keep on practicing.', 'amazing job!', 'I'm so proud of you.' WhatsApp has a feature that allows other users to react to your messages using certain emojis, so participants would leave a 'like emoji' or 'heart emoji' on my messages indicating that they appreciate the support and the compliments.

      4.2.3.2. Technical Problems' Messages. Another pertinent theme in the participants' messages was technical problems. As mentioned before, the participants could speak freely in a private setting without worrying about being judged, and the researcher could see if they were experiencing any issues that would prevent them from completing the training. Although there was a generally good attitude toward the training website, some participants faced problems at the beginning of the training. Once the training was launched, three participants (Alexander, Penelope, and Scarlett) sent messages indicating that the website was not working on their phones. The messages were: 'Actually, I faced a problem.', 'The website is not working on my phone.' A participant sent a screenshot, 'When I try to record myself, I get this.' Participants were requested to provide a screenshot of the issue, and it revealed that the website was displaying a red screen with the message: "We are unable to use your microphone." I used multiple devices, including my laptop and phone, to access the website, try it, and check if it generally functions to avoid any problems later in the study. However, participants faced an issue and were frustrated that the website did not operate on their phones, and I was equally as frustrated. To solve the problem, I created an account, logged in to the website using my phone (iPhone XR), noticed that the website worked perfectly, and recorded my speech without any problem. Then, I tried to use it on another phone (an Android device) and noticed that on the top left of the screen, the user should allow the website to access their microphone to enable them to record themselves. Hence, I screen-recorded

the phone when allowing the website to access the microphone and speaker and I sent it to the participants. After a short while, the participants said that the website was functional after following my instructions. They left messages saying, 'It is finally working for me,' and 'Now it is working, thank you.' All the screenshots of the messages sent by me and the participants are in appendices.

4.2.3.3. Negative Messages. As indicated, the participants left positive comments about the website; their messages revealed that using the ASR website to practice pronunciation was fun and helpful. However, it was only sometimes accurate. The participants left frustrated messages when the website gave them low pronunciation scores. They sent messages saying, 'Oh! I'm getting low scores." me too!' And they start having conversations with each other about this. A participant said, 'I don't like it when I get 59.' These messages demonstrate that the participants tried their best to achieve higher training scores, but the website gave them lower scores for some reason. This is because numerous other factors may interfere while using an ASR tool. These factors may be related to background noise, a poor microphone, or a slow internet connection, which can prevent the website from successfully saving the sound. However, I continued providing support and encouraging them to keep training and that they should give little importance to the scores because the primary goal of this training was to assist them in improving their pronunciation and the scores are just a form of guidance and do not affect their level in general.

Overall, the WhatsApp group was a helpful tool that allowed me to maintain regular contact with all the participants. It offered them a secure place to share their opinions, ask questions about the website, and get assistance with any challenges. Considering that none of the participants had

received similar pronunciation training, the messages and reactions revealed that they enjoyed using the ASR website to practice pronunciation. This tool also helped increase the participant's awareness of using technological tools to improve their pronunciation.

4.3. Analysis of the Post-training Interview

A semi-structured interview was conducted with the experimental group of 12 participants to answer the third research question. It included questions regarding the training and the ASR website to elicit their perceptions and thoughts. The interview is considered a conventional data-gathering tool used in qualitative research. Merriam & Tisdell (2015) claimed that using the qualitative design is highly beneficial as it allows the researchers to gather more data. Additionally, Kvale (1996) argued that: "using interview techniques, the researcher commonly aims to obtain the perspective of the interviewee by interpreting the meaning of the described phenomena" (p. 6). I employed the interview to gain insight into participants' opinions and attitudes concerning the training. Participants' perceptions of online ASR tools, the ASR website scoring and feedback, obstacles, and technical problems while using the ASR website are the main themes under which the data in this section is organized.

Table 40

Interview Themes

Sections	Themes
The Website Effectiveness	- Access to the website - Lessons and activities available on the website - The website scoring and feedback
Pronunciation and Fluency Improvments	- Obstacles and technical problems that were faced while using the website - Participants improved their pronunciation - Participants improved their fluency
Participants' Feelings and Thoughts on the Website	- Conventional learning or online learning? - Technological / Online Platforms and Tools Usage in the Future - Teachers' internet usage

Table 40 shows how the participants 'answers were organized and sorted into sections and themes.

First, I wanted to collect participants' answers regarding the website's effectiveness, so we asked them questions revolving around the website access, the activities available on the website, the feedback and scores provided, and the deficiencies of accessing and using the website.

Concerning website access, all participants stated that access to the online website was easy. However, they needed to be made aware that they should create accounts, and once the

researcher explained the process, they used their emails and passwords to create their accounts and easily log into the website.

I asked the participants whether they liked the activities they participated in during the training. All participants said that they enjoyed the activities on the website and found them very beneficial. Camilla, Scarlett, Mila, Levi, Alexander, Owen, Ezra, and Lorrie expressed that they enjoyed all the activities they took part in, quoting them, "I enjoyed all the activities." Another believed that the activities were straightforward to accomplish. Ella said, "I enjoyed all the activities because they were effortless."

Other participants had unique preferences. They felt that some activities were particularly fascinating. Aria and Penelope said, "The activities I enjoyed most were the poem and the dialogue." In contrast, Luna said, "I enjoyed the listen and repeat activities. Knowing how to pronounce the word and then trying to pronounce it myself is good. It is beneficial." The responses given by the participants show that the pronunciation and fluency activities that the participants took part in were beneficial, enjoyable, and easy to accomplish.

Participants were then asked questions regarding the website feedback during the test and the training and if they felt that the website had fairly and accurately measured their pronunciation and fluency levels. Scarlett, Penelope, Owen, and Lorrie thought "the scores were too easy to understand and were very clear." Camilla said, "They were accurate and helped me improve each time." The scoring algorithm of the website was easy to interpret, and the participants understood how it works perfectly. As Luna highlighted, "The scores were clear and precise. It gives you a

score out of 10. And sometimes, when you don't pronounce a word correctly, it highlights it in red and gives you the appropriate pronunciation."

The participants were also asked whether they found the feedback accurate and helpful. Alexander, "They were good techniques and helped to level up." Mila, "They were very accurate and showed me where I made mistakes and how to correct them." And Ezra found the feedback to be precise and accurate "Because the percentages were accurately determining my grades." However, some participants expressed frustration because they thought the website was sometimes unfair. Ella said, "Very effective, it helps me to check my mistakes and to see my scores and level, actually that's helped me a lot to discover a lot of things, but sometimes I think that the scores were not fair." And Aria said, "For me, I didn't find the scores fair. Every time I pronounce a word, I don't get a good score." As well as Levi, "Sometimes I pronounce correctly, but then I get low scores, and it makes me sad."

Positive feedback was received from all respondents indicating that the website was simple to use on both PCs and mobile devices and that they faced no obstacles or technical issues when using it. Overall, the participants' impressions of the website's effectiveness revealed that it was beneficial and worked perfectly on their devices. The guided activity set was enjoyable, helpful, and highly developed. Additionally, the scores and feedback were fair, accurate, and appropriately measured their improvements. Also, the feedback provided was motivating and made them improve their pronunciation and fluency.

Then I wanted to investigate whether the participants thought the website helped them improve their pronunciation and fluency. They highlighted the extent to which they believed the

website was functional and practical. They said the website and the training helped them improve their pronunciation and fluency. The website has a wide range of activities exposing participants to new vocabulary. Hence, Levi said, "It helped me learn new words." And Aria highlighted, "The training helped me improve my vocabulary."

The website allows the users to record themselves, listen to their recordings, and re- record themselves if they are not satisfied with the score. Accordingly, this kind of repetition is a valuable technique for improving pronunciation. Therefore, Luna expressed, "I saw improvements in my pronunciation because of the listen and repeat activities." And Penelope, "I improved my pronunciation because of repeating." Further, all the other participants spoke positively about how it helped them improve their speaking skills. Alexander said, "The training makes you feel like you are a native speaker." Hence, participating in such training can help students improve their speaking level.

Interview answers revealed participants' perceptions of the website and their attitudes and thoughts on using online tools to develop their skills and abilities. All participants said they preferred online learning to traditional learning since it was more convenient and available whenever and wherever they wanted. Participants were also questioned about their motivation to use more technological tools to develop specific language skills following the training. Most respondents said they would use online tools since they are convenient and accommodating. Further, each participant was requested to state what kind of tool they were willing to use. Levi, "Yes, it does make me motivated to use social media like Facebook and YouTube because it helps me to understand the lessons very well with contacting the teacher." Mila, "Yes, I will use all the

apps that can help me improve my skills." And Luna, "Sure, I want to try other technological tools and applications that make me learn and develop my level with a funny way; I would like to use the technology that helps me to improve my speaking skills and my listening because I like these two skills." While Alexander said, "Yes, I will use more technological tools in the future. There are plenty of Apps on the phone that can help me improve my speaking skills and grammar and learn new vocabulary." Also, Lorrie expressed his feelings towards the website used in training and how he enjoyed this kind of technology "Yes, it is a high technology and very beneficial." All answers were very positive, and all respondents confirmed that they would for sure use technological tools that can help them improve their skills and abilities.

A respondent stated, "I am not sure about it, but I don't have a specific one. If I find it helpful, I'll use it." Accordingly, many EFL students must be aware of the numerous internet resources, including tools, websites, and applications they may use to improve their skills, specifically speaking. Additionally, most respondents wished their teachers would use technology more in the classroom since it has numerous advantages, might save them time and effort, and help them develop their skills more contemporary manner. Luna, "Teachers could use blended learning, virtual classrooms, and educational websites to make learning more fun and appealing." Alexander, "I wish my teachers used PowerPoint and games, give us internet homework, and use tablets in the classroom." All participants know their teachers could use millions of options and tools to make learning more accessible and fun. Further, other respondents praised the website operated in training. They said they would continue using it in the future and would also prefer if their teachers could recommend other "listen and speak" tools to help them improve their speaking and pronunciation.

Overall, the findings obtained from interviewing the experimental group participants confirm the intervention's noticeable impact and support the results of the pre-posttest. The participants' improved results, WhatsApp messages, and responses to the interview questions all demonstrated how much they were delighted to engage in the training and utilize the ASR tool to improve their pronunciation. The interview highlighted how crucial it is to give students suitable materials and training to assist their improvements. Hence, teachers should consider implementing more technological tools to promote learning.

Conclusion

Analyzing data was carried out and reported through different phases; each phase revealed an answer to the research questions and tested the hypotheses. The questionnaire allowed me to get teachers 'positive insights on the importance of teaching pronunciation and the benefits of technology in helping their students promote their speaking skills and communicative competence. Then, the training phase allowed the experimental group participants to train their pronunciation and fluency through an online website. The scores and completion rate of the tasks and activities that are available on the website reveal that the participants enjoyed them and found them helpful. Further, the analysis and comparison of the pre-posttest scores showed significant pronunciation and fluency improvements among the experimental group participants who trained using an ASR online tool. This confirmed that the treatment was effective. Finally, the semistructured interview with the participants from the experimental group offered respondents feedback on the training. Their positive responses indicate that they appreciated the training and that the ASR tool utilized was valuable and practical. They, therefore, believe that it helped them improve their pronunciation and fluency and anticipate using more technological tools in the future.

General Conclusion

This mixed-method research investigated the effects of Automatic Speech Recognition (ASR) in improving EFL learners' pronunciation. It also sought teachers' perceptions of using technology to teach and assess pronunciation, and learners' attitudes toward pronunciation training using ASR technology. Speaking tests with the control and experimental groups. These tests assessed learners' pronunciation and speaking skills using the online website. They served as a measure to establish learners' oral proficiency levels before and after the intervention. Second, the experimental phase represents the core of this research as it directly tests the impact of the ASR technology on learners' oral proficiency. During this phase, the researcher introduced the experimental group to discrete training using the ASR website while withholding it from the control group. Last, a post-training interview was executed in the last phase which is the post-experimental phase. It attempted to check the experimental group participants' satisfaction with the training and their overall thoughts and opinions on the ASR technology.

Both qualitative and quantitative techniques were employed to gather and analyze data. The results regarding the analysis of the questionnaire revealed that teachers recognize the importance of teaching pronunciation as they believe it contributes to learners' intelligibility and communicative abilities both in the classroom and in real-life situations. They also acknowledged the significance of exposing learners to authentic input and corrective feedback to help them improve their pronunciation. Furthermore, teachers clarified that incorporating technology into pronunciation instruction has a positive impact on learners. However, integrating such technology requires training, effort, and planning before introducing it to the learners. They also emphasized

that it is crucial to provide adequate assistance to use technological tools and platforms and properly guide learners to become more autonomous and self-reliant.

The results showed a statistically significant difference between the outcomes of the experimental group and the control group. The initial hypothesis of this study suggested that using an ASR tool would help EFL learners improve their pronunciation and fluency. To test the hypothesis, several calculations and analyses were conducted to confirm or deny them. The pre-test and post-test scores were gathered, and then a series of statistical measures were carried out. Before conducting statistical tests, it was important to assess normality to ensure the validity of statistical analyses. Hence, the Shapiro-Walk test results confirmed that the data were drawn from a normally distributed population. Then, Levene's test was used to confirm the equality of variances across the groups. After testing for normality and homogeneity, independent samples t-tests were performed on SPSS Statistics.

The results showed that the mean scores of the experimental and control groups were statistically different, indicating that the experimental group improved their pronunciation compared to the control group, which did not show any improvement. Although the scores of the experimental group were slightly improved, the difference in the scores remained evident. According to the findings, there were statistically significant differences between the experimental and control groups' mean scores in favor of the experimental group. The experimental and control groups differ because each group was treated differently; the experimental group participated in ASR training, and the control group did not. The experimental group improved their pronunciation and fluency through training with the ASR online website. According to these results, the

Alternative Hypothesis H1 was statistically accepted, and the Null Hypothesis H0, implying that the treatment was ineffective, was rejected. Hence, ASR technologies are an effective tool for helping learners improve their pronunciation.

The responses collected from the interview conducted with the experimental group's participants revealed their positive perceptions of the technology. They found that the ASR website was easy to use. They said that the scores obtained from the website were clear and accurate. Additionally, they expressed that technology-enhanced learning could make lessons more enjoyable and effective. Learners also expressed their intentions to use the website even after the training, and they will always try to search for new advanced tools to improve other skills.

The study findings confirmed that advanced technologies allow learners to receive instant feedback on how they sound, enabling them to spot and correct their mispronunciations. They also provide a consistent and objective assessment of pronunciation which enables learners to monitor their development over time. On top of that, learners can receive ASR-based pronunciation practice at any time and from any location, giving them convenience and flexibility.

Overall, ASR technology has several benefits for teaching and assessing pronunciation, such as immediate feedback, individualization, accessibility, engagement, and progress tracking. When this technology is appropriately integrated into language learning, it can significantly help learners improve their pronunciation and their language competency.

4.4. Limitations

Identifying the study's limitations is crucial for understanding its scope and potential obstacles that may be faced while conducting research in this area.

This study was limited to first-year students at the Department of English and Literature at Biskra University. The study's sample of 12 participants per group may limit the generalizability of the findings. Hence, selecting a larger sample size in future studies would provide more significant statistical results.

The training period of four weeks may have been very short and insufficient and only measured the short-term effects of the experimental intervention on learners' pronunciation. In future research, conducting longitudinal studies may allow researchers to track improvements in participants' pronunciation over an extended period. This will enable the researcher to get insights into the long-term effects of the experimental intervention. The constraint of simply examining the short-term impacts shown in quasi-experimental research can be addressed with the help of longitudinal designs.

Voluntary sampling was used to recruit participants, only participants who volunteered participated in the study. Future research may randomly select the study's sample. This will allow the researcher to get better insights and evident results. Further, an ASR-based website was used in this study. Hence, future studies may search for other available ASR-based technologies and check their effectiveness in enhancing other skills.

Replicating the study using different samples, settings, and methodologies can help validate the findings and enhance their generalizability. Replication studies offer possibilities to support or contradict earlier findings.

4.5. Implications

Considering the implications of the study, researchers can offer a more thorough interpretation of their results and contribute to the field of language education and research.

The study's findings may have implications for language teaching practices, especially in the area of pronunciation instruction. Insights gained from the study may suggest more effective pronunciation teaching methods and materials.

The study highlighted the potential benefits of integrating technology into language learning. This could encourage teachers to explore technological tools and methods for teaching and assessing pronunciation. However, receiving training and adequate planning is essential before introducing the technology to the learners. Being familiar with the technology allows educators to provide immediate and sufficient help for their students.

This study suggests areas for future research such as investigating the long-term effects of pronunciation interventions, exploring the effectiveness of different instructional ASR-based technologies, or examining the impact of using other technological tools on other language skills.

The findings of this study showed that using an ASR tool to teach, assess, and improve pronunciation is effective. The learners were introduced to the ASR website in a training period

separated from their conventional classes and were only used in their free time. The results indicate that with this kind of training and such highly developed technology, learners can improve their English pronunciation. This study was the first that explored EFL learners' use of ASR technology in an Algerian context. Hence, pedagogical implications and recommendations for further research were provided to guide the future use of ASR tools in Algerian contexts where pronunciation is taught conventionally, and feedback is limited.

List of References

AbuSeileek, A. F. (2017). An integrative approach based on using annotations and computer-mediated feedback to support learners 'proficiency in the English language. *International Journal of Learning Technology*, *12*(2), 165. https://doi.org/10.1504/ijlt.2017.086383

Aliaga-García, C. (2007). Aliaga-García The role of phonetic training in L 2 speech learning: 1 The role of phonetic training in L 2 speech learning.

Arora, S., & Singh, R. (2012). Automatic Speech Recognition: A Review. International Journal Of Computer Applications (0975 – 8887), Volume 60(No.9).

Baker, A. (1982). *Tree Or Three? Student's Book: An Elementary Pronunciation Course*.

Ball, M. J., & Rahilly, J. (2014). *Phonetics: The Science of Speech*. Routledge.

Batliner, A., Schuller, B., Seppi, D., Steidl, S., Devillers, L., Vidrascu, L., Vogt, T., Aharonson, V., & Amir, N. (2011). The Automatic Recognition of Emotions in Speech. *Cognitive Technologies*, 71–99. https://doi.org/10.1007/978-3-642-15184-2_6

Beatty, K. (2013). *Teaching & Researching: Computer-Assisted Language Learning*. Routledge.

Berg, B. (2009). Qualitative Research Methods for the Social Sciences. Boston: Allyn & Bacon, pp.101-157.

Blake, B. J. (2008). *All About Language: A Guide*. OUP Oxford.

Bøhn, H., & Hansen, T. (2017). Assessing Pronunciation in an EFL Context: Teachers ' Orientations towards Nativeness and Intelligibility. *Language Assessment Quarterly*, *14*(1), 54–68. https://doi.org/10.1080/15434303.2016.1256407

Boyd, D., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Brinton, D. M., Celce-Murcia, M., & Goodwin, J. M. (1996). *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*.

Brown, A. (2014). *Pronunciation and Phonetics: A Practical Guide for English Language Teachers*. Routledge.

C., Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. Cambridge University Press.

Capraro, R. M., & Capraro, M. M. (2002). Myers-Briggs Type Indicator Score Reliability Across: Studies a Meta-Analytic Reliability Generalization Study. *Educational and Psychological Measurement*, *62*(4), 590–602. https://doi.org/10.1177/0013164402062004004

Carr, P. (2019). *English Phonetics and Phonology: An Introduction*. John Wiley & Sons.

Carrier, M. (2017). Automated Speech Recognition in language learning: Potential models, benefits, and impact.Training, Language and Culture, 1(1), 46-61. doi:10.29366/2017tlc.1.1.3

Carroll, J. M. (1962). The Prediction of Success in Intensive Foreign Language Training. *Training Research and Education University of Pittsburgh Press*.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching Pronunciation Hardback with Audio CDs (2): A Course Book and Reference Guide*. Cambridge University Press.

Chiu, T., Liou, H., & Yeh, Y. (2007). A Study of Web-based oral activities enhanced by Automatic Speech Recognition for EFL College Learning. *Computer Assisted Language Learning*, *20*(3), 209–233. https://doi.org/10.1080/09588220701489374

Chun, D. M. (2002). *Discourse Intonation in L2: From Theory and Research to Practice*.

John Benjamins Publishing.

Clark, J., Yallop, C., & Fletcher, J. (2007). *An Introduction to Phonetics and Phonology*. Wiley-Blackwell.

Clark, John, and Colin Yallop. 1995. "An introduction to phonetics and phonology." 2nd edition. Oxford: Blackwell.

Cohen, L., Manion, L., & Morrison, K. (2017). *Research Methods in Education*. Routledge.

Collins, B., & Mees, I. M. (2013). *Practical Phonetics and Phonology: A Resource Book for Students*. Routledge.

Creswell, J. W., & Clark, V. L. P. (2007). Designing and Conducting Mixed Methods Research. *Australian and New Zealand Journal of Public Health*, *31*(4), 388. https://doi.org/10.1111/j.1753-6405.2007.00096.x

Crotty, M. (1998). *The Foundations of Social Research: Meaning and Perspective in the Research Process*. SAGE.

Cruttenden, A. (1994). Phonetic and prosodic aspects of Baby Talk. *Cambridge University Press EBooks*, 135–152. https://doi.org/10.1017/cbo9780511620690.008

Crystal, D. (1985). A Dictionary of Linguistics and Phonetics (4th Ed.). Oxford: Blackwell.

Cucchiarini, C., Neri, A., & Strik, H. (2009). Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback. *Speech Communication*, *51*(10), 853–863. https://doi.org/10.1016/j.specom.2009.03.003

Dalton, D. F. (2002). Some Techniques for Teaching Pronunciation. *The Internet TESL Journal*.

Davis, K., Biddulph, R. & Balashek, S. (1952). "Automatic recognition of spoken digits".

Delahunty, G. P., & Garvey, J. J. (2010). *The English Language: From Sound to Sense*. Parlor Press LLC.

Denscombe, M. (2017). *EBOOK: The Good Research Guide: For Small-Scale Social Research Projects*. McGraw-Hill Education (UK).

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility : Evidence from four L1s. Studies in Second Language Acquisition, 19, 1 – 16.

Derwing, T. M., & Munro, M. J. (2014). Once you have been speaking a second language for years, it's too late to change your pronunciation. In L. Grant (Ed.), Pronunciation m yths (pp. 3 4 – 5 5). Champaign: The University of Michigan Press.

Derwing, T. M., & Munro, M. J. (2015). Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research. Philadelphia, PA: John Benjamins.

Derwing, T. M., Munro, M. H. G., & Thomson, R. I. (2007). A Longitudinal Study of ESL Learners 'Fluency and Comprehensibility Development. *Applied Linguistics*, *29*(3), 359–380. https://doi.org/10.1093/applin/amm041

Derwing, T. M., Munro, M. H. G., Thomson, R. H., & Rossiter, M. J. (2009). THE RELATIONSHIP BETWEEN L1 FLUENCY AND L2 FLUENCY DEVELOPMENT. *Studies in Second Language Acquisition*, *31*(04), 533. https://doi.org/10.1017/s0272263109990015

Derwing, T. M., Rossiter, M. J., Munro, M. H. G., & Thomson, R. I. (2004). Second Language Fluency: Judgments on Different Tasks. *Language Learning*, *54*(4), 655–679. https://doi.org/10.1111/j.1467-9922.2004.00282.x

Dobrovolsky, M., & Katamba, F.X. (2011). 16 Phonetics: The Sounds of Language.

Dudley, H., & Tarnoczy, T. H. (1950). The speaking machine of Wolfgang von

Kempelen. Journal of the Acoustical Society of America, 22, 151-166. doi:10.1121/1.1906583.

Elimat, A. K., & AbuSeileek, A. F. (2014a). Automatic speech recognition technology is an effective means for teaching pronunciation. *The JALT CALL Journal*, *10*(1), 21–47. https://doi.org/10.29140/jaltcall.v10n1.166

Elimat, A. K., & AbuSeileek, A. F. (2014b). Automatic speech recognition technology is an effective means for teaching pronunciation. *The JALT CALL Journal*, *10*(1), 21–47. https://doi.org/10.29140/jaltcall.v10n1.166

Elting, S., & Firkins, A. (2006a). Dramatizing poetry in the second language classroom. *English Teaching-Practice and Critique*, *5*(3), 127–136. http://files.eric.ed.gov/fulltext/EJ843834.pdf

Elting, S., & Firkins, A. (2006b). Dramatizing poetry in the second language classroom. *English Teaching-Practice and Critique*, *5*(3), 127–136. http://files.eric.ed.gov/fulltext/EJ843834.pdf

Elting, S., & Firkins, A. (2006c). Dramatizing poetry in the second language classroom. *English Teaching-Practice and Critique*, *5*(3), 127–136. http://files.eric.ed.gov/fulltext/EJ843834.pdf

Errattahi, R., El Hannani, A., & Ouahmane, H. (2015). Automatic Speech RecognitionErrors Detection and Correction: A Review. *International Conference on Natural Language and Speech Processing*.

Errattahi, R., Hain, T., & Ouahmane, H. (2019). System-independent ASR error detection and classification using Recurrent Neural Network. *Computer Speech & Language*, *55*, 187–199. https://doi.org/10.1016/j.csl.2018.12.007

Eskenazi, M. (1999). Using a computer in foreign language pronunciation training: What advantages? CALICO Journal, 16(3), 447–69.

Flanagan, J. L. (1972). *Speech Analysis: Synthesis and Perception, 2., (expanded) Ed*.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65. https://doi.org/10.1016/s0095-4470(19)30537-6

Fletcher, J. E. (1989). Input-output analysis and tourism impact studies. *Annals of Tourism Research*, *16*(4), 514–529. https://doi.org/10.1016/0160-7383(89)90006-6

Fraser, H. (2000). *Coordinating Improvements in Pronunciation Teaching for Adult Learners of English as a Second Language*.

Gardner, R. C., & Lambert, W. E. (1972). Attitudes and Motivation in Second Language Learning. Rowley, MA: Newbury House Publishers.

Gerosa, M., & Giuliani, D. (2004). Investigating automatic recognition of non-native children's speech. *Conference of the International Speech Communication Association*. https://doi.org/10.21437/interspeech.2004-572

Gibran, K. (1991). *The Prophet*. Christian Large Print.

Gilakjani, A. P. (2012). A Study of Factors Affecting EFL Learners 'English Pronunciation Learning and the Strategies for Instruction. *International Journal of Humanities and Social Science*.

Gilakjani, A. P., & Rahimy, R. (2019). Factors influencing Iranian teachers 'use of computer-assisted pronunciation teaching (CAPT). *Education and Information Technologies*. https://doi.org/10.1007/s10639-018-09851-6

Gilbert, J. B. (2008). *Teaching Pronunciation: Using the Prosody Pyramid*.

Gilbert, J. B. (2012). *Clear Speech Teacher's Resource and Assessment Book: Pronunciation and Listening Comprehension in North American English*. Cambridge University Press.

Gold, B., Morgan, N., & Ellis, D. (2011). *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*. John Wiley & Sons.

Google Help. (n.d.). https://support.google.com/

Google Voice. (n.d.). https://voice.google.com/

Gulzar, T., Singh, A., Rajoriya, D., & Farooq, N. (2014). A Systematic Analysis of Automatic Speech Recognition: An Overview. *International Journal of Current Engineering and Technology*.

Guskaroska, A. (2019). *ASR as a tool for providing feedback for vowel pronunciation practice*. Iowa State University.

Gut, U. (2009). *Introduction to English Phonetics and Phonology*. Peter Lang.

Harmer, J. (2010). *The Practice of English Language Teaching*.

Hewings, M. (2004). *Pronunciation Practice Activities: A Resource Book for Teaching English Pronunciation*. Ernst Klett Sprachen.

Hickey, R. (2014). *A Dictionary of Varieties of English*. John Wiley & Sons.

Hincks, R. (2003). *Speech technologies for pronunciation feedback and evaluation*. *ReCALL*, *15*(1), 3–20. https://doi.org/10.1017/s0958344003000211

Hincks, R. (2015). Technology and learning pronunciation. In M. Reed & J. Levis (Eds), The Handbook of English pronunciation (pp. 505–519). Malden, MA: John Wiley & Sons.

Hughes, A., Trudgill, P., & Watt, D. (2013). *English Accents and Dialects: An Introduction to Social and Regional Varieties of English in the British Isles, Fifth Edition*. Routledge.

Isaacs, T., & Thomson, R. I. (2013). Rater Experience, Rating Scale Length, and Judgments of L2 Pronunciation: Revisiting Research Conventions. *Language Assessment Quarterly*, *10*(2), 135–159. https://doi.org/10.1080/15434303.2013.769545

Isaacs, T., & Trofimovich, P. (2017). Second Language Pronunciation Assessment: Interdisciplinary Perspectives. *Multilingual Matters EBooks*, 3–11.

J.W.Forgie and C.D.Forgie, 1959, Results obtained from a vowel recognition computer program, J.Acoust.Soc.Am., 31(11), pp.1480-1489.

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

Juang, B., & Furui, S. (2000). Automatic recognition and understanding of spoken language - a first step toward natural human-machine communication. *Proceedings of the IEEE*, *88*(8), 1142–1165. https://doi.org/10.1109/5.880077

Kaiser, D. J. (2017). iPronounce: Understanding pronunciation apps. Online Webinar, 8 June 2017. Recording and slides available at http://ollren.org/events/past-events#s-lg-boxwrapper17792711

Kang, O., & Ginther, A. (2017). *Assessment in Second Language Pronunciation*. Taylor & Francis.

Kang, O., Thomson, R. I., & Murphy, J. M. (2017). *The Routledge Handbook of Contemporary English Pronunciation*. Routledge.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68.

https://doi.org/10.1016/j.bushor.2009.09.003

Karpagavalli, S., & Chandra, E. (2016). A Review on Automatic Speech Recognition Architecture and Approaches. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, *9*(4), 393–404. https://doi.org/10.14257/ijsip.2016.9.4.34

Katyal, A., Kaur, A., & Gill, J. (2014). Automatic Speech Recognition: A Review. International Journal Of Engineering And Advanced Technology (IJEAT), Volume-3(Issue-3).

Kelly, G. (2000). *How to Teach Pronunciation*.

Kelly, G. (2001). *How To Teach Pronunciation*. Pearson Education India.

Kelly, L. G. (1969). *25 Centuries of Language Teaching: An Inquiry Into the Science, Art, and Development of Language Teaching Methodology, 500 B.C.-1969*. Rowley, Mass : Newbury House Publishers.

Kenworthy, J. (1990). *Teaching English Pronunciation*.

Koponen, M. & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), Perspectives on Fluency (pp. 5–24). Ann Arbor, MI: University of Michigan Press.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. In *System* (Vol. 32, Issue 2, pp. 145–164). Elsevier BV. https://doi.org/10.1016/j.system.2004.01.001

Kreidler, C. W. (2004). *The Pronunciation of English: A Course Book*. Wiley-Blackwell.

Laroy, C. (2013). *Pronunciation - Resource Books for Teachers*. Oxford University Press.

Lenneberg, E. H. (1967). The Biological Foundations of Language. *Hospital Practice*, *2*(12), 59–67. https://doi.org/10.1080/21548331.1967.11707799

Levis, J. (2007). Computer technology in teaching and researching pronunciation. Annual

Review of Applied Linguistics, 27, 184-202.

Levis, J., & Suvorov, R. (2012). Automatic Speech Recognition. The Encyclopedia Of Applied Linguistics. doi: 10.1002/9781405198431.wbeal0066

Li, M., Han, M., Han, Z., Mo, Y., Chen, X., & Liu, X. (2017). Improving English Pronunciation via Automatic Speech Recognition Technology. International Symposium On Educational Technology.

MacGregor, L., Corley, M., & Donaldson, D. (2009). Not all disfluencies are equal: The effects of disfluent repetitions on language comprehension. *Brain and Language*, *111*(1), 36–45. https://doi.org/10.1016/j.bandl.2009.07.003

Maniruzzaman, M. (2008). *Teaching EFL Pronunciation: Why, What, and How?* GRIN Verlag.

McCrocklin, S. (2014). The potential of Automatic Speech Recognition for fostering pronunciation learners' autonomy (Doctor dissertation). Iowa State University.

McCrocklin, S. (2015). Automatic speech recognition: Making it work for your pronunciation class. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference (ISSN 2380-9566), Santa Barbara, CA (pp. 126-133). Ames, IA: Iowa State University.

McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of Automatic Speech Recognition. *System*, *57*, 25–42. https://doi.org/10.1016/j.system.2015.12.013

McCrocklin, S. (2018). Learners 'Feedback Regarding ASR-based Dictation Practice for Pronunciation Learning. *The CALICO Journal*, *36*(2), 119–137. https://doi.org/10.1558/cj.34738

Mohsin, N. (2012). Improving pronunciation of secondary school teachers through computer-

assisted learning: An Experience. Proceedings of the European Conference LINQ, Brussels, 134–142.

Morley, J. (1991). The Pronunciation Component in Teaching English to Speakers of Other Languages. *TESOL Quarterly*, *25*(3), 481. https://doi.org/10.2307/3586981

Muijs, D. (2004). Doing quantitative research in education with SPSS. London: Sage.

Munro, M. H. G., & Derwing, T. M. (1995). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning*, *45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Neri, A., Cucchiarini, C., Strik, H. (2002) Feedback in computer-assisted pronunciation training: technology push or demand pull? Proc. 7th International Conference on Spoken Language Processing (ICSLP 2002), 1209-1212, doi: 10.21437/ICSLP.2002-246

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The Pedagogy-Technology Interface in Computer Assisted Pronunciation Training. *Computer Assisted Language Learning*, *15*(5), 441–467. https://doi.org/10.1076/call.15.5.441.13473

O'Leary, Z. (2017). *The Essential Guide to Doing Your Research Project*. SAGE.

Ogden, R. (2017). *Introduction to English Phonetics*. Edinburgh University Press.

Olejnik, S., & Algina, J. (2000). Measures of Effect Size for Comparative Studies: Applications, Interpretations, and Limitations. *Contemporary Educational Psychology*, *25*(3), 241–286. https://doi.org/10.1006/ceps.2000.1040

Patkowski, M. S. (1990). Age and Accent in a Second Language: A Reply to James Emil Flege. *Applied Linguistics*, *11*(1), 73–89. https://doi.org/10.1093/applin/11.1.73

Pennington, M. C., & Rogerson-Revell, P. (2018). *English Pronunciation Teaching and Research: Contemporary Perspectives*. Springer.

Pica, T. (1984). L1 Transfer and L2 Complexity as Factors in Syllabus Design. *TESOL Quarterly*, *18*(4), 689. https://doi.org/10.2307/3586583

Powney, J., & Watts, M. (2018). *Interviewing in Educational Research*. Routledge.

Probst, K., Ke, Y., & Eskenazi, M. (2002). Enhancing foreign language tutors – In search of the golden speaker. *Speech Communication*. https://doi.org/10.1016/s0167-6393(01)00009-7

Purushotma R. (2005). Commentary: You're not studying, you're just…. Language Learning & Technology, 9(1), 80–96.

Reed, M., & Levis, J. (2015). *The Handbook of English Pronunciation*. John Wiley & Sons.

Riyahifar, Z. (2016). Unmasking Factors Affecting Learning Pronunciation: Critical Period or Sensitive Period? Retrieved from http://www.psp-ltd.com/JIEB_47_4_2016.pdf

Roach, P. (2000). English Phonetics and Phonology. Cambridge: Cambridge University Press.

Roach, P. (2002). A Little Encyclopedia of Phonetics (pp. 93-98). UK: University of Reading.

Roach, P. J. (2000). *English Phonetics and Phonology: A Self-Contained, Comprehensive Pronunciation Course*.

Rogers, H. (2000). *The Sounds of Language: An Introduction to Phonetics*. Routledge.

Rosyid, A. (2016). FACTORS AFFECTING EFL LEARNERS IN LEARNING ENGLISH PRONUNCIATION. *Jurnal Ilmiah Pendidikan*.

Samudravijaya, K. Automatic Speech Recognition. Retrieved 1 September 2020, from http://www.iitg.ac.in/samudravijaya/tutorials/asrTutorial.pdf

Schmitt, N. (2010). *An Introduction to Applied Linguistics*. Routledge.

Segalowitz, N. (2010). Cognitive Bases of Second Language Fluency. In *Routledge eBooks*. Informa. https://doi.org/10.4324/9780203851357

Sidgi, L. F. S., & Shaari, A. J. (2017). The Usefulness of Automatic Speech Recognition (ASR) Eyespeak Software in Improving Iraqi EFL Students 'Pronunciation. *Advances in Language and Literary Studies*, *8*(1), 221. https://doi.org/10.7575/aiac.alls.v.8n.1p.221

Speechace. (2022, June 21). *Home – Speechace | Pronunciation and fluency assessment via speech recognition*. Speechace | Pronunciation and Fluency Assessment via Speech Recognition. https://www.speechace.com/

Spezzini, S. (2010). Effects of Visual Analogies on Learner Outcomes: Bridging from the Known to the Unknown. *International Journal for the Scholarship of Teaching and Learning*, *4*(2). https://doi.org/10.20429/ijsotl.2010.040211

Spezzini, S., Barratt, L., & Carter, D. (2018). Pronunciation Assessment. The TESOL Encyclopedia Of English Language Teaching, 1-7. doi: 10.1002/9781118784235.eelt0694

Svendsen, T. (2003). Speech Technology: Past, Present and Future. Telektronikk, Volume 99(No. 2), 6-17.

Swamy, S., & KV, R. (2013). An Efficient Speech Recognition System. *Computer Science and Engineering : An International Journal*, *3*(4), 21–27. https://doi.org/10.5121/cseij.2013.3403

Szpyra-Kozłowska, J. (2014). *Pronunciation in EFL Instruction: A Research-Based Approach*. Multilingual Matters.

Takenouchi, T. (2022). The Effects of Pronunciation Instruction Using Speech Recognition Software for Adult Learners of English. *The Journal of the Practical English Phonetics*.

Tatawy, M. E. (2002). Corrective Feedback in Second Language Acquisition. *DOAJ (DOAJ: Directory of Open Access Journals)*. https://doi.org/10.7916/salt.v2i2.1645

The Journal of the Acoustic Society of America, 24(6), 637-642.

Thornbury, S. (2006). *An A-Z of ELT: A Dictionary of Terms and Concepts Used in English Language Teaching*. MACMILLAN.

TOD, A. (2006). Interviewing. In: GERRISH, K. and LACEY, A., (eds.) Research process in nursing. Wiley-Blackwell, 337-352.

Tottie, G. (2002). *An Introduction to American English*. Blackwell Publishing.

Towell, R., Hawkins, R. E., & Bazergui, N. (1996). The Development of Fluency in Advanced Learners of French. *Applied Linguistics*, *17*(1), 84–119. https://doi.org/10.1093/applin/17.1.84

Trudgill, P., & Hannah, J. (2017). *International English: A Guide to Varieties of English Around the World*. Taylor & Francis.

Tsagris, M., & Pandis, N. (2021). Normality test: Is it really necessary? *American Journal of Orthodontics and Dentofacial Orthopedics*, *159*(4), 548–549. https://doi.org/10.1016/j.ajodo.2021.01.003

Underhill, A. (2005). *Sound Foundations: Learning and Teaching Pronunciation*.

Ur, P. (2012). *A Course in English Language Teaching*. Cambridge University Press.

Volín, J., & Skarnitzl, R. (2010). The strength of foreign accent in Czech English under adverse listening conditions. *Speech Communication*. https://doi.org/10.1016/j.specom.2010.06.009

Wachowicz, K., & Scott, B. (1999). Software that listens: It's not a question of whether, it's a question of how. CALICO Journal, 16(3), 253–76.

Wallace, L. (2016). Using Google web speech as a springboard for identifying personal pronunciation problems. In J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds). Proceedings of

the 7th Pronunciation in Second Language Learning and Teaching Conference, ISSN 2380-9566, Dallas, TX, October 2015 (pp. 180-186). Ames, IA: Iowa State University.

Wells, J. C., & Wells, J. C. (1982). *Accents of English: Volume 1*. Cambridge University Press.

Witt, S. M., & Young, S. (2000). Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Communication*, *30*(2–3), 95–108. https://doi.org/10.1016/s0167-6393(99)00044-8

Wong, R. (1987). *Teaching Pronunciation: Focus on English Rhythm and Intonation*.

Wright, L. W., Algeo, J., & Kyt, M. (2000). *The Development of Standard English, 1300-1800: Theories, Descriptions, Conflicts*. Cambridge University Press.

Yates, L., & Zielinski, B. (2009). *Give it a Go: Teaching Pronunciation to Adults*.

Yu, D., & Deng, L. (2014). *Automatic Speech Recognition: A Deep Learning Approach*. Springer.

Yule, G. (2014). *The Study of Language*. Cambridge University Press.

Zhang, Y. (2009). Reading to Speak: Integrating Oral Communication Skills. *English Teaching Forum*, *47*(1), 32–34. http://files.eric.ed.gov/fulltext/EJ923446.pdf

Zimmer, M. (2018). Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity. *Social Media and Society*, *4*(2), 205630511876830. https://doi.org/10.1177/2056305118768300

APPENDICES

Appendix A: Teachers 'Questionnaire

Dear teachers,

This questionnaire is part of our Ph.D. research regarding the effectiveness of using ASR-based technology to teach and improve the pronunciation of second-year students at Mohamed Khider Biskra University. Hence, we invite oral expression and phonetics teachers to participate in this study. The primary purpose of this questionnaire is to obtain your opinions and attitudes towards pronunciation teaching and how it is dealt with and taught in our department. Your contribution will undoubtedly be of great help to us.

Your answers will be strictly confidential and will be used for academic research purposes only.

Section 1. Personal Information

1. How long have you been teaching English?

   a. Less than 5 years   b. 5 to 10 years   c. More than 10 years

2. Have you already taught second-year level?

   a. Yes   b. No

Section 2. Teaching Oral Expression/Phonetics

1. As a teacher of phonetics, what is the overall goal for teaching Phonetics to first-year students:

   1. Identify and produce target sounds

   2. Master prosodic features Rhythm, stress, pitch, and intonation

   3. The ability to communicate

   4. Other (please specify)

2. As a teacher of oral expression, what is the overall goal of teaching oral expression to first-year students:

a. To allow students to speak the target language correctly

b. To enable students to communicate inside the classroom

c. To allow students to use the target language in real-life situations

d. Other (please specify)

3. When planning your lesson, do you decide the content of teaching on your own, or

Do you have other considerations?

4. If your students struggle to speak, do you provide immediate feedback and notice their

errors?

Section 3. Teaching Pronunciation

1. How important is pronunciation teaching?

Not at all important   Slightly important   Important   Fairly important   Very important

2. Do you assess your students 'pronunciation?

a. Yes   b. No

- If yes, what exactly do you assess and how?

3. In your opinion, why is pronunciation considered the "Cinderella area" of language

teaching?

4. Do you think providing authentic input and instant feedback is crucial to improving students 'pronunciation?

a. Yes   b. No

Please specify why? …………………………………………………………

5. As far as first-year students are concerned, how would you evaluate their level of pronunciation?

a. Poor   b. Fair   c. Good   d. Excellent

Section 4. Technology and Pronunciation

1. Do you think that using technology to teach pronunciation can be effective?

a. Yes   b. No

2. Have you ever used a technological tool to teach pronunciation?

a. Yes   b. No

- If yes, what tools did you use (online/offline) (software, application)?

3. Before selecting and incorporating new technology, what options do you want this technology to offer your student to improve their pronunciation?

a. Authentic input   b. Corrective feedback   Other

4. What were your students 'attitudes toward using technology inside the classroom?

5. Have you heard of ASR-based technologies before?

a. Yes   b. No

- If Yes, explain .

Section 5. Teachers 'Opinions on Automatic Speech Recognition Technology

My research study includes pronunciation training using an ASR website to improve students 'pronunciation. Do you think teachers should incorporate technology inside and outside the classroom/design additional training online?

Thank you for your time.

Appendix B: Participants 'Information Card

Instruction : Please fill in this information card if you are willing to participate in the study. The collected data will be treated confidentially and will be used for educational purposes only. (Read the consent form first and then fill in your information.)

Full name: ……………………………..

Gender:

Age: ……… years old

How old were you when you started studying English? ………… years old

What was your baccalaureate exam score? ………….

What was your mark in English? ………..

Was English your first choice as a university major?   ☐ Yes   ☐ No

Why?
………………………………………………………………………………………………

What is your favorite module? ……………….

What is the aspect of language that you want to improve?

☐ Speaking   ☐ Listening   ☐ Reading   ☐ Writing

Have you ever had any specific English language training outside of school?

……………………………………………………………………………………………

How would describe your technical skills (technology skills)?

☐ Poor   ☐ Average   ☐ Good   ☐ Very Good   Excellent

Phone number: ………………

E-mail: …………………………….

Group: ……..

Thank you.

Appendix C: Consent Form

Project Title: The effectiveness of Using Automatic Speech Recognition Technologies on Students 'Pronunciation - A case study of first-year EFL learners at Biskra University.

The objective of the study: Our study investigates the effectiveness of using the ASR technology website to teach and assess learners 'pronunciation and help them improve their speaking skills.

Introduction

(This consent may contain words that you do not understand. You will find contact details below to ask the researcher to explain any words or information that you do not clearly understand.)

You are being asked to participate in a research study. This research is going to be conducted during the second semester of the scholastic year 2021 - 2022. When you are invited to participate in research, you have the right to be informed about the study procedures so that you can decide whether you want to consent to participation. You have the right to know what you will be asked to do so that you can decide whether or not to be in the study. Your participation is <u>voluntary</u>. You do not have to be in the study if you do not want to. You will find below all the information you need concerning this research. Please read carefully and then tick the boxes below.

1. **Your role:** Your role as a participant in this project is to take part in pronunciation training separated from your main courses at the university. You are going to be requested to access an API using your email and personal information and then participate in a set of activities that aim at improving your pronunciation. Additionally, you will be participating in a set of tests and interviews to collect your thoughts on the website and assess your progress.

2. **Confidentiality:** Your personal information will be collected in this study and will be encoded and you will not be identified. Your specific information will be

securely collected and kept in a separate, secure location on the researcher's PC or mobile phone. In addition, this study includes recordings (audiotapes and voice files) that will be taken throughout different stages of the study. Hence, you will give your permission for their use as a part of the Ph.D. thesis.

I voluntarily agree to participate in this research study. ☐

I have had the purpose and nature of the study explained to me in writing and I have had the opportunity to ask questions about the study. ☐

I understand that participation involves tests, recordings, and interviews ☐

I understand that I will not benefit directly from participating in this research. ☐

I agree to my voice being audio-recorded. ☐

I understand that all information I provide for this study will be treated confidentially. ☐

I understand that in any report on the results of this research, my identity will remain anonymous. This will be done by changing my name and disguising any details of my interview that may reveal my identity or the identity of people I speak about. ☐

I understand that signed consent forms and original audio recordings will be retained on the researcher's PC or mobile phone. ☐

I understand that I am free to contact any of the people involved in the research to seek further clarification and information. ☐

The researcher: Souici Roumaissa – PhD student at Mohamed Khider University.

Contact: souiciroumaissa@outlook.com

Supervisor: Pr. Chelli Saliha

Signature of research participant: ………………………..   Date: ………………

Signature of the researcher: ………………………..   Date: ………………
Thank you for taking the time to read this consent form.

Thank you for your participation.

Yours sincerely,

Souici Roumaissa

Appendix D: Speaking Test Answers

You are going to participate in a speaking test. Hence, your voice will be recorded and saved on an online platform (speechace.com) and then your speaking proficiency and pronunciation will be scored. The data collected will only be used for research purposes. Thank you for taking part in the testing.

Instructions:

- Read the answers out loud and do your best to read with a clear voice.

- You have a preparation time of 30 seconds so you can read the answers once before starting to record. If you want a longer period, just ask.

- Once you start the recording and you do not feel satisfied with your outcome, you have the chance to ask for repetition if you have to. (You can re-take the test once only).

__Question:__ What are the advantages of the internet?

__Answer__:

The internet is one of the most powerful creations in the world. It offers people endless knowledge and entertainment. Today, the Internet plays a critical role in many areas of human life. It has multiple advantages such as: Abundant information and resources that allow people to get information about any topic and offer an answer to any type of question, and provides people with endless access to watch movies, and videos, play games online, and listen to music. Further, it enables people to communicate with others from around the world and stay in constant communication with one another. The Internet is the most powerful weapon in the 21st century.

Appendix E: Control Group Scores

Participant	Pre-Test Scores	Post-Test Scores
Alice	Speechace score **6.7/9** — Pronunciation 7.6/9, Fluency 5.5/9, Vocabulary 7.3/9, Grammar 6.7/9	Speechace score **6.9/9** — Pronunciation 7.5/9, Fluency 6.0/9, Vocabulary 7.2/9, Grammar 6.5/9
Amelia	Speechace score **7.2/9** — Pronunciation 7.6/9, Fluency 6.5/9, Vocabulary 7.5/9, Grammar 7.2/9	Speechace score **7/9** — Pronunciation 7.5/9, Fluency 6.8/9, Vocabulary 7.5/9, Grammar 6.8/9
Ava	Speechace score **7.1/9** — Pronunciation 7.4/9, Fluency 6.3/9, Vocabulary 7.4/9, Grammar 7.5/9	Speechace score **7/9** — Pronunciation 7/9, Fluency 5.6/9, Vocabulary 7.7/9, Grammar 7.7/9
Grace	Speechace score **0/9** — Pronunciation 7.6/9, Fluency 7/9, Vocabulary 7.1/9, Grammar 6.7/9	Speechace score **7.3/9** — Pronunciation 7.7/9, Fluency 7/9, Vocabulary 7.9/9, Grammar 6.5/9
Isabella	Speechace score **4.6/9** — Pronunciation 7.8/9, Fluency 6.5/9, Vocabulary 6.9/9, Grammar 6.5/9	Speechace score **0/9** — Pronunciation 7.9/9, Fluency 6.6/9, Vocabulary 7/9, Grammar 6.5/9
Liam	Speechace score **7.5/9** — Pronunciation 8/9, Fluency 7.8/9, Vocabulary 7.5/9, Grammar 6.9/9	Speechace score **7.3/9** — Pronunciation 8.1/9, Fluency 7.4/9, Vocabulary 7.3/9, Grammar 6.5/9

Participant	Pre-Test Scores	Post-Test Scores
Lucas	Speechace score 6.9/9 — Pronunciation 7/9, Fluency 6.4/9, Vocabulary 7.2/9, Grammar 6.5/9	Speechace score 6.9/9 — Pronunciation 7/9, Fluency 6.5/9, Vocabulary 6.6/9, Grammar 7.2/9
Madison	Speechace score 7.2/9 — Pronunciation 7.7/9, Fluency 6.4/9, Vocabulary 7.3/9, Grammar 7.3/9	Speechace score 7.2/9 — Pronunciation 7.6/9, Fluency 6.3/9, Vocabulary 7.1/9, Grammar 7.6/9
Mia	Speechace score 4.6/9 — Pronunciation 7.3/9, Fluency 6.3/9, Vocabulary 7.0/9, Grammar 7.6/9	Speechace score 6.8/9 — Pronunciation 6.5/9, Fluency 6.6/9, Vocabulary 6.6/9, Grammar 7.6/9
Noah	Speechace score 6.9/9 — Pronunciation 7.2/9, Fluency 6.4/9, Vocabulary 7.1/9, Grammar 7/9	Speechace score 6.8/9 — Pronunciation 7.2/9, Fluency 6.5/9, Vocabulary 7.1/9, Grammar 6.5/9
Olivia	Speechace score 7/9 — Pronunciation 7.3/9, Fluency 6.4/9, Vocabulary 7.1/9, Grammar 7.1/9	Speechace score 7/9 — Pronunciation 7.1/9, Fluency 6.5/9, Vocabulary 7/9, Grammar 7.3/9
Sophia	Speechace score 7.1/9 — Pronunciation 7.7/9, Fluency 6.6/9, Vocabulary 7.1/9, Grammar 7.1/9	Speechace score 6.9/9 — Pronunciation 8.1/9, Fluency 6.3/9, Vocabulary 7.3/9, Grammar 5.3/9

Appendix F: Experimental Group Scores

Participant	Pre-Test Scores	Post-Test Scores
Alexander	Speechace score **0.0/9** — Pronunciation 7.8/9, Fluency 6.0/9, Vocabulary 7.1/9, Grammar 6.2/9	Speechace score **6.7/9** — Pronunciation 7.7/9, Fluency 6.5/9, Vocabulary 6.2/9, Grammar 6.3/9
Aria	Speechace score **4.9/9** — Pronunciation 7.7/9, Fluency 7.2/9, Vocabulary 7.5/9, Grammar 6.8/9	Speechace score **7.3/9** — Pronunciation 7.9/9, Fluency 7.5/9, Vocabulary 7.5/9, Grammar 6.3/9
Camilla	Speechace score **4.6/9** — Pronunciation 8/9, Fluency 8/9, Vocabulary 7.3/9, Grammar 6.6/9	Speechace score **7.5/9** — Pronunciation 8.4/9, Fluency 7.6/9, Vocabulary 7.6/9, Grammar 6.5/9
Ella	Speechace score **6.9/9** — Pronunciation 7.5/9, Fluency 6.2/9, Vocabulary 7.2/9, Grammar 6.6/9	Speechace score **7.6/9** — Pronunciation 8.2/9, Fluency 7.6/9, Vocabulary 7.4/9, Grammar 7.3/9
Ezra	Speechace score **7.4/9** — Pronunciation 8.4/9, Fluency 7.5/9, Vocabulary 7.2/9, Grammar 6.6/9	Speechace score **7.5/9** — Pronunciation 8.3/9, Fluency 7.5/9, Vocabulary 7.8/9, Grammar 6.4/9
Levi	Speechace score **6.8/9** — Pronunciation 7.6/9, Fluency 5.8/9, Vocabulary 7.1/9, Grammar 6.6/9	Speechace score **7.6/9** — Pronunciation 8.5/9, Fluency 8.0/9, Vocabulary 7.4/9, Grammar 6.5/9

Participant	Pre-Test Scores		Post-Test Scores	
Lorrie	Speechace score **6.8/9**	Pronunciation 7.7/9 Fluency 5.5/9 Vocabulary 7.6/9 Grammar 6.5/9	Speechace score **7/9**	Pronunciation 8.1/9 Fluency 6/9 Vocabulary 7.4/9 Grammar 6.6/9
Luna	Speechace score **4.9/9** Scoring guide	Pronunciation 8.3/9 Fluency 7.5/9 Vocabulary 7.3/9 Grammar 6.4/9	Speechace score **7.8/9**	Pronunciation 8.5/9 Fluency 8.4/9 Vocabulary 7.5/9 Grammar 5.9/9
Mila	Speechace score **7.4/9** Scoring guide	Pronunciation 7.8/9 Fluency 7.3/9 Vocabulary 7.3/9 Grammar 7.1/9	Speechace score **7.8/9**	Pronunciation 8.3/9 Fluency 8.4/9 Vocabulary 7.3/9 Grammar 7.3/9
Owen	Speechace score **7.1/9** Scoring guide	Pronunciation 7.7/9 Fluency 7.0/9 Vocabulary 7.2/9 Grammar 6.4/9	Speechace score **7.3/9**	Pronunciation 8.0/9 Fluency 7.0/9 Vocabulary 7.4/9 Grammar 6.6/9
Penelope	Speechace score **6.9/9**	Pronunciation 7.8/9 Fluency 6.0/9 Vocabulary 7.4/9 Grammar 6.5/9	Speechace score **0/9**	Pronunciation 8.2/9 Fluency 7.5/9 Vocabulary 6.7/9 Grammar 6/9
Scarlett	Speechace score **7.5/9** Scoring guide	Pronunciation 8.4/9 Fluency 7.7/9 Vocabulary 7.5/9 Grammar 6.5/9	Speechace score **7.6/9**	Pronunciation 8.5/9 Fluency 8/9 Vocabulary 7.6/9 Grammar 6.5/9

Appendix G: A Sample of Participants 'Recording

Fluency | Pronunciation   137 Words per minute ⓘ   ⏸ 4 Bad pauses ⓘ

▶ Internet is one of the most powerful Creations in the world. It offers people and less knowledge and entertainment 🔶🔶 today. The internet plays a critical role in many areas of human life, which has multiple advantages 🔶 such as abundant of information and resources. That allows people to get information about any topic 🔶 and offers an answer to any type of question. Provides people with endless access to watch movies, videos, play games online, listen to music 🔶🔶🔶 further. It enables people to communicate with others from around the world, and stay in constant communication with, with one. Another internet is the most powerful weapon in the twenty, the twenty-first century.

8.7

Fluency | Pronunciation   86% Accuracy ⓘ

▶ Internet is one of the most powerful nation in the world. It's offers four people, and less knowledge and entertainment today. The internet plays a critical role in many areas of human life. It has multiple advantages such as abundance of information and resources, that allows people to get information about any topic and offers an answer to any type of question. Provides people with endless set to watch movies videos, play games online, listen to music for there. In fact, any both people to communicate with others from around the world and stay in constant communication, with one. Another internet is the most powerful weapon in the 21-century.
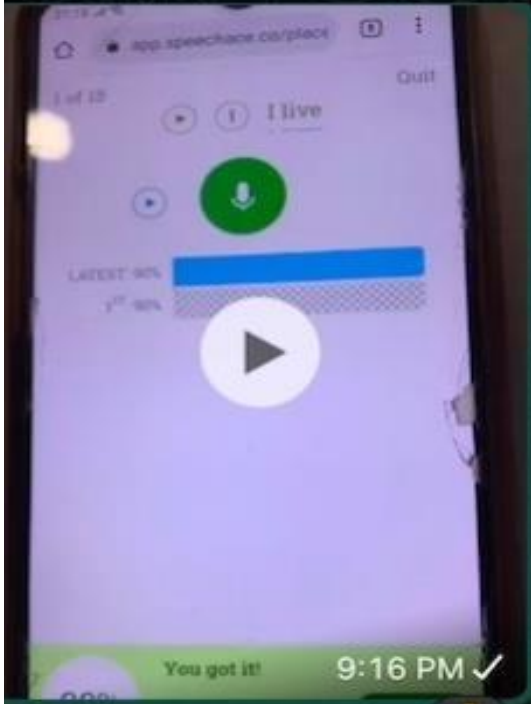
7.7

Appendix H: Screenshots of WhatsApp Messages
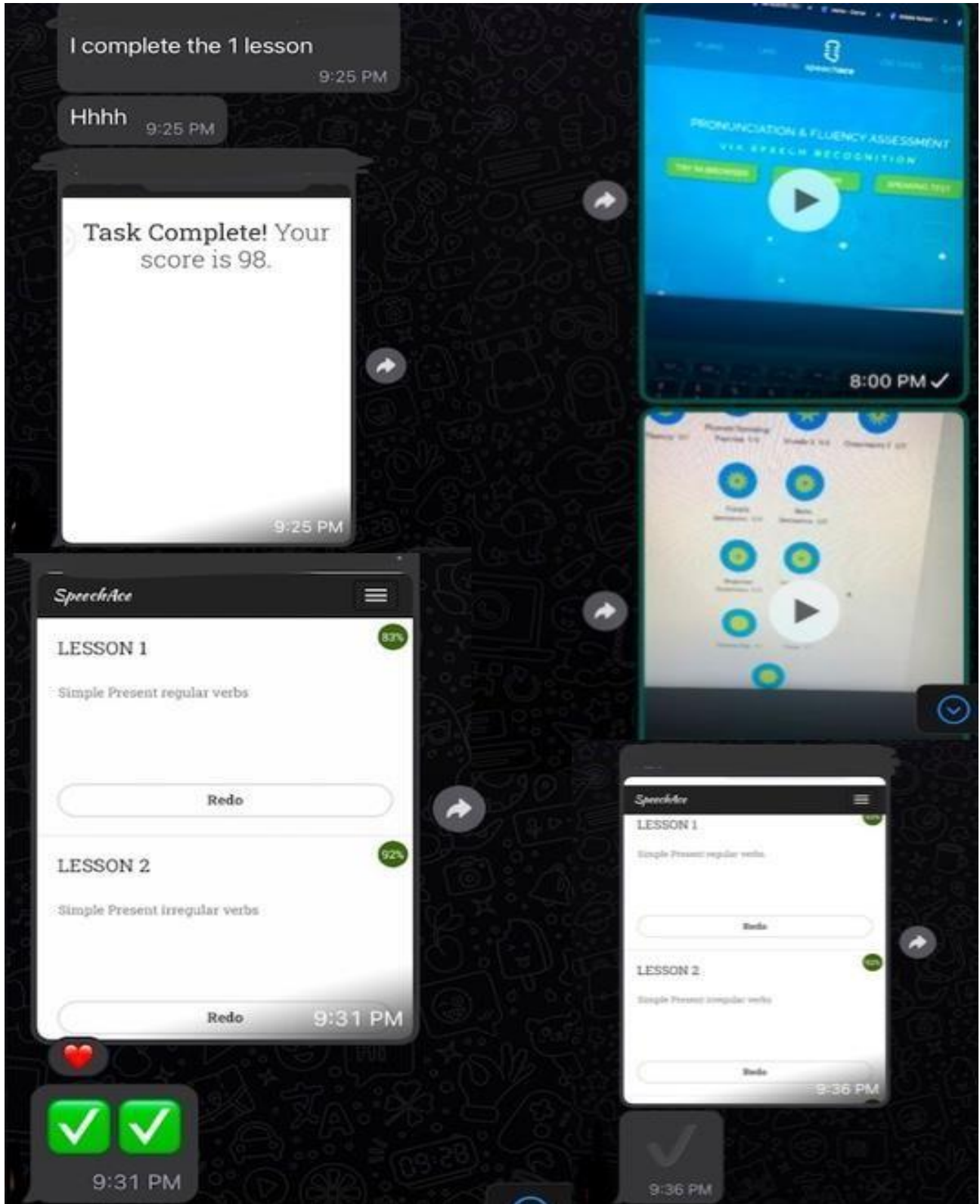
Actually i faced a problem

9:08 PM

How can I create an account?
Tell me

8:26 PM

Everyone, I want you to
create accounts on the
website I will send you. (If you
can access the accounts we
already opened, do it. If not,
please open new accounts)
(Remember the email and
password and send them to
me in a private message here
on WhatsApp). Log in into the
site and I will show you
everything in a video and
what you should work on.

7:57 PM ✓

You got it!   9:16 PM ✓

And me it works now 🙂

9:20 PM

Task Complete! Your
score is 93.

You
🎥 Video

Aaaaa Thanks

8:36 PM

👍

Me too 😂

9:26 PM

When i try to use the microphone it appears this

9:11 PM

For me it works

9:19 PM

I live

You
I don't understand the problem

me too 😑

9:10 PM

يتعذّر على الموقع الإلكتروني هذا طلب الإذن

يرجى إغلاق أي علامات محادثات أو تركيبات من التطبيقات الأخرى، ثم إعادة المحاولة مرة أخرى.

أعد المحاولة   إلغاء

9:08 PM

When you used your mobile, is it working normally?

9:08 PM ✓

Please share with us some screenshots to see if it works fine 😊

9:08 PM ✓

📷 Photo

Excellent job 👏👏
👏 are you enjoying it?

9:57 P

Okay i will start 🔋

8:24 PM

And good luck 8:25 PM

Yes i do 😍

9:57 PM

Now

9:18 PM

It works 9:18 PM

Thank u 9:18 PM

Same 😊

9:58 PM

Appendix I: Post-Training Interview

<u>Question 1:</u> You have been training your pronunciation using an ASR-based technology, do you prefer training your pronunciation in the classroom or online?

<u>Question 2:</u> How would you describe/rate your experience of pronunciation training using the ASR-based technology?

<u>Question 3:</u> Do you think that this training helped you improve your pronunciation? How?

<u>Question 4:</u> What did you feel/think about the feedback and scores that were provided by the ASR technology during the test and the training?

<u>Question 5:</u> Do you think that the scores were fair and measured your pronunciation and fluency level correctly?

<u>Question 6:</u> Did you find difficulties in understanding/interpreting the feedback and the scores/percentages that were provided by the ASR technology? Please specify why?

<u>Question 7:</u> Did you like/enjoy the set of activities that you participated in during the training? What kind of activities did you like/enjoy the most?

Question 8: Did you encounter any obstacles/difficulties while using the ASR technology? What are these obstacles related to?

Question 9: Do you think that the ASR technology helped you make progress in your speaking fluency as well? How?

Question 10: Now that you trained your pronunciation using ASR technology, are you motivated to use more technological tools, platforms, software, or applications to improve different skills?
What kind of technology do you want to use? Why?

Question 11: Do you think that your teachers should incorporate technology more in the classroom? What kind of technology? And in what module?

Résumé

L'instruction efficace en prononciation anglaise nécessite une exposition à une entrée authentique, de la pratique et des retours. De nos jours, les technologies avancées peuvent être utilisées pour enseigner et évaluer la prononciation. Par conséquent, cette recherche mixte avait pour objectif d'étudier l'effet de la technologie de reconnaissance automatique de la parole (ASR) sur la prononciation des étudiants en anglais langue étrangère (EFL) au Département d'anglais et de littérature de l'Université de Biskra. Dans ce but, un questionnaire semi-structuré a été administré à un échantillon de quatre enseignants d'expression orale et de phonétique dans la phase initiale, en plus de l'utilisation d'un dispositif quasi-expérimental impliquant un groupe témoin et un groupe expérimental, chacun comprenant douze participants (n=12). L'intervention a été suivie d'un entretien semi-structuré mené avec les participants du groupe expérimental. Les quatre enseignants ont convenu qu'il est important d'intégrer la technologie dans la classe, et ils ont souligné qu'une planification et une formation adéquates sont essentielles avant d'introduire un outil technologique aux étudiants. De plus, les résultats de l'intervention ont révélé l'importance de l'utilisation de l'ASR pour améliorer la prononciation des étudiants. En outre, les réponses positives à l'entretien ont complété les résultats et ont également montré que les participants ont apprécié l'utilisation de la technologie ASR. Les résultats de la recherche ont souligné les avantages potentiels de l'intégration de la technologie ASR dans l'enseignement de la prononciation pour faciliter le processus d'apprentissage et conduire à de meilleurs résultats.

Mots-clés : Prononciation ; Reconnaissance automatique de la parole ; Technologie ; Apprenants en anglais langue étrangère (EFL)

الملخص

يتطلب التعليم الفعّال للنطق باللغة الإنجليزية التعرض للإدخال الأصيل، والممارسة، وملاحظات تصحيحية. في الوقت الحاضر، يمكن استخدام التقنيات المتقدمة لتعليم وتقييم النطق. لذا، كانت هذه الدراسة التي تعتمد على المنهج المختلط تهدف إلى التحقيق في تأثير تكنولوجيا التعرف التلقائي على الكلام (أي اس ار) على نطق طلاب اللغة الإنجليزية كلغة أجنبية في قسم اللغة الإنجليزية والأدب في جامعة بسكرة. ولهذا الغرض، تم تنفيذ استبيان شبه منظم لعينة من أربعة مدرسين للتعبير الشفوي والصوتيات في المرحلة الأولية، بالإضافة إلى استخدام تصميم شبه تجريبي يشمل مجموعة ضابطة وأخرى تجريبية، كل منهما يضم اثنتي عشرة مشاركًا (١٢ =ع) تلت التدخل مقابلة شبه منظمة أُجريت مع مشاركي المجموعة التجريبية. واتفق الأساتذة الأربعة على أن تضمين التكنولوجيا في الفصل الدراسي أمر مهم، وشددوا على أهمية التخطيط الكافي والتدريب قبل تقديم أداة تكنولوجية للطلاب. بالإضافة إلى ذلك، كشفت نتائج التدخل عن أهمية استخدام تقنية التعرف التلقائي على الكلام لتحسين نطق الطلاب. وعلاوة على ذلك، أكدت الردود الإيجابية في المقابلة النتائج وأظهرت أيضًا أن المشاركين استمتعوا باستخدام تكنولوجيا تقنية التعرف التلقائي على الكلام. أبرزت نتائج البحث الفوائد المحتملة لدمج تكنولوجيا تقنية التعرف التلقائي على الكلام في تعليم النطق لتسهيل عملية التعلم وتحقيق نتائج تعلم أفضل.

الكلمات الرئيسية: النطق؛ التعرف التلقائي على الكلام؛ التكنولوجيا؛ طلاب اللغة الإنجليزية كلغة أجنبية (EFL)