Talking with Machines: Exploring AI as a Tool to Improve the English Pronunciation of Spanish-Catalan Teenagers

by

Marina Palomo Martín

Submitted to the

Department of English and German Studies

in partial fulfillment of the requirements for the degree of

Master's in Teaching and Learning English as a Foreign / Second Language

at

UNIVERSITAT ROVIRA I VIRGILI

June 13, 2024

Signature of Student:

MARINA PALOMO I MARTÍN - DNI 23899050A

Digitally signed by MARINA PALOMO I MARTÍN - DNI 23899050A
Date: 2024.06.13 19:22:22 +02'00'

………………………………

Certified by Dr. /Ms./Mr. **Joaquín Romero Gallego,** Master's Thesis Supervisor

Signature of Supervisor:

Joaquín Romero

Digitally signed by Joaquín Romero
Date: 2024.06.13 19:50:30 +02'00'

………………………………

UNIVERSITAT ROVIRA I VIRGILI

MATLEFS

Master's in Teaching and Learning English as a Foreign/Second Language
2023-2024

Master's Final Project
ORIGINAL WORK FORM

STUDENT

Last Name(s): Palomo Martín

First Name: Marina

I hereby state that all the work presented as part of my Final Master's Project is original with no exception and that I have not, voluntarily or otherwise, misused or misreported any previously published information. I am aware that any failures to comply with these statements will automatically result in disqualification of my final paper and that I will not be able to obtain any credits for it.

MARINA PALOMO I MARTÍN - DNI 23899050A

Digitally signed by MARINA PALOMO I MARTÍN - DNI 23899050A
Date: 2024.06.13 19:22:47 +02'00'

Signature of Student

In Tarragona, June 13, 2024.

Abstract

This study explores the potential of artificial intelligence (AI) to enhance pronunciation skills in English as a Foreign Language (EFL) learners, focusing on American English vowel sounds. An introduction to AI is given at the outset of the research, followed by a thorough investigation of the importance of pronunciation instruction and learning, the intricacies of the English vowel system, and the difficulties encountered by Spanish-Catalan learners. Subsequently, the role of AI in second language (L2) learning and methods for integrating AI into classroom practices are discussed. To assess the efficacy of the AI-powered tools, Siri and the Speakometer app are examined in terms of the speech perception and production of 24 Spanish-Catalan teenagers. Participants were split into two groups, receiving identical explicit instruction over four sessions. However, one group practiced traditional methods, and the other used the Speakometer app. Before and after the intervention, the participants engaged with Siri while being recorded. A post-questionnaire was administered to assess their experiences with pronunciation learning. The findings imply that AI-powered applications do not significantly enhance pronunciation learning or the development of a more native-like accent. Additionally, the type of stimuli and vowel sounds produced affect AI comprehension. Based on this study, it can be said that the efficiency level required for effective pronunciation practice in EFL settings has not yet been reached by existing AI tools.

Keywords: *AI, EFL, pronunciation instruction, vowel sounds, AI-powered tools for pronunciation*

Acknowledgments

First and foremost, I would like to express my deepest appreciation to my supervisor, Dr. Joaquín Romero, for his unwavering patience, support, and guidance throughout this thesis. Thank you for introducing me to the world of phonetics and phonology two years ago. Without your inspiration, none of this would have been possible.

I am also profoundly grateful to Escola Elisabeth, Yasmina, and all the students who eagerly participated in this study. Bringing pronunciation knowledge into the EFL classroom was a rewarding experience, and it was delightful to see teenagers enjoy learning about it.

To my parents and sister, thank you for being my anchor, always believing in me, and loving me unconditionally. This work is dedicated to you. Additionally, my heartfelt thanks go to Eloy for staying by my side and bringing love, peace, and positivity into my life. You make everything so much easier.

I had the pleasure of sharing this journey with Adrián. I truly enjoyed our discussions about pronunciation and the friendship we have built. Thank you for your companionship and support.

Lastly, I would like to thank you, the reader, for taking the time to go through these pages. Your interest and attention are greatly appreciated.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AI	Artificial Intelligence
ASR	Automatic Speech Recognition
CAPT	Computer-Assisted Pronunciation Training
EFL	English as a Foreign Language
F1	First formant
F2	Second formant
IPA	Intelligent Personal Assistant
L1	First or native language
L2	Second or foreign language
MRI	Magnetic Resonance Imaging
NLP	Natural Language Processing
PAM-L2	Perceptual Assimilation Model of Second Language Speech
SLA	Second Language Acquisition
UTI	Ultrasound Tongue Imaging

1. Introduction

The landscape of language learning is currently experiencing a profound and transformative shift. The rapid advancements in artificial intelligence (AI) and the widespread availability of online platforms are reshaping how individuals approach language acquisition. Nevertheless, amidst these changes, pronunciation instruction remains largely overlooked (Park, 2015). As a result, even advanced second or foreign language (L2) learners often encounter challenges in oral communication, struggling to overcome their strong non-native accents, which can hinder their fluency and potentially lead to misunderstandings (Saito et al., 2015). Although the emergence of AI-powered tools and applications presents new possibilities for pronunciation practice, several questions remain unanswered: Can AI truly enhance L2 learners' speech production and make it sound more native-like? Is AI prepared for didactic purposes, or is its comprehension limited?

1.1. Artificial Intelligence

Understanding what AI is begins with exploring the concept of intelligence itself. Intelligence is a multifaceted and debated term with no universally accepted definition. Some view intelligence as the human mental and emotional ability to reason, solve problems, understand concepts, and learn (Morandín-Ahuerma, 2022), while others believe that multiple intelligences exist, varying among individuals (Gardner, 1983). This ambiguity naturally extends to AI.

John McCarthy, who coined the term in the late 1950s, described AI as “the science and engineering of making intelligent machines” (McCarthy, 2007, p. 2). In essence, AI is a term to describe a set of technologies that enable machines to simulate and perform tasks that typically require human intelligence and problem-solving skills without explicit human guidance (Morandín-Ahuerma, 2022). It is also seen as a branch of computer science focused on creating intelligent entities that learn from available data and use that knowledge to make predictions (Russell & Norvig., 2010). Russell and Norvig (2010) propose four categories of AI placed along two dimensions: the goals of AI (matching human performance or ideal rationality) and their capabilities (reasoning/thinking or acting/behaving) (Bringsjord & Govindarajulu, 2018).

1.1.1. The Anatomy of AI

Given the intricacy of AI's inner workings, this study will only delve into the fundamental building blocks that drive these technologies. AI encompasses a wide range of techniques and processes, each representing a specialized subfield that contributes to the overarching science of AI (CSU Global, 2021). As outlined by Sofia et al. (2020), these subfields can be grouped into two main areas based on the system's capabilities: (1) Reasoning and decision-making, which involves transforming real-world data into a machine-readable format, and (2) Learning and perception, where knowledge is extracted from data in both structured and unstructured formats, such as text, speech, images, and sounds.

Within the area of learning and perception, there is machine learning, a core element of AI that empowers computers to learn from data and previous computations and make increasingly accurate predictions through pattern recognition (Choi et al., 2020; Janiesch et al., 2021; Sofia et al., 2020). Over the past decades, machine learning has witnessed remarkable advancements, including the development of artificial neural networks into deep neural networks with enhanced learning capabilities (Janiesch et al., 2021).

Inspired by the structure and function of the human brain, artificial neural networks comprise interconnected nodes, or artificial neurons, that process information by transmitting signals through weighted connections. These weights are fine-tuned during training, allowing the network to learn and improve. Deep learning, a subfield of machine learning, arises from constructing artificial neural networks with an extensive number of neurons arranged in multiple layers (Choi et al., 2020; Janiesch et al., 2021; Levis & Suvorov, 2012; Lippmann, 2023; Montesinos et al., 2022).

Even though both machine learning and deep learning utilize neural networks, the main distinction lies in their architectural structure. The former employs shallower networks consisting of an input layer, one or two hidden layers, and an output layer (Choi et al., 2020), whereas the latter harnesses deep neural networks that frequently comprise tens to hundreds of hidden layers. This increased depth allows deep learning to handle large datasets more effectively, outperforming traditional machine learning algorithms in processing text, image, speech, and audio data (Janiesch et al., 2021).

Two other essential components within AI's internal structure are natural language processing (NLP) and automatic speech recognition (ASR). To understand and simulate how humans communicate, AI relies on NLP, which employs methods to understand and generate natural language (Adamopoulou & Moussiades, 2020). However, this task is highly complex, as language can sometimes be ambiguous and variable. During interactions, people usually use idioms and collocations whose meaning is not necessarily related to the meaning of their parts, adding an extra layer of complexity for AI (Coheur, 2020).

On the other hand, ASR refers to the technology that decodes and transcribes spoken language. This process involves taking acoustic input, analyzing it, and converting it into text to enable AI to understand user queries and respond appropriately (Dizon, 2017; Levis & Suvorov, 2012; Pieraccini, 2012). ASR works with natural language understanding components in speech interfaces to interpret user intentions or commands (Clark et al., 2019).

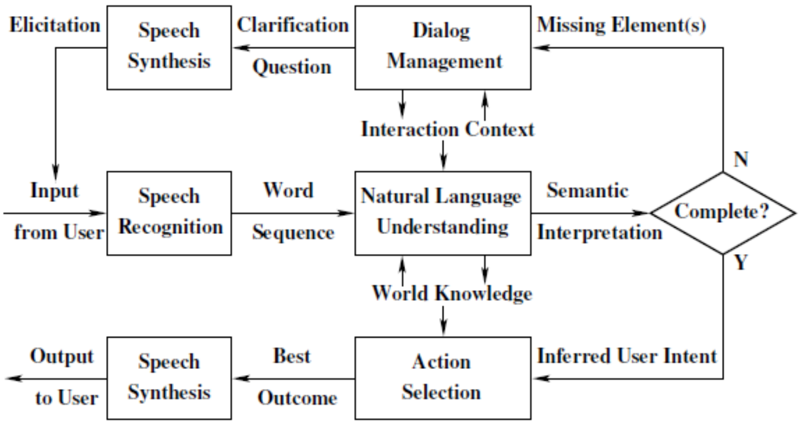
1.1.2. Types of AI and their Applications

Generally speaking, there are two main types of AI: weak or narrow AI and strong or general AI. As Liu (2021) and Martinez (2019) described, weak AI is designed to replicate or simulate specific tasks but requires human interaction to work. In contrast, strong AI is a hypothetical form of AI with generalized human cognitive abilities that can find solutions without human intervention. Today, all existing instances of AI fall under the category of weak AI because a strong AI that fully mirrors human-like intelligence has not yet been achieved (Sofia et al., 2020). Common everyday applications of weak AI include image and face recognition systems, machine translation tools, self-driving cars, e-learning platforms, language learning apps, and chatbots (Sofia et al., 2020; Svenningsson & Faraon, 2019).

Chatbots went a step further by creating personal voice assistants, also known as virtual agents, digital assistants, or intelligent personal assistants (IPAs). These software agents are voice recognition technologies built into smartphones or dedicated home speakers that facilitate human-like interactions and can handle various tasks (Adamopoulou & Moussiades, 2020; Hoy, 2018). Apple's Siri, Amazon's Alexa, Google's Assistant, and Microsoft's Cortana are a few of the most well-known examples of IPAs.

IPAs follow a basic input-process-output cycle. The systems receive data from humans or other sources (input), manipulate it to extract valuable insights (process), and then generate meaningful information that can be used (output) (Paschen et al., 2020). Building on Bellegarda's (2013) model (see Figure 1), AI interaction follows a defined sequence. First, the user's spoken input is converted into a sequence of words, which undergoes natural language understanding techniques to extract the semantic meaning. Dialogue management employs the interaction context to address potential information gaps and prompt the user for missing details. Lastly, by combining the extracted meaning with its knowledge of available tasks, the system determines the most appropriate action and delivers the corresponding output to the user.

Figure 1. *Overview of IPA-User Interaction*



Note. Figure retrieved from Bellegarda (2013).

1.2. Pronunciation in the EFL Classroom

One of the most challenging aspects of second language acquisition (SLA) is pronunciation, and still, it is highly underemphasized in language and teacher-training programs (Darcy, 2018). As Jarosz (2019) observes, EFL curricula within educational settings tend to ignore phonetic instruction, presuming L2 students will naturally pick up pronunciation skills over time. Many EFL teachers are reluctant to deal with pronunciation because they do not know how to do it and lack the resources; they believe there is not enough time to do so, or they consider that the effort invested in teaching pronunciation and providing feedback usually yields unsatisfactory outcomes. Thus, they tend to prioritize other areas of language learning (Darcy, 2018; Macdonald, 2002; Levis & Grant, 2003; Park, 2015; Szpyra-Kozłowska, 2014). Unfortunately, this neglect can disadvantage students' learning, leading to incorrect pronunciation of English sounds, impaired perception of English speech, and even withdrawal from using the language altogether (Baran-Łucarz, 2014).

1.2.1. The Importance of Pronunciation in EFL Learning

The ultimate goal of SLA is to equip learners with the skills necessary to communicate effectively in the target language, whether with native or non-native speakers. Notwithstanding, studies indicate that non-native speakers often experience anxiety when using English, regardless of their language proficiency, mainly due to self-perceived deficiencies in pronunciation abilities (Baran-Łucarz, 2014; Liu, 2006).

For successful oral exchange, non-native speakers' speech must be intelligible. Setter and Jenkins (2005) argue that if pronunciation hinders word comprehension, the chances of communication breakdown increase. Similarly, Derwing and Munro (2005) highlight that foreign-accented speech can affect both speakers and listeners in terms of perception and production. Jarosz (2019) even asserts that native speakers sometimes find pronunciation errors more annoying than lexical or grammatical mistakes. This underscores the importance of form alongside content, as it can greatly shape first impressions. As a result, strong accents can trigger negative biases, potentially leading to social marginalization and discrimination, impacting non-native speakers' social and professional spheres (Derwing & Munro, 2005; Szpyra-Kozłowska, 2014).

To prevent these issues and foster mutual understanding, it is vital to implement well-structured pronunciation instruction (Jarosz, 2019). A common misconception is that exposing learners to spoken English will automatically lead to correct pronunciation. While listening activities play a significant role in the process, they should be supplemented with more targeted practice. Celce-Murcia et al. (2010) propose a five-stage framework for teaching pronunciation communicatively. The model starts by raising learners' awareness of a specific pronunciation aspect through description and analysis. This is followed by listening discrimination or perception activities. Subsequently, learners engage in oral exercises that progress through controlled, guided, and ultimately communicative exercises to solidify the learned pronunciation feature. In this learning process, EFL teachers should act as appropriate models for their students, aiding them in grasping the sound system of English and providing feedback on their speech production. They should emphasize the importance of pronunciation, as phonetic errors can lead to misunderstandings and even communication breakdowns (Szpyra-Kozłowska, 2014). The focus should be not only on fluency (i.e., the capacity to produce language without undue pause or hesitation) but also on accuracy (i.e., conforming to target language norms), as the two are intertwined (Celce-Murcia et al., 2010).

Learning pronunciation can help EFL learners become better speakers and bolster their confidence to engage in L2 conversations (Szpyra-Kozłowska, 2014). Hence, it is incumbent upon educators to weave pronunciation instruction into their teaching framework, utilizing methods such as explicit instruction, fostering awareness, and integrating speaking-oriented activities (Levis & Grant, 2003). While learners should be encouraged to reach their full potential, the primary aim of incorporating pronunciation teaching within EFL environments should not merely be attaining native-like pronunciation. Instead, the focus should pivot towards improving intelligibility (i.e., the listener's competence to discern and interpret specific phonetic and linguistic components) and comprehensibility (i.e., the extent to which the listener can grasp the speaker's intended meaning) (Alharthi, 2024; Celce-Murcia et al., 2010; Derwing & Munro, 2005; Szpyra-Kozłowska, 2014).

1.2.2. The Sound System of American English: Vowels

The foundations of any language's pronunciation lie in the fields of phonetics and phonology, both of which examine the mechanics of linguistic sound but from different perspectives. As Odden (2013) states, phonetics studies the physical manifestation of speech sounds, encompassing areas such as acoustics (waveforms and formants), duration, and articulation (involvement of vocal tract resonances and muscles in producing sound). In contrast, phonology deals with the mental representation of these sounds and the abstract rules governing them. Phonological features can be categorized into segmental (i.e., vowels and consonants) and suprasegmental, which transcends individual sounds (e.g., stress, pitch, rhythm, tone, or intonation).

Describing vowel sounds presents a greater challenge than consonants due to the absence of articulator contact (Celce-Murcia, 2010; Ladefoged & Johnson, 2014). All English vowel sounds are voiced, but their characteristics are determined by the positioning of the tongue, jaw, and lips. Subtle adjustments in the shape of the oral cavity affect the airflow, resulting in distinct resonances that differentiate each vowel sound (Barber et al., 2012). To adequately describe vowel sounds, researchers like Avery and Ehrlich (2013) and Celce-Murcia (2010) emphasize the importance of four parameters:

(1) Height: The vertical placement of tongue within the mouth (high, mid, or low),

(2) Frontness: The horizontal position of the tongue (front, central, or back),

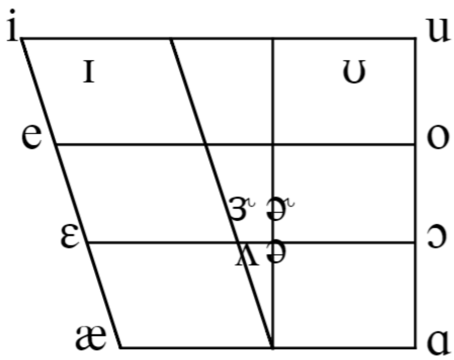
(3) Tenseness: Whether the muscles are tense or lax,

(4) Rounding: The position of the lips (rounded or unrounded/spread).

Even so, it is important to acknowledge the controversial nature of tenseness. Scholars like Raphael and Bell-Berti (1975) have studied this topic and found that muscular tension does not consistently correlate with the production of tense-lax distinctions. Besides articulation differences, another distinguishing factor that sets tense and lax vowels apart is their distribution within words. Lax vowels require a consonant following them and cannot appear at the end of words, whereas tense vowels do not have this limitation (Wayland, 2018).

To examine the range of tongue positions involved in vowel production, the "vowel chart, quadrilateral, or space" is used. However, despite the existence of the parameters mentioned above, the vowel chart classifies vowels based on height and frontness (see Figure 2). General or Standard American English has a vowel inventory of at least 14 sounds, each represented by specific symbols from the International Phonetic Alphabet. These sounds include /i/, /ɪ/, /e/, /ɛ/, /æ/, /ɑ/, /o/, /ɔ/, /ʊ/, /u/, /ʌ/, and /ə/, as well as the two r-colored vowels (i.e., vowels affected by the /r/ sound) /ɝ/ and /ɚ/ (N.B., /ə/ and /ɚ/ are exclusively used in unstressed position). Additionally, there are three diphthongs: /aɪ/, /ɔɪ/, and /aʊ/ (Avery & Ehrlich, 2013). Among these vowel sounds, six of them are rounded, namely /o/, /ɔ/, /ʊ/, /u/, /ɝ/, and /ɚ/. In terms of tenseness, the sounds /i/, /e/, /ɔ/, /o/, /u/, and /ɝ/ are considered tense, while /ɪ/, /ɛ/, /æ/, /ɑ/, /ʊ/, /ʌ/, /ə/, and /ɚ/ fall into the lax category (Romero, n.d.-b).

Figure 2. *General American English Vowel Chart*



Note. Vowel chart retrieved from Romero (n.d.-b).

The acoustics of vowel production can be explained through the source-filter theory of speech, which postulates that sounds are generated and then shaped or filtered by the vocal tract acting as a resonating tube (Huckvale, 2017). Vowel sounds, as noted by Ladefoged and Johnson (2014), are produced with different resonances simultaneously: an inherent pitch and multiple harmonics that confer upon it a distinguishing quality (i.e., the way the vowels sound to the ear) (Kenyon, 1997). These harmonics, called “formants,” can help distinguish one vowel from another. Formants are concentrations of acoustic energy at certain frequencies that can be measured in Hertz (Hz). The source of sound is the vibration of the vocal folds, whereas the vocal tract acts as a filter that modulates the quality of the vowel (Huckvale, 2017; Ladefoged & Johnson, 2014).

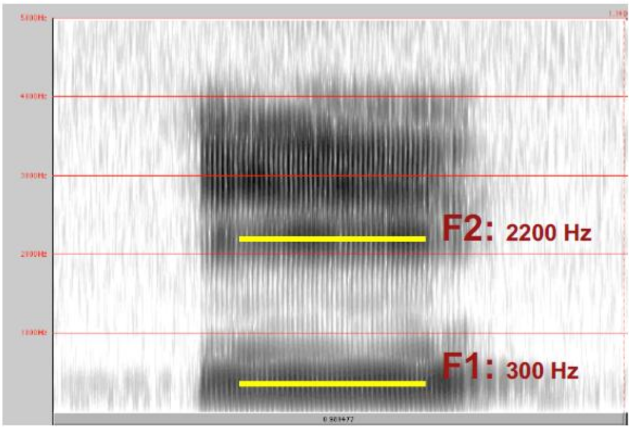
Regarding vowel quantity or length, Kenyon (1997) explains that in phonetics, it refers to duration. In British English phonetic transcription, the symbol [ː] is placed after vowels to denote that they are long. However, since vowel length is seldom distinctive in American English, the length mark can be disregarded. Kenyon (1997) notes that the length of a vowel in American English depends on its phonological context, more precisely, what comes after the vowel. For instance, vowels will be longer before a voiced consonant, shorter before a voiceless consonant, and longer when followed by a sonorant.

When acoustically describing vowel sounds, the most important formants are the first three. The first formant (F1) is inversely related to vowel height: the higher the vowel, the lower the F1, and vice versa. The second formant (F2) is somewhat related to the degree of the vowel’s frontness. Front vowels demonstrate high F2, whereas back vowels exhibit low F2. However, F2 is influenced by frontness and lip rounding. (Ladefoged & Johnson, 2014; Ogden, 2009).

To mitigate the effect of lip rounding, F2 can be considered in relation to F1. The degree of frontness is best represented by the difference between F1 and F2 frequencies. “The closer they are together, the more “back” a vowel sounds” (Ladefoged & Johnson, 2014, p. 208). The third formant (F3), on the other hand, plays a minor role in distinguishing vowels. For all English vowels, except /ɝ/, the frequency of F3 can be predicted from the frequencies of F1 and F2 (Ladefoged & Johnson, 2014).

Vowel formants can be observed and analyzed using spectrograms, which are three-dimensional graphs that represent time on the horizontal axis, frequency (Hz) on the vertical axis, and amplitude (i.e., intensity) in a grey scale (Ladefoged & Johnson, 2014). By examining the dark horizontal bands in the spectrogram, which indicate concentrated energy at specific frequencies, it is possible to determine the vowel sound produced by a speaker (Ladefoged & Johnson, 2014). In spectrograms, formants are counted from the bottom upwards. The lowest band corresponds to F1, and the band immediately above it corresponds to F2 (see Figure 3 for an example of F1 and F2 in a spectrogram).

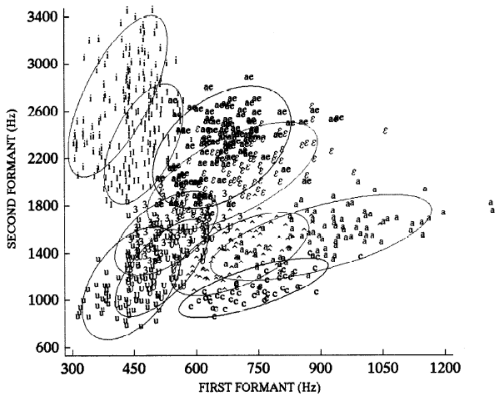
Figure 3. *Vowel* /i/ *Displayed in a Spectrogram*



Note. Image retrieved from Romero (n.d.-a).

Formant frequencies, however, can vary among individuals and are influenced by factors like gender, age, dialect, speaking style, speech rate, and the speaker’s anatomy (e.g., the shape and length of the vocal tract). There can even be differences within the same region or dialect (Cavalcanti et al., 2021; Ladefoged & Johnson, 2014). In a study by Hillenbrand et al. (1995), the average formant frequencies of vowels were calculated based on data collected from 46 men, 48 women, and 46 children from Michigan, Illinois, Wisconsin, Minnesota, Ohio, and Indiana. The findings revealed significant variability in the production of F1 and F2, as depicted in Figure 4.

Figure 4. *Values of F1 and F2 for American Men, Women, and Children*



Note. Figure retrieved from Hillenbrand et al. (1995) showing F1 and F2 production of men, women, and children for 10 vowels with ellipses fit to the data ("ae" = /æ/, "a" = /ɑ/, "c" = /ɔ/, "ʌ" = /ʌ/, and "3" = /ɝ/). Measurements for /e/ and /o/ were omitted.

Considering the variability of formant frequencies, it becomes imperative to establish standardized reference values for comparing and evaluating vowel production in American English (see Table 1).

Table 1. *Standard Formant Frequencies for American English Vowels*

	/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	/ʌ/	/ɝ/
F1	300	375	475	600	750	800	625	525	450	300	700	850
F2	2200	2000	1900	1900	1800	1150	800	1200	1150	1200	1250	1350

Note. Table adapted from Romero (n. d.-a).

1.2.3. Spanish-Catalan Speakers Learning EFL Vowel Pronunciation

The vowel systems of Spanish and Catalan exhibit characteristics that can potentially influence the learning experience of English, which has a broader range of vowel sounds (Coe, 2001; Gómez González & Sánchez Roura, 2016).

In Spanish, there are five monophthongal vowel sounds (/i/, /e/, /a/, /o/, and /u/) characterized by being tense, short, and exhibiting minimal variation as a result of contextual phonetic processes (Martínez Celdrán & Elvira-García, 2019). These vowels are generally produced higher and more towards the front of the mouth compared to English vowels (Gómez González & Sánchez Roura, 2016). On the other hand, Eastern Catalan has a broader range of vowel sounds, encompassing the Spanish vowels and three additional ones: /ɛ/, /ɔ/, and /ə/. Nonetheless, the vowel inventory of both languages is smaller than the English vowel system, making it difficult for Spanish-Catalan learners of EFL to perceive or pronounce new phonemes (Garita Sánchez et al., 2019).

Many studies have investigated the acquisition of L2 speech, with particular emphasis on the Perceptual Assimilation Model (PAM) (Best, 1994) and its extension, PAM-L2 (Best & Tyler, 2007). These models posit that upon encountering unfamiliar sounds in the L2, learners filter and categorize these sounds through the phonological system of their L1. Essentially, learners assimilate new L2 sounds into the most articulatorily-similar native phoneme (Cebrian, 2006; Rallo Fabra & Romero, 2012; Rallo Fabra & Tyler, 2021; Yavaş, 2011).

As a result of this phonological assimilation process, Spanish-Catalan speakers often identify the English /æ/ as their native /a/ sound. Additionally, they perceive the English vowel pairs /i/-/ɪ/ and /u/-/ʊ/ as their native /i/ and /u/. The latter assimilation arises due to the lack of tenseness as a distinctive feature in either Spanish or Catalan, making it complicated for them to discern between tense-lax vowel pairs (Cebrian, 2006; Gómez González & Sánchez Roura, 2016).

Rallo Fabra and Romero (2012) scrutinized the disparities in formant frequencies between the two languages in a comparative study of American English and Eastern Catalan vowel inventories. They found that the Catalan /i/ vowel sound is intermediate between English /i/ and /ɪ/ sounds and that the formant frequencies of English /ɪ/ closely resemble those of the Catalan /e/. Their study also noted that the English /æ/ sound is acoustically located between the Catalan /ɛ/ and /a/ sounds and that the English sounds /ʌ/ and /ɑ/ do not seem to have a Catalan equivalent. In Spanish and Catalan, /a/ is a central vowel with higher F2 frequencies than the English sounds /ʌ/ and /ɑ/. Just like in English, it is important to remember that formant frequencies will depend on a list of factors (see Table 2 for a comparison of Spanish and Catalan vowel sounds).

Moreover, Rallo Fabra (2022) further underscored that Spanish-Catalan bilingual speakers are exposed to written English input long before they can speak the language, which has profound implications for their pronunciation skills. Without corrective feedback from teachers, pronunciation errors become internalized and fossilized.

Table 2. *Formant Frequency Values for Spanish and Catalan Vowels*

	F1		**F2**	
Vowel	*Spanish*	*Catalan*	*Spanish*	*Catalan*
/i/	313	334	2200	2078
/e/	457	450	1926	1839
/ɛ/	–	581	–	1700
/a/	699	730	1471	1358
/ɔ/	–	608	–	1125
/o/	495	489	1070	1047
/u/	349	394	877	960

Note. This table is adapted from Martínez Celdrán (1995) and Recasens (2014). This thesis takes these articles' F1 and F2 values for Spanish and Catalan vowels as standard.

When comparing the orthographic transparency among the three languages, Rallo Fabra (2022) asserts that Spanish and Catalan demonstrate high transparency with consistent one-to-one sound-spelling correspondence. In contrast, English has an opaque orthography, resulting in multiple pronunciations for a given spelling (e.g., the grapheme <i> is pronounced as /aɪ/ in *high* but as /ɪ/ in *hit*). This mismatch between graphemes and phonemes can present challenges for Spanish-Catalan EFL learners when pronouncing English words (Rallo Fabra, 2022).

A specific example of this challenge is the vowel pair /ɝ/-/ɔr/. In American English, r-coloring occurs when certain orthographic vowel + /r/ sequences occur within the same syllable. "The vowel anticipates and glides towards the central /r/ position and takes on some of the retroflex quality of /r/" (Celce-Murcia et al., 2010, p. 127). Since Spanish and Catalan do not have r-colored vowels in their inventories, they often replace the /ɝ/ sound with the vowel sound corresponding to the spelling plus an /r/ (e.g., pronouncing the word *world* with /or/ or /ɔr/) (Coe, 2001; Flege & Wayland, 2019). Moreover, Spanish-Catalan speakers struggle to discern between /ɝ/ and /ɔr/, especially in words spelled with -or- (e.g., *word, work,* or *world*) and in minimal pairs such as *worm* /wɝm/ vs. *warm* /wɔrm/.

1.2.4. Imaging Techniques in Pronunciation Teaching

In vowel pronunciation teaching, EFL teachers frequently concentrate on the symbols from the International Phonetic Alphabet, the four phonological parameters, and diagrams to illustrate these abstract features. However, explaining precise tongue movements during sound production can be challenging. This is where tools like magnetic resonance imaging (MRI) and ultrasound tongue imaging (UTI) come into play, as they offer real-time, anatomically detailed visuals of articulator movements during speech (Bliss et al., 2017; Holt, 2018; Ladefoged & Johnson, 2014). These tools can aid L2 learners in better understanding tongue placement for producing English vowel sounds and, consequently, improve their pronunciation.

With MRIs, students can only observe another individual's speech production through resources like the Seeing Speech website (Lawson et al., 2018b). The advantage of MRIs is their ability to capture "not only lingual, labial and jaw motion, but also articulation of the velum, pharynx and larynx, and structures such as the palate and pharyngeal wall” (Toutios et al., 2016, p. 2428).

In contrast, UTI can be used directly in the classroom, allowing students to see their own tongue movements during speech production and providing visual feedback (Bliss et al., 2017). Alternatively, teachers can use the Dynamic Dialects website for UTI videos (Lawson et al., 2018a). However, unlike MRIs, UTI cannot image through bone or air, only providing visualization of the tongue (Gick et al., 2008).

1.3. The Role of AI in Pronunciation Teaching and Learning

In today's world, technology has become an integral part of every aspect of human life, and SLA is no exception. Among the various technological advancements, AI has garnered significant attention for its potential to enhance and support students' learning in the EFL context (Dakakni & Safa, 2023). However, amidst this technological revolution, L2 teachers and students ponder: Should they adhere to traditional approaches or embrace the opportunities presented by AI? And how effective and pedagogically valuable are these AI-based sources in improving pronunciation skills in a new language? (Rogerson-Revell, 2021).

1.3.1. Integrating AI in Pronunciation Instruction

Over the past decade, the field of computer-assisted language learning (CALL) has witnessed tremendous growth, particularly in the realm of computer-assisted pronunciation training (CAPT) tools, with the appearance of new mobile applications and IPAs that leverage the power of AI and ASR technology (Dizon, 2017; Neri, Mich, et al., 2008; Rogerson-Revell, 2021).

AI-based systems offer plenty of benefits for EFL learners, including individualized attention, the possibility to compare their pronunciation with that of native speakers, automatic feedback, self-regulated learning, and the flexibility to practice the L2 at a time and location of their choosing (Dizon, 2017; El Shazly, 2021; Fouz-González, 2020). Furthermore, AI tools have the potential to provide learners with an authentic means of communication (Moussalli & Cardoso, 2016) and facilitate improvements in their perception and production of segmental and suprasegmental features (Fouz-González, 2020).

Numerous studies evidence the beneficial role of AI in increasing student motivation and willingness to communicate in a foreign language. Moussalli and Cardoso (2016) reported high levels of enjoyment from part of language students during their interactions with the IPA Alexa. Likewise, Neri, Mich, et al. (2008) found that students valued and profited from using CAPT tools, particularly in improving their L2 pronunciation. Dakakni and Safa (2023) and El Shazly (2021) further stressed the advantages of IPAs, noting their ability to analyze users' abilities, needs, and interests and customize the learning accordingly. This approach fosters learner confidence and a stronger desire to speak the L2, ultimately leading to improved learning outcomes.

Notwithstanding these positive results, AI-powered tools do not always outperform traditional teacher-led pronunciation training. As stated by Dizon (2017), a study by Neri, Cucchiarini et al. (2008) with L2 Dutch learners found no significant pronunciation improvement in an experimental group that used CAPT tools compared to a control group that used traditional methods. The researchers suggested that more intensive training could have yielded better results.

A prevalent concern among EFL teachers is the potential threat AI poses to their job security (Dizon, 2017). However, despite these systems' sophisticated features, their effectiveness in improving learner outcomes remains under investigation. AI can be susceptible to errors, such as failing to detect pronunciation mistakes or misidentifying correct pronunciations (Rogerson-Revell, 2021). Another challenge with AI is that ASR technology can have difficulties understanding non-native or accented speech, leading to communication breakdowns and learner frustration during pronunciation practice (Dizon, 2017; Moussalli & Cardoso, 2019).

These limitations imply that, although AI-powered tools can be valuable additions to the EFL classroom and hold promising prospects for the future, a skilled language instructor remains essential for guiding students, addressing individual needs, and providing corrective feedback whenever needed while they interact with AI (Neri, Mich, et al., 2008; Rogerson-Revell, 2021). As Rogerson-Revell (2021) states when talking about AI resources, "the increasing attractiveness and availability of such tools does not of course ensure their pedagogic value or effectiveness" (p. 190). Thus, CAPT tools must be used carefully since while they may be fancy-looking and impress students and teachers at first, they could eventually fail to meet sound pedagogical requirements (Neri, Cucchiarini, et al., 2008).

1.3.2. AI-Powered Tools for Pronunciation Practice

The most straightforward way to bring AI-powered pronunciation training into the classroom is through the students' smartphones, which presents several benefits. Mobile language learning applications, commonly called apps, present an intuitive interface and are frequently available for free. They offer extended language exposure, empowering learners to practice conveniently at their own pace, anytime and anywhere, both inside and outside the EFL classroom (Rosell-Aguilar, 2017; Shortt et al., 2021).

A noteworthy feature of these apps is the inclusion of gamification elements, such as reward systems and progress indicators, making the learning process more entertaining, thereby increasing learner motivation. Some apps even provide audio-visual materials from native speakers, further supporting users in their language acquisition journey. Popular language learning apps like Duolingo, Busuu, Babbel, Rosetta Stone, or Memrise provide a wide range of activities for L2 learners, including vocabulary, reading and writing, grammar learning, listening, speaking, and translation. However, these apps often lack specific exercises geared toward pronunciation training (Aini et al., 2023; Rosell-Aguilar, 2017; Shortt et al., 2021).

Although they might not be as widely known, there are apps specifically designed for L2 pronunciation practice. Examples include Speak English, ELSA Speak, Speakometer, or Pronunciation Power (Aini et al., 2023; Fouz-González, 2020; Humardhiana, 2022; Rogerson-Revell, 2021; Rosell-Aguilar, 2017). These apps allow users to choose between British or American English pronunciation for practice sessions. By employing AI and ASR, they analyze users' speech patterns, provide feedback on pronunciation accuracy, and display phonetic transcriptions, acting as virtual 'pronunciation coaches' (Rogerson-Revell, 2021). One such app, ELSA Speak, was tested by Kholis (2021), who reported that it positively impacted students' pronunciation skills and motivation. Participants found the app convenient and effective, enjoying their experience using it.

Nonetheless, it is crucial to acknowledge that these AI-powered pronunciation apps are not without their limitations. The range of topics and conversations they offer may be somewhat restricted, and access to additional materials often requires purchasing the app's premium version (Humardhiana, 2022). Furthermore, like any tool that uses AI, the efficacy of these apps is still under investigation (Aini et al., 2023).

Another way to practice pronunciation using AI is through online dictionaries and machine translation tools (Rosell-Aguilar, 2017). Students can use speech-to-text features to input words or sentences into these systems, verifying whether their pronunciation is accurately recognized. Some dictionaries also offer phonetic transcription and the option to hear native speakers with different English accents pronounce the searcher words.

Lastly, as mentioned in the previous section, IPAs have significant pedagogical potential in SLA (Moussalli & Cardoso, 2019). Several studies (e.g., Dizon, 2017; Moussalli & Cardoso, 2016; 2019) have evaluated the capability of IPAs to understand L2 speech and their pedagogical potential in L2 education through learner-generated commands, questions, or interactive storytelling.

These AI systems can provide ample input and output opportunities, fostering meaningful human-machine interactions. In oral interactions between L2 students and IPAs, as noted by Molden (2015) and Moussalli and Cardoso (2019), if the AI system detects a mispronunciation, it can respond with comments, questions or even provide incorrect responses. This implicit form of feedback can be instrumental in raising learners' awareness of their mistakes and improving their pronunciation accuracy (Dizon, 2017).

In a study by Liakin et al. (2014) comparing two control groups of students learning French pronunciation with a human teacher and an experimental group using ASR software, they found that the use of AI had positive effects on the acquisition of a French vowel sound, as the experimental group outperformed the control ones, though not significantly. Moreover, the results of Liakin et al. (2014) pointed towards an optimistic direction regarding using mobile ASR in speech perception.

Molden (2015), however, points out that when speakers provide sufficient context, IPAs might automatically correct minor mispronunciations or grammatical errors. This, in turn, can lead L2 students to believe they produced the correct utterances even if they did not.

1.3.2.1. Siri: The Popular IPA

Apple's Siri is one of the most influential IPAs, revolutionizing user interaction with smart devices. Integrated into all Apple devices, Siri performs various tasks and retrieves information through voice commands. Launched in 2011, it remains one of the oldest and most widely used IPAs, continuously evolving and incorporating new functionalities (Alharthi, 2024; Hoy, 2018).

Siri's capabilities range from interpreting spoken queries to answering questions, scheduling appointments, making calls, sending text messages, setting alarms, controlling other apps, and accessing the Internet (Apple, n.d.; Hoy, 2018). Activated by the trigger phrase "Hey Siri," it employs deep neural networks, NLP, and other AI components to convert spoken input into text, enabling it to understand user commands and generate meaningful responses. Furthermore, Siri can access personal information stored on the device to provide tailored information and interpret speech within its linguistic context (Hoy, 2018; Moussalli & Cardoso, 2019).

Although Siri is not designed for SLA, it holds great potential in this area (Alharthi, 2024; Hirai & Kovalyova, 2023; Liakin et al., 2014). Moussalli and Cardoso (2019) contend that Siri's user-friendly interface can assist in pronunciation, offering students extensive exposure and numerous opportunities for output practice. Its real-time speech-to-text transcription allows language teachers to evaluate the accuracy of ASR technology in recognizing students' speech (Hirai & Kovalyova, 2023).

Alharthi (2024) explored the potential of using Siri to improve pronunciation through voice commands and claimed that Siri's context awareness and ability to correct minor grammatical errors automatically make it suitable for pronunciation practice. Given Siri's main goal of achieving intelligibility, this aligns perfectly with the objectives of pronunciation learning. Alharthi's (2024) research concluded that integrating Siri into pronunciation courses improved EFL students' academic performance.

However, Dizon (2021) cautions that a major limitation of using Siri and other IPAs is their irregular capacity to comprehend non-native speech. For example, Dizon's (2017) study showed that Alexa only fully understood 50% of the commands given by Japanese EFL learners. Similarly, Chen et al. (2020) found that Google Assistant struggled to comprehend the spoken utterances of EFL participants. Daniels and Iwago's (2017) research provided mixed results, with Google Assistant successfully transcribing 82% of the target words and sentences, while Siri had a lower mean accuracy rate of 66%.

1.3.2.2. Speakometer App: An AI-based Pronunciation Tool

Speakometer is an innovative mobile app powered by AI designed to enhance users' pronunciation skills in American or British English, with a particular emphasis on segmental practice (Speakometer, n.d.). According to a review by Kurt (2022), the app is built on three essential pillars for any CAPT tool: auditory and visualization features, ASR, and the provision of accurate feedback for learners.

Utilizing AI algorithms and ASR, Speakometer evaluates users' English pronunciation and offers feedback through written motivational messages accompanied by a visual representation called a "speakometer." The speakometer displays four colors representing different ratings, ranging from bad (red) to perfect (green) performance (Kurt, 2022). In addition, the Speakometer app provides phonetic transcription to further assist with pronunciation (Speakometer, n.d.).

After pronouncing a word, users can listen to their recording and compare it to a native speaker's pronunciation. This feature raises L2 students' awareness and serves as a self-feedback tool. Users can retry a word or proceed to the next one as desired. Each exercise set comprises ten practice words, and upon completion, Speakometer presents an overall score out of 100 points (Kurt, 2022).

Speakometer is available for both Apple and Android devices and can be downloaded for free, although certain in-app features may require payment. Compared to other pronunciation practice apps, Speakometer stands out for its extensive range of practice exercises available in the free version, including the *Top 200 English words, Voiced vs. Voiceless Consonants,* or *Phonetic Mastery.* However, the app's shortcomings include the absence of suprasegmental training and the lack of meaningful contextualization during word practice (Kurt, 2022).

1.4. Research Questions and Hypotheses

Drawing upon the background information gathered on AI, L2 pronunciation teaching and learning, and the integration of AI into the EFL classroom, this study aims to address the following research questions:

RQ1: Does the use of AI help EFL students improve their English pronunciation?

RQ2: How does the nature of stimuli, such as minimal pairs and sentences, influence AI comprehension?

RQ3: Does the use of different vowels impact the pronunciation accuracy of EFL students and, consequently, AI understanding?

RQ4: To what extent does the use of AI help the production of EFL students become more native-like?

The corresponding hypotheses for these research questions are as follows:

H1: The use of AI helps EFL students improve their English pronunciation due to its adaptability and ease of use.

H2: The nature of stimuli influences AI comprehension, with the system demonstrating a better understanding of sentences over minimal pairs.

H3: The pronunciation accuracy of EFL students varies depending on the vowels they produce, which, in turn, affects the AI understanding.

H4: The integration of AI into pronunciation instruction significantly contributes to the attainment of native-like production in EFL learners, particularly in improving their production of vowel sounds.

2. Method

2.1. Research Design

An experimental study was conducted with Spanish-Catalan bilingual high school students learning EFL to address the research questions. The study comprised a control group and an experimental group, employing a pretest-posttest design. During the interim period between the tests, both groups participated in four English pronunciation classes. Despite sharing identical instructional content, the groups engaged in distinct activities to practice pronunciation. The control group followed traditional methods, whereas the experimental group used the AI-driven pronunciation app Speakometer. The main objective was to determine whether students demonstrated improved learning outcomes with AI-based methods compared to traditional approaches. Additionally, the significance of integrating pronunciation instruction within the EFL classroom was assessed.

2.2. Participants

The sample consisted of 24 Spanish-Catalan teenagers (n = 24), aged between 14 and 15, studying the 3rd year of ESO (i.e., Educación Secundaria Obligatoria, or Secondary Education) at Escola Elisabeth, a semi-private school located in Salou, Tarragona, Spain. According to the CEFR levels, the students had a B1 English proficiency level.

For the experimental part of the study, participants were selected from two different classes to prevent cross-contamination of information. After that, the students were divided into a control group (n = 12) and an experimental group (n = 12). The gender distribution among the participants was evenly split, with 12 boys and 12 girls. The control group had more boys, while the experimental group was predominantly female. All participants were native Spanish-Catalan bilinguals born in Spain, and, on average, they began learning English between the ages of 4 and 5. However, out of the 24 students, only 7 received English lessons outside of regular school hours.

2.3. Materials

Firstly, students had to complete a consent form via Microsoft Forms on their laptops to participate in the study (see *Appendix A).* During the tests, a smartphone (an iPhone 15) was used to access Siri, while an iPhone 11 and a Sony PCM-M10 voice recorder served to record the students' voices. Printed sheets with minimal pairs and sentences were shown to the students (see *Appendix B*), and a separate sheet with the target vowel sounds was used to evaluate Siri's comprehension of each student's speech (refer to *Appendix C*). Additionally, some background information questions were included for the pretest (see *Appendix D).*

For the intervention phase, PowerPoint slides (see *Appendix E)* were created and presented using a projector and speakers provided by the school. Occasionally, a whiteboard and markers were used to illustrate words and vowel sounds. For the control group, cards containing different vowel sounds were prepared (see *Appendix F),* while the experimental group used their smartphones to access the pronunciation app Speakometer.

Lastly, a post-questionnaire was created using Microsoft Forms to know how the participants felt after the pronunciation classes (see *Appendices G & H*). It was administered online, and the students used their laptops to complete it. Afterwards, the data was analyzed using the resulting graph from Microsoft Forms

2.4. Procedure

After obtaining ethics approval[1] and the participants' consent, 24 students were randomly selected from different classes within the same high school level (3rd of ESO) and assigned to control and experimental groups. Individual pretests were conducted, with each student leaving their respective classrooms to meet the researcher in a separate room. The students were asked to provide some background information, and they were recorded while reading the prepared set of minimal pairs and sentences aloud to Siri, which was configured to English (United States). Simultaneously, Siri's interpretation of the students' utterances was noted using printed sheets with the target vowel sounds. Each recording session lasted approximately 2:30 minutes.

Following the pretests, the instruction phase began, comprising four sessions of 25-30 minutes each. Both groups received identical instruction on American English vowel sounds during these sessions through PowerPoint presentations. Each class focused on two vowel sounds known to be challenging for Spanish-Catalan speakers: /i/-/ɪ/, /u/-/ʊ/, /æ/-/ʌ/, and /ɝ/-/ɔr/. Active student participation was encouraged, minimizing teacher talk time to maximize student engagement in sound perception and production.

In the first class, students were introduced to a comparison of the Spanish, Catalan, and English vowel systems, along with their symbols from the International Phonetic Alphabet. They learned about vowel classification based on tongue position and the phonological parameters of height, frontness, and tenseness, exploring how they apply to the first pair of vowel sounds. MRIs of vowel sounds from the website Seeing Speech (Lawson et al., 2018) were shown to facilitate understanding of tongue positioning.

[1] To ensure the privacy of the participants, this study obtained the approval from the Ethics Committee for Research in People, Society, and the Environment (CEIPSA) under the reference **CEIPSA-2024-TFM-0009.**

Lastly, a table illustrating sound distribution and common spellings for each vowel sound was presented. This presentation structure was replicated in subsequent sessions, each with a brief review of key concepts from previous classes.

The second class introduced the phonological parameter of rounding and the vowels /u/ and /ʊ/. The third class focused on /æ/ and /ʌ/, using clips of YouTube videos featuring native speakers to aid in distinguishing these sounds (Learn Academic English, 2020; San Diego Voice and Accent, 2020). The final class, lasting an entire hour, introduced the r-colored vowel sound. Instead of MRIs, videos of UTI were utilized to demonstrate the sounds /ɝ/ and /ɔr/, followed by YouTube clips (Sounds American, 2017; 2018). Each class concluded with a practice session for the learned vowel sounds, which varied between the two groups.

The control group engaged in whole-class activities, using two cards for each vowel sound and words displayed on a screen. They were prompted to articulate the words and try to guess the corresponding vowel sounds, receiving feedback on the correct vowel sound and pronunciation. Conversely, the experimental group participated in pair activities using their smartphones. They had to download the Speakometer app, set their native language to either Spanish or Catalan and the practice accent to American English. They would always click on "Learn," then "Short and long vowels," and select the vowel sounds to practice, listening and recording themselves producing the words. At the end of the exercise, they shared their performance and the difficulties encountered.

Upon finishing the intervention, the students were re-recorded, reading the same minimal pairs and sentences to assess their progress from the pretest to the posttest. The study concluded with a brief post-questionnaire with ten statements using a Likert Scale to gauge the participants' general opinions and perceptions of the course and English pronunciation learning.

2.5.Data analysis

Upon gathering the necessary data, the analysis was divided into two main sections: speech perception and production. The perception analysis involved eight minimal pairs and eight sentences featuring the vowel sounds /i/, /ɪ/, /æ/, /ʌ/, /u/, /ʊ/, /ɝ/, and /ɔr/. Minimal pairs were scored on a scale from 0 to 2 based on Siri's comprehension (0 = No vowel sound was detected; 1 = One of the two vowels was understood; and 2 = Both vowels were understood). Each student could get a maximum score of 16.

As for sentences, each of them contained four vowel sounds, which were scored on a scale from 0 to 4, with a maximum score of 32. Scores were then converted to percentages for consistent weighting and comparison. Once this was done, the scores were calculated and regarded as indicators of speech perception results. The data was coded in Microsoft Excel and analyzed using linear mixed models in JASP to evaluate the impact of the intervention on students' speech accuracy from pretest to posttest[2]. The dependent variable was the percentage identification of the scores, and the fixed effects variables included group (control and experimental), time (pretest and posttest), and stimulus (minimal pairs and sentences).

Due to the large data volume, the production or acoustic analysis had to be narrowed to four vowels: /i/, /ɪ/, /æ/, and /ʌ/, instead of the original eight. Students' voice recordings were analyzed one by one using Praat software (Boersma & Weenink, 2023), with text grids added to indicate the word and its corresponding vowel sound (see Figure 5 for an example of what was done in Praat).

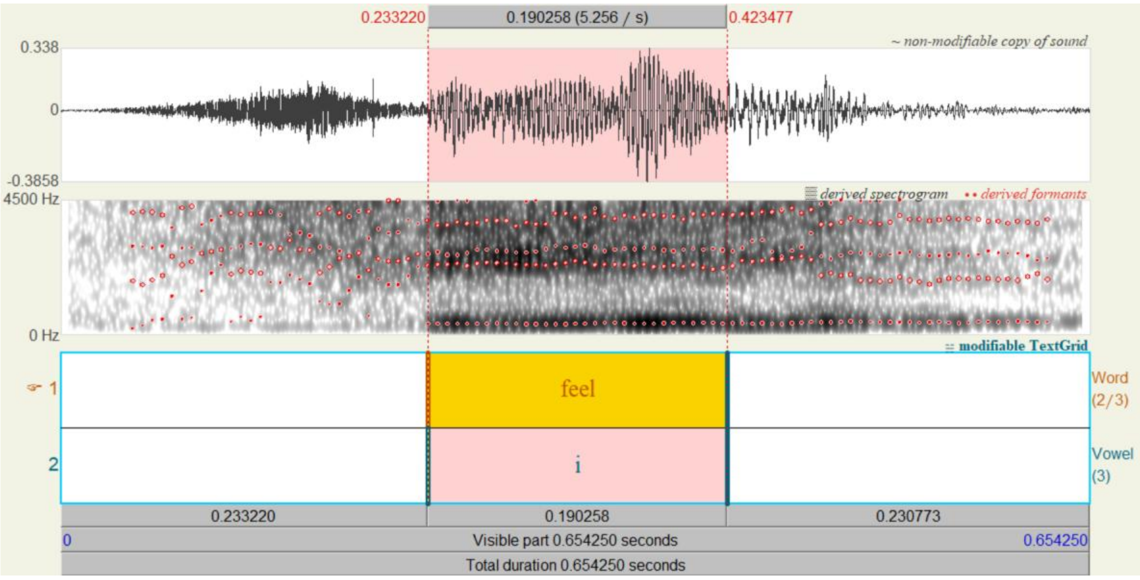
[2] Despite the small sample size of 24 participants, a Shapiro-Wilk test was deemed unnecessary as each participant produced 48 target vowel sounds per test.

Since words are random variables, only one word per vowel from the minimal pairs and sentences was selected for analysis. All target sounds consistently appeared in a consonant-vowel-consonant (CVC) position. Some students were excluded from either the minimal pairs or sentences in the pretest or posttest when they produced vowel sounds that deviated significantly from the target ones, or if their audio quality was insufficient for reliable analysis.

After the initial analysis in Praat, the reading of the F1 and F2 values of each vowel sound were obtained through a Praat script. The extracted formant frequencies were then organized into Excel files and analyzed in JASP using two different mixed models: one for F1 and another for F2.

To provide a benchmark for comparison, fictional "native speakers" representing the expected or standard formant frequencies of American English vowel sounds were introduced in both the control and experimental groups. Finally, scatterplots or vowel charts were generated in Excel to represent the production results in a visual way.

Figure 5. *An Example of Praat's Acoustic Analysis of Vowel Sounds*



3. Results

The study's results are divided into three sections. Section 3.1 presents the perception test results, indicating Siri's comprehension of students' speech during the pretests and posttests. Section 3.2 analyzes the students' production when reading minimal pairs and sentences and includes vowel charts to visually represent the results. Finally, Section 3.3 shows a graph with the post-questionnaire results. For every subsection, the data will be presented with a general ANOVA table, followed by their corresponding line graphs and, finally, the table with the descriptive statistics.

3.1. Perception Results

As can be observed from Table 3 below, the results were not significant for *group* ($F(1,22) = 2693$; $p = 0.115$), *time* ($F\ (1,22) = 3.631\times10^{-28}$; $p = 1.000$), or any of the interactions between the variables. The only effect that was highly significant in the perception test was *stimulus* ($F(1,22) = 625.429$; $p = < .001$). This is clearly illustrated in Figure 6, which plots the mean percentage identification for the two groups from the pretest to the posttest for the two types of stimuli.

Table 3. *ANOVA Summary for Perception*

Effect	df	F	p
Group	1, 22.00	2.693	0.115
Time	1, 22.00	3.631×10^{-28}	1.000
Stimulus	1, 22.00	625.429	< .001
Group ✻ Time	1, 22.00	0.459	0.505
Group ✻ Stimulus	1, 22.00	0.544	0.469
Time ✻ Stimulus	1, 22.00	0.835	0.371
Group ✻ Time ✻ Stimulus	1, 22.00	1.202	0.285

Figure 6. *Plot of Perception*

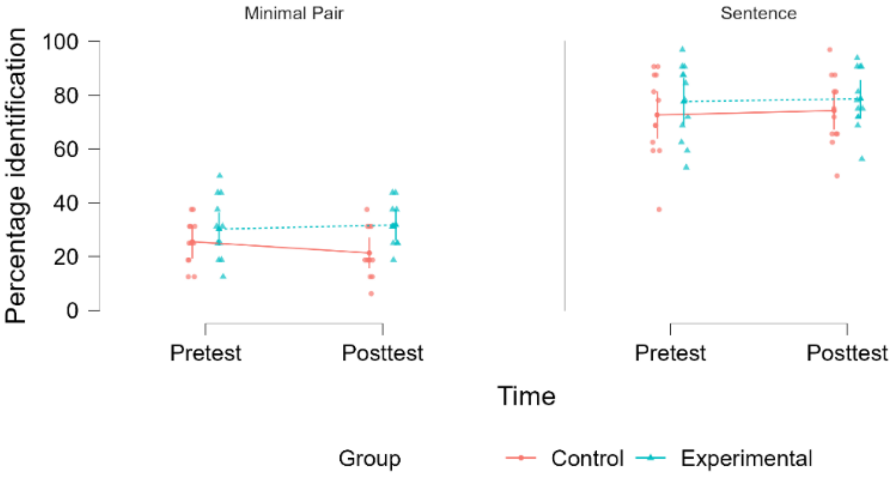


Table 4. *Descriptive Statistics for Perception*

	Percentage identification	
GROUP	**Control**	**Experimental**
Mean	48.438	54.557
Std. Deviation	27.958	26.264
STIMULUS	**Minimal Pair**	**Sentence**
Mean	27.214	75.781
Std. Deviation	10.027	13.622
TIME	**Pretest**	**Posttest**
Mean	51.497	51.497
Std. Deviation	27.098	27.501

Siri's understanding of the control and experimental groups was relatively stable, with minimal changes over time. Notably, Siri's perception showed a higher percentage of identification in sentences than in minimal pairs. The experimental group (*M*= 54.557, *SD* = 26.264) was slightly better understood by the IPA than the control group (*M* = 48.438, *SD* = 27.958), though these results were not significant. This pattern is also evident in the descriptive statistics in Table 4, where the largest significant difference is between the mean scores for the types of stimuli: minimal pairs (*M* = 27.214, *SD* = 10.027) and sentences (*M* = 75.781, *SD* = 13.622) (see Table 4).

3.2. Production Results

3.2.1. *Production of High Front Tense Vowel* /i/

The only effect that was significant in the F1 production of the high front tense vowel /i/ was *nativeness* ($F(1, 44.00) = 117.256$; $p < .001$), although *stimulus* ($F(1, 44.00) = 3.594$; $p = 0.065$) and the two-way interaction between *stimulus* and *nativeness* ($F(1, 44.00) = 3.594$; $p = 0.065$) approached the significance threshold (see Table 5).

Non-native groups had higher F1 values (control group: $M = 350.677$, $SD = 70.528$; experimental group: $M = 353.917$, $SD = 67.540$) than the standard values for native speakers ($M = 300$). There were no significant differences between the means of the control and experimental groups, and the data points were relatively close to the mean (see Table 7).

Table 5. *ANOVA Summary of F1 Production of High Front Tense Vowel*

Effect	df	F	p
Group	1, 44.00	0.112	0.739
Time	1, 44.00	1.410	0.241
Stimulus	1, 44.00	3.594	0.065
Nativeness	1, 44.00	117.256	< .001
Group ＊ Time	1, 44.00	0.886	0.352
Group ＊ Stimulus	1, 44.00	0.041	0.841
Time ＊ Stimulus	1, 44.00	0.092	0.763
Group ＊ Nativeness	1, 44.00	0.112	0.739
Time ＊ Nativeness	1, 44.00	1.410	0.241
Stimulus ＊ Nativeness	1, 44.00	3.594	0.065

Note. For the sake of brevity, the 3-way and 4-way interactions are not displayed, as none of them are significant. The same will be true for the other ANOVA Summary tables.

Figure 7. *Plot of F1 Production of High Front Tense Vowel*

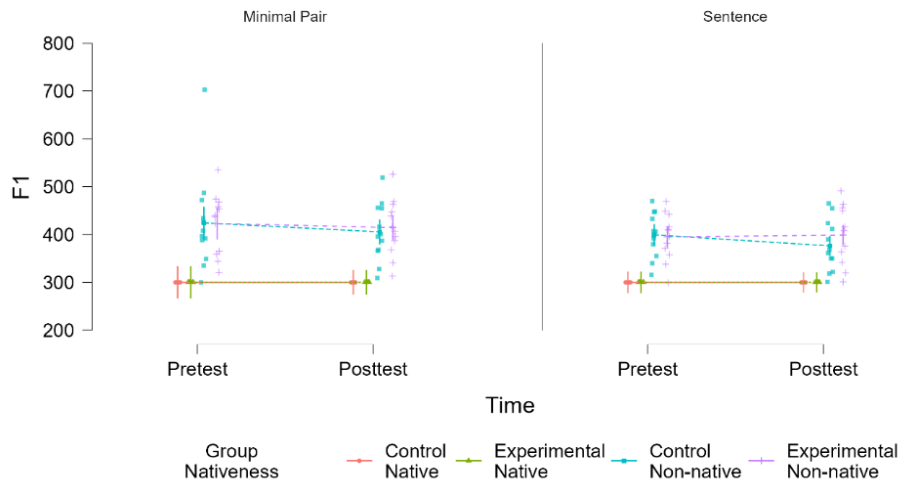


Table 6. *ANOVA Summary of F2 Production of High Front Tense Vowel*

Effect	df	F	p
Group	1, 88.00	2.080	0.153
Time	1, 44.00	0.104	0.749
Stimulus	1, 44.00	0.029	0.867
Nativeness	1, 88.00	0.973	0.327
Group ✻ Time	1, 44.00	0.055	0.815
Group ✻ Stimulus	1, 44.00	0.158	0.693
Time ✻ Stimulus	1, 88.00	1.148	0.287
Group ✻ Nativeness	1, 88.00	2.080	0.153
Time ✻ Nativeness	1, 44.00	0.104	0.749
Stimulus ✻ Nativeness	1, 44.00	0.029	0.867

Figure 8. *Plot of F2 Production of High Front Tense Vowel*

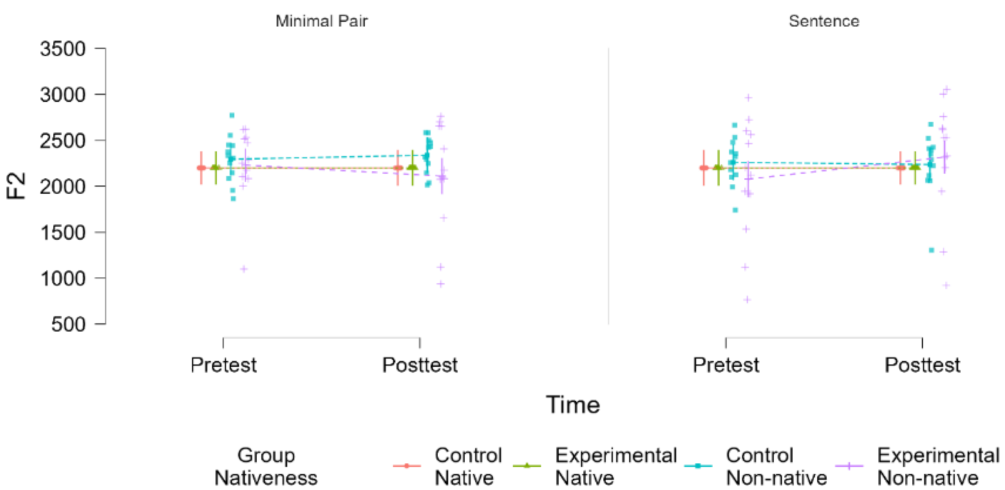


Table 7. *Descriptive Statistics for F1 and F2 Production of High Front Tense Vowel*

	F1		F2	
GROUP	**Control**	**Experimental**	**Control**	**Experimental**
Mean	350.677	353.917	2241.406	2192.229
Std. Deviation	70.528	67.540	189.308	408.834
STIMULUS	**Minimal Pair**	**Sentence**	**Minimal Pair**	**Sentence**
Mean	358.406	346.188	2221.854	2211.781
Std. Deviation	77.410	58.936	280.418	354.287
TIME	**Pretest**	**Posttest**	**Pretest**	**Posttest**
Mean	355.208	349.385	2208.094	2225.542
Std. Deviation	73.362	64.359	299.457	338.195
NATIVENESS	**Native**	**Non-native**	**Native**	**Non-native**
Mean	300.000	404.594	2200.000	2233.635
Std. Deviation	0.000	63.354	0.000	451.257

As shown in Table 6, no significant differences were found for the F2 production of /i/ (all effects had $p > .05$). Non-native students (M = 2233.635, SD = 451.257) produced F2 values close to the native ones (M = 2200), though with great standard deviation. Unlike F1 production, the F2 values of control (M = 2241.406, SD = 189.308) and experimental (M = 2192.229, SD = 408.834) non-native students showed substantial variability, which was particularly notable in the experimental non-native group across both the minimal pairs and sentences, as well as in the pretest and posttest (see Figure 8 and Table 7).

3.2.2. *Production of High Front Lax Vowel* /ɪ/

Like the F1 results for /i/, the only significant effect in the F1 production of the high front lax vowel /ɪ/ was *nativeness* (F (1, 45.63) = 17.210; $p < .001$). However, the two-way interaction between *group* and *time* (F (1, 49.85) = 3.497, p = 0.067) was close to being significant (see Table 8). Non-native students tended to produce higher F1 values than the standard native ones (non-native: M = 419.863, SD = 86.100; native: M = 375) (see Table

10). There seemed to be no significant difference from the pretest (*M* = 397.532, *SD* = 80.542) to the posttest (*M* = 397.333, *SD* = 44.622).

Table 8. *ANOVA Summary of F1 Production of High Front Lax Vowel*

Effect	df	F	p
Group	1, 45.63	0.489	0.488
Time	1, 49.85	0.007	0.932
Stimulus	1, 53.14	1.821	0.183
Nativeness	1, 45.63	17.210	< .001
Group ✻ Time	1, 49.85	3.497	0.067
Group ✻ Stimulus	1, 53.14	0.204	0.653
Time ✻ Stimulus	1, 133.55	0.446	0.505
Group ✻ Nativeness	1, 45.63	0.489	0.488
Time ✻ Nativeness	1, 49.85	0.007	0.932
Stimulus ✻ Nativeness	1, 53.14	1.821	0.183

Figure 9. *Plot of F1 Production of High Front Lax Vowel*

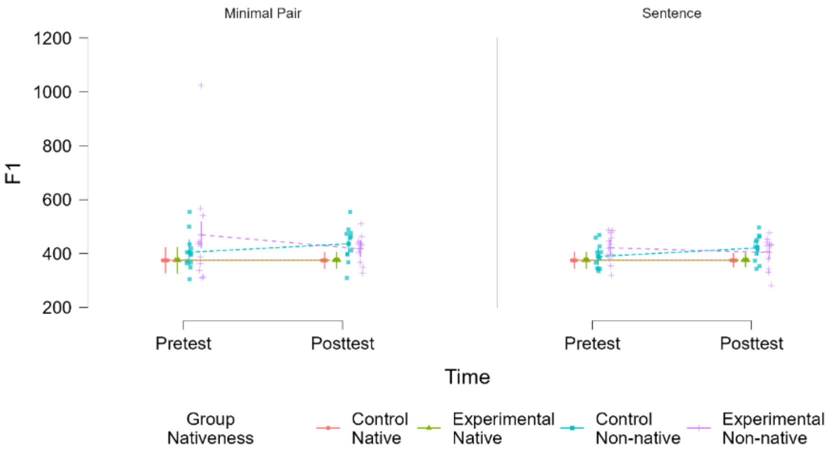


Table 9. *ANOVA Summary of F2 Production of High Front Lax Vowel*

Effect	df	F	p
Group	1, 44.51	0.201	0.656
Time	1, 83.27	0.081	0.776
Stimulus	1, 80.33	2.929	0.091
Nativeness	1, 44.51	2.631	0.112
Group ✻ Time	1, 83.27	0.094	0.760
Group ✻ Stimulus	1, 80.33	0.861	0.356
Time ✻ Stimulus	1, 130.50	0.097	0.756
Group ✻ Nativeness	1, 44.51	0.201	0.656
Time ✻ Nativeness	1, 83.27	0.081	0.776
Stimulus ✻ Nativeness	1, 80.33	2.929	0.091

Figure 10. *Plot of F2 Production of High Front Lax Vowel*

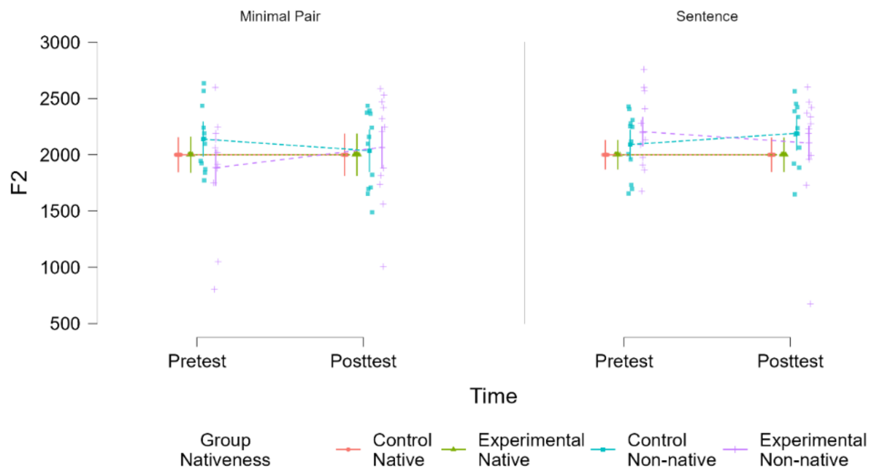


Table 10. *Descriptive Statistics for F1 and F2 Production of High Front Lax Vowel*

	F1		F2	
GROUP	**Control**	**Experimental**	**Control**	**Experimental**
Mean	393.885	401.053	2056.990	2033.394
Std. Deviation	44.944	80.195	213.009	328.158
STIMULUS	**Minimal Pair**	**Sentence**	**Minimal Pair**	**Sentence**
Mean	403.074	391.906	2016.223	2073.802
Std. Deviation	82.625	39.949	290.250	258.733
TIME	**Pretest**	**Posttest**	**Pretest**	**Posttest**
Mean	397.532	397.333	2041.096	2049.448
Std. Deviation	80.542	44.622	264.864	286.986
NATIVENESS	**Native**	**Non-native**	**Native**	**Non-native**
Mean	375.000	419.863	2000.000	2090.632
Std. Deviation	0.000	86.100	0.000	385.394

The results of F2 production of the high front lax vowel show that there were no significant differences in any of the effects nor in their interaction ($p > .05$). While the main effect of *stimulus* (F (1, 80.33) = 2.929; p = 0.091) approached significance, it did not reach the conventional threshold (see Table 9). The dispersion of data points for F2 production was considerably higher than in F1 production in every single effect, as can be seen in Figure 10 and Table 10. Interestingly, the F2 values of the control non-native group in the minimal pairs started high and appeared to move closer to the standard values over time (Figure 10). Conversely, the opposite trend emerged in sentences, where they started lower and then produced higher values than the standard ones in the posttest. As for the experimental non-native group, they did the opposite: they started low in the minimal pairs pretest but increased over time. However, they began producing high values in the sentences and remained high in the posttest, although slightly lower than the expected F2 values.

3.2.3. *Production of Low Front Vowel* /æ/

The analysis of the low front vowel /æ/ revealed no statistically significant effects of *group*, *time*, *stimulus*, *nativeness*, or any of the interactions for F1 production. This is evident from the p-values in Table 11, all above the threshold ($p > .05$). As seen in Figure 11, both non-native groups showed relatively stable F1 production across the testing period. Yet, it is worth noting that some non-native students from both groups deviated from the expected F1 value, especially in the minimal pairs, as indicated by the scattered data points on the graph and the mean and standard deviation details for minimal pairs (M = 754.115, SD = 106.713) (see Table 13). Despite this, the absence of significant effects suggests that students did not significantly differ from native speakers in their F1 production.

Table 11. *ANOVA Summary Results for F1 Production of Low Front Vowel*

Effect	df	F	p
Group	1, 44.00	0.184	0.670
Time	1, 44.00	2.829	0.100
Stimulus	1, 44.00	2.027	0.162
Nativeness	1, 44.00	1.474	0.231
Group ✻ Time	1, 44.00	0.771	0.385
Group ✻ Stimulus	1, 44.00	0.257	0.615
Time ✻ Stimulus	1, 44.00	2.635	0.112
Group ✻ Nativeness	1, 44.00	0.184	0.670
Time ✻ Nativeness	1, 44.00	2.829	0.100
Stimulus ✻ Nativeness	1, 44.00	2.027	0.162

Figure 11. *Plot of F1 Production of Low Front Vowel*

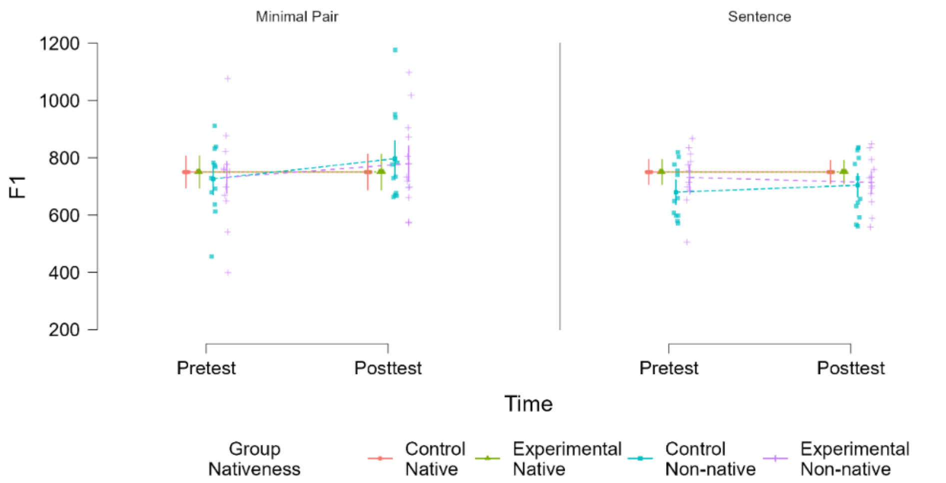


Table 12. *ANOVA Summary of F2 Production of Low Front Vowel*

Effect	df	F	p
Group	1, 44.01	0.348	0.558
Time	1, 45.32	2.241	0.141
Stimulus	1, 50.88	0.526	0.472
Nativeness	1, 44.01	111.656	< .001
Group ✻ Time	1, 45.32	0.007	0.932
Group ✻ Stimulus	1, 50.88	1.066	0.307
Time ✻ Stimulus	1, 88.00	0.250	0.618
Group ✻ Nativeness	1, 44.01	0.348	0.558
Time ✻ Nativeness	1, 45.32	2.241	0.141
Stimulus ✻ Nativeness	1, 50.88	0.526	0.472

Figure 12. *Plot of F2 Production of Low Front Vowel*

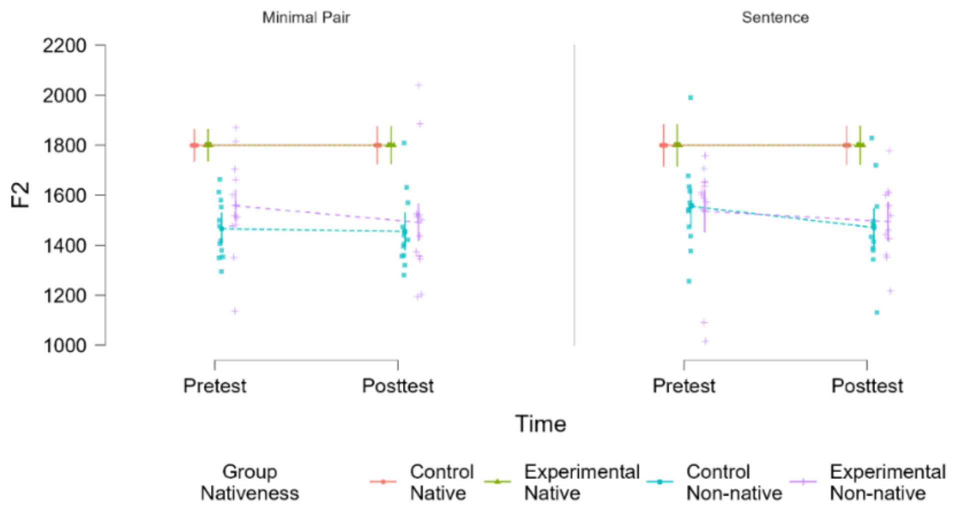


Table 13. *Descriptive Statistics for F1 and F2 of Low Front Vowel*

	F1		F2	
GROUP	**Control**	**Experimental**	**Control**	**Experimental**
Mean	738.260	744.385	1643.594	1660.125
Std. Deviation	89.003	93.749	192.713	202.131
STIMULUS	**Minimal Pair**	**Sentence**	**Minimal Pair**	**Sentence**
Mean	754.115	728.531	1646.406	1657.313
Std. Deviation	106.713	70.789	201.487	193.587
TIME	**Pretest**	**Posttest**	**Pretest**	**Posttest**
Mean	733.500	749.146	1664.531	1639.188
Std. Deviation	87.197	94.882	188.309	205.787
NATIVENESS	**Native**	**Non-native**	**Native**	**Non-native**
Mean	750.000	732.646	1800.000	1503.719
Std. Deviation	0.000	128.753	0.000	183.793

Regarding F2 production, the only significant effect was *nativeness* ($F(1, 44.01) = 111.656$, $p < .001$), indicating that non-native students deviated significantly from the native-like F2 production (see Table 12). This is further illustrated in Figure 12 and Table 13, where the expected F2 ($M = 1800$) contrasts sharply with the Spanish-Catalan students' production ($M = 1503.719$, $SD = 183.793$).

There was considerable variation within the non-native groups ($M = 1503.719$, $SD = 183.793$), with some students in the experimental group ($M = 1660.125$, $SD = 202.131$) producing F2 values as low as 1000 Hz, while others exceeded the expected value. Both control and experimental non-native students tended to produce low F2 values.

3.2.4. *Production of Low Central Vowel /ʌ/*

In the F1 production of the low central vowel /ʌ/, none of the variables were statistically significant since they all had $p > 0.05$ (see Table 14). As observed in Figure 13 and further detailed in Table 16, there was a high degree of data dispersion from both non-native groups ($M = 676.184$, $SD = 121.934$), indicating a high standard deviation, particularly in the minimal pairs (*M* = 681.902, *SD* = 102.633) during the pretest (*M* = 681.444, *SD* = 101.174), showing a wide range of F1 values, from as low as 300Hz to as high as 900Hz. Despite the lack of significant effects, there was a clear trend of increasing F1 values from the pretest to the posttest among non-native speakers. However, these values either fell short or exceeded the standard F1 value.

Table 14. *ANOVA Summary of F1 Production of Low Central Vowel*

Effect	df	F	p
Time	1, 44.52	2.928	0.094
Group	1, 40.76	0.083	0.775
Stimulus	1, 43.25	1.338	0.254
Nativeness	1, 40.76	3.078	0.087
Time ＊ Group	1, 44.52	0.069	0.794
Time ＊ Stimulus	1, 42.45	0.916	0.344
Group ＊ Stimulus	1, 43.25	1.544	0.221
Time ＊ Nativeness	1, 44.52	2.928	0.094
Group ＊ Nativeness	1, 40.76	0.083	0.775
Stimulus ＊ Nativeness	1, 43.25	1.338	0.254

Figure 13. *Plot of F1 Production of Low Central Vowel*

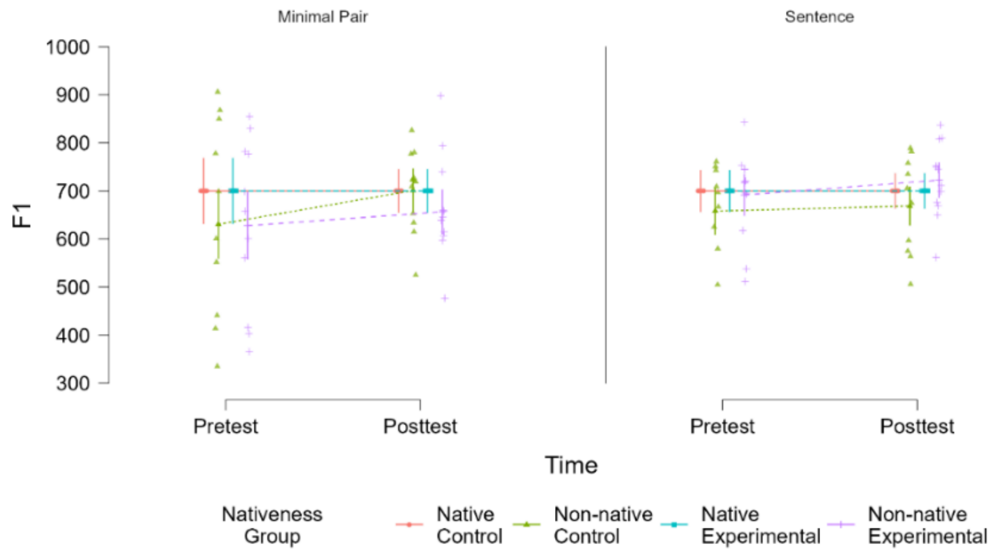


Table 15. *ANOVA Summary of F2 Production of Low Central Vowel*

Effect	df	F	p
Group	1, 36.56	2.122	0.154
Time	1, 107.32	0.116	0.735
Stimulus	1, 42.41	1.904	0.175
Nativeness	1, 36.56	24.588	< .001
Group ✻ Time	1, 107.32	0.278	0.599
Group ✻ Stimulus	1, 42.41	3.462	0.070
Time ✻ Stimulus	1, 123.48	0.283	0.596
Group ✻ Nativeness	1, 36.56	2.122	0.154
Time ✻ Nativeness	1, 107.32	0.116	0.735
Stimulus ✻ Nativeness	1, 42.41	1.904	0.175

Figure 14. *Plot of F2 Production of Low Central Vowel*

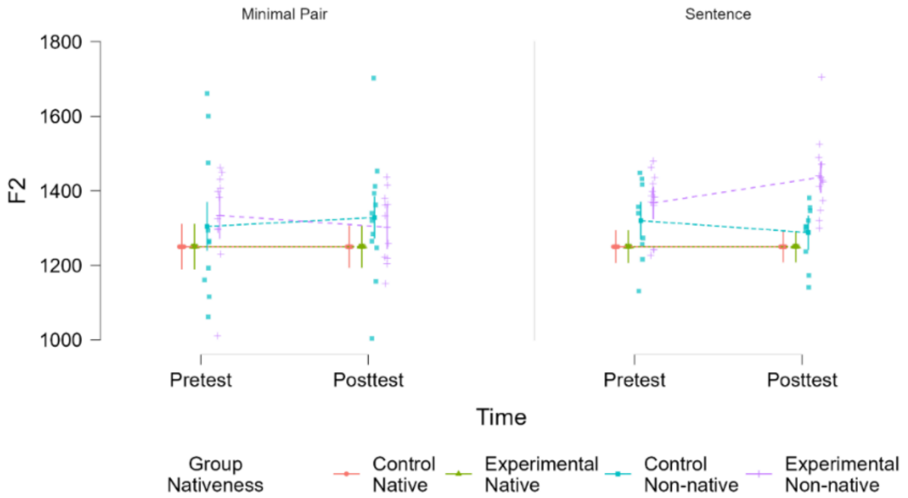


Table 16. *Descriptive Statistics for F1 and F2 Production of Low Central Vowel*

	F1		F2	
GROUP	**Control**	**Experimental**	**Control**	**Experimental**
Mean	687.416	689.872	1278.124	1305.096
Std. Deviation	85.513	84.289	102.825	97.334
STIMULUS	**Minimal Pair**	**Sentence**	**Minimal Pair**	**Sentence**
Mean	681.902	695.527	1283.293	1300.758
Std. Deviation	102.633	61.234	108.398	91.987
TIME	**Pretest**	**Posttest**	**Pretest**	**Posttest**
Mean	681.444	695.677	1289.500	1294.376
Std. Deviation	101.174	64.625	100.534	101.303
NATIVENESS	**Native**	**Non-native**	**Native**	**Non-native**
Mean	700.000	676.184	1250.000	1338.299
Std. Deviation	0.000	121.934	0.000	131.579

In contrast, the analysis of F2 production revealed a significant effect of *nativeness* (F (1, 36.56) = 24.588; $p < .001$), and the interaction between *group* and *stimulus* approached statistical significance (F(1,42.41) = 3.462; p = 0.070) (see Table 15). Comparing the standard F2 production of native speakers (M = 1250) with that of non-native students (M = 1338.299, SD = 131.579), it becomes evident that the F2 values of Spanish-Catalan students were much higher than expected, with considerable individual variation.

This trend is also reflected in the dispersed data points (see Figure 14) and the high standard deviations in both minimal pairs (M = 1283.293, SD = 108.398) and sentences (M = 1300.758, SD = 91.987). During the production of sentences in the posttest, the experimental non-native group produced notably high F2 values, whereas the control non-native group demonstrated F2 values that closely approximated those of native speakers.

3.2.4.1. Vowel Charts

Vowel charts were created using Excel to represent the vowel sounds produced by the students more intuitively. Figures 15 to 18 display the same data from the production results but by plotting the formant frequencies in the vowel space, alongside the standard values (SV) for each vowel sound. This latter component allows for a comparison between the results from both the control and experimental groups with the American English values they should have obtained.

It was observed that both non-native groups deviated from native pronunciation in either the F1 or F2 values, and the data points showed a high degree of dispersion. There were no significant differences between the pretest and posttest, the groups, or the stimuli used. However, for the four vowel sounds, the main effect of *nativeness* was found to be significant. For the vowel sounds /i/ and /ɪ/, this happened in the F1 production, where the non-native groups produced significantly higher values than expected. Conversely, for /æ/ and /ʌ/, the most substantial differences were in the F2 values, indicating a significant deviation from the expected frontness.

Figure 15. *Vowel Chart: Minimal Pairs Pretest*

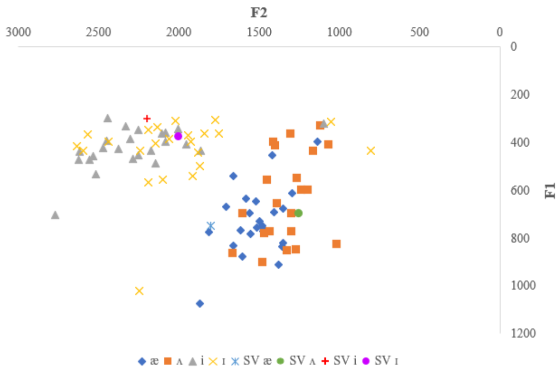


Figure 16. *Vowel Chart: Minimal Pairs Posttest*

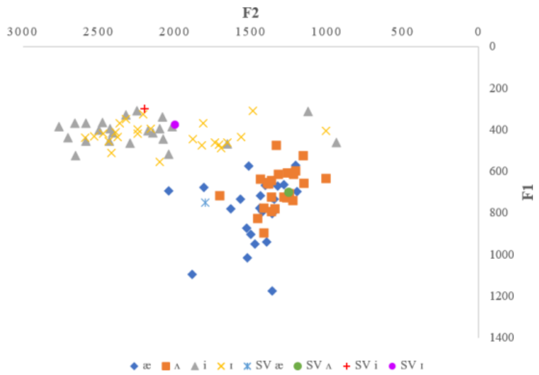


Figure 17. *Vowel Chart: Sentences Pretest*

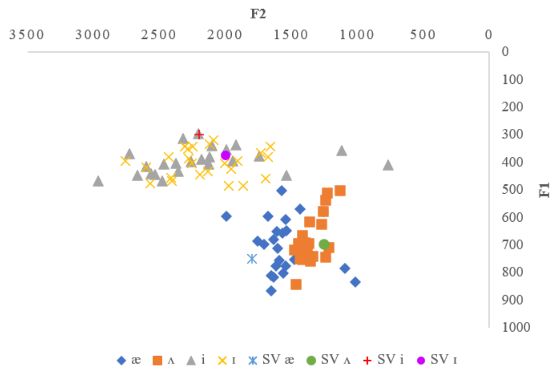
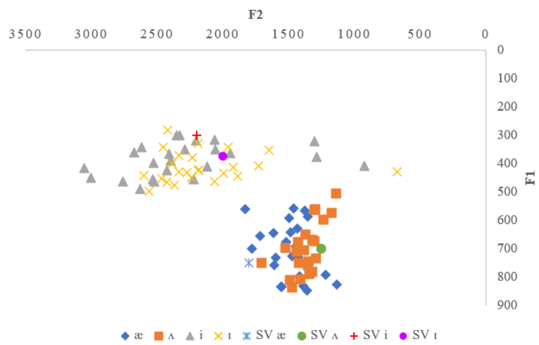


Figure 18. *Vowel Chart: Sentences Posttest*

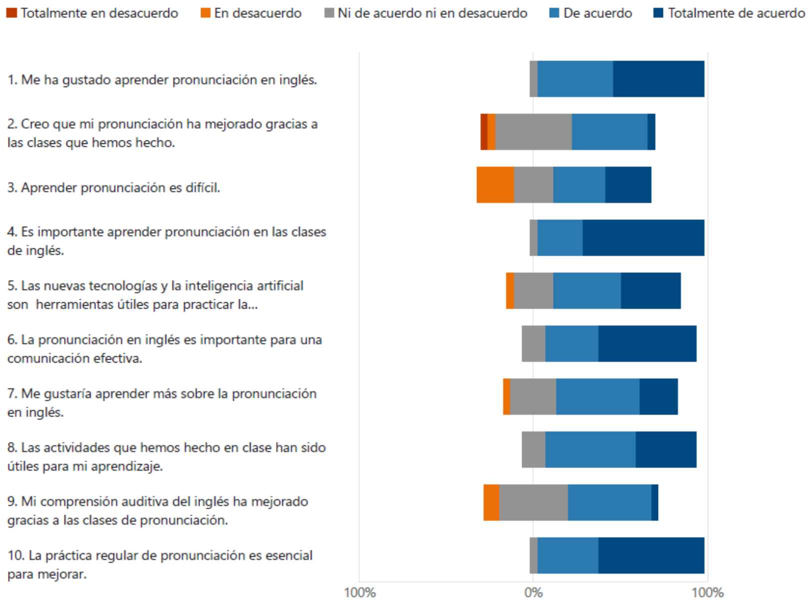


3.3.Post-Questionnaire Results

A 5-point Likert scale was used to measure the general students' opinions and satisfaction with English pronunciation learning. The scale ranged from 1 (Strongly disagree) to 5 (Strongly agree) and included 10 items covering various aspects of the sessions they received. The results of the post-questionnaire, which were analyzed using Microsoft

Forms, are presented in Figure 19. The graph displays the responses to ten statements presented to 23 of the students, displaying the percentages for each answer. One participant had to be removed from the results because they did not follow the instructions correctly, thus becoming an outlier that could potentially alter the results.

Figure 19. *Post-Questionnaire Results*



Note. Graph retrieved from Microsoft Forms. For a translation of the post-questionnaire, refer to *Appendix H.*

When asked about their enjoyment of learning English pronunciation, half of the students (52.2%) strongly agreed, while the remaining students either agreed (43.5%) or remained neutral (4.3%). As for perceiving improvements in their pronunciation, responses varied, with most students either agreeing (43.5%) or expressing neutrality (also 43.5%) regarding their progress.

The participants' opinions on the difficulty of learning pronunciation displayed diverse responses. Some students disagreed (21.7%), while an equal number adopted a neutral standpoint (21.7%). Conversely, 30.4% agreed, and 26.1% strongly agreed that learning pronunciation was challenging. Furthermore, a significant majority (69.6%) strongly agreed that learning pronunciation in English classes is important.

The use of new technologies and AI as tools for practicing English pronunciation garnered positive feedback. A considerable number of students agreed (39.1%), with an additional 34.8% strongly agreeing, while a minority maintained a neutral position (21.7%). Only a small fraction of participants (4.3%) expressed disagreement with this statement. Additionally, 56.5% of students believed that a firm grasp of English pronunciation is indispensable for effective communication, with 47.8% expressing a desire to delve further into the intricacies of English pronunciation.

Statement number eight evoked a predominantly positive response, with most students agreeing (52.2%) or strongly agreeing (34.8%) that the pronunciation activities done in class benefit their language learning. Concerning improving students' perception and auditory comprehension of English following the pronunciation sessions, nearly half of the participants (47.8%) agreed, while 39.1% held a neutral position, 8.7% disagreed, and merely 4.3% strongly agreed.

4. Discussion

The main purpose of the present study was to investigate the potential impact of introducing AI into the language classroom to improve the pronunciation skills of teenage Spanish-Catalan EFL learners. Additionally, the learners' opinions on pronunciation learning within SLA were gathered. The discussion will be structured into three main parts. Firstly, a thorough analysis and interpretation of the major findings from the results section will be provided. Secondly, attention will be given to addressing the research questions and their corresponding hypotheses. Finally, the study's limitations will be listed, and suggestions for further research in this domain will be put forth.

4.1. Major Findings

4.1.1. Siri's Perception

The perception test results revealed that Siri's understanding of the eight vowel sounds varied drastically depending on the type of stimulus, which was the only significant effect ($p < .001$). The IPA would comprehend sentences ($M = 75.781$) considerably better than minimal pairs ($M = 27.214$). Regarding group understanding, Siri demonstrated a slightly better understanding of the students from the experimental group ($M = 54.557$) compared to the control group ($M = 48.438$). Even so, the difference between the two groups was minimal and not statistically significant ($p = 0.115$), and it cannot be attributed to the use of the Speakometer app. Furthermore, the intervention did not affect the IPA's understanding of the EFL learners' speech, as evidenced by the equal means obtained in the pretest ($M = 51.497$) and posttest ($M = 51.497$).

4.1.2. Students' Production

Based on the production results, there were no discernable changes in any of the four vowel sounds between the pretest and posttest, as the effect of *time* was never statistically significant. The most notable difference was observed between high and low vowels. For the high front tense and lax vowels /i/ and /ɪ/, there was a significant difference in F1 production for the main effect of *nativeness* ($p < .001$), with no significant difference in F2 production. Conversely, the F1 production results for the low front and low central vowels /æ/ and /ʌ/ revealed no significant differences in any of the main effects or their interactions, though the effect of *nativeness* was significant in F2 production ($p < .001$). This means that the students were producing the right height for high vowels but the wrong frontness, and the other way around for low vowels.

4.1.2.1. High Front Tense Vowel Production Discussion

In the production analysis of the high front tense vowel /i/, the mean F1 values of both groups seemed to indicate they were producing a slightly lower vowel sound than the native value (control: $M = 350.677$; experimental: $M = 353.917$; native: $M = 300$). This suggests that the participants were producing F1 values closer to the standard Catalan values proposed by Recasens (2014) (F1 = 334) than to the American English ones proposed by Romero (n.d.-a) (F1 = 300).

Although there were no significant differences in frontness (F2) based on main effects or their interactions, there was a notable standard deviation, particularly in the experimental group ($M = 2192.229$, $SD = 408.834$). Similarly, the effect of *nativeness* was not significant due to the non-native pronunciation ($M = 2233.635$, $SD = 451.257$) being close to native-like ($M = 2200$), though with a significant standard deviation.

4.1.2.2. High Front Lax Vowel Production Discussion

The F1 values for the high front lax vowel /ɪ/ produced by non-native students exceeded those of native speakers (control: $M = 393.885$; experimental: $M = 401.053$; native: $M = 375$). The formant frequencies produced by Spanish-Catalan students were closer to the Spanish /e/ (F1 = 457) and the Catalan /e/ (F1 = 450) sounds rather than the English /ɪ/ (F1 = 375). F2 production showed no significant differences, but the effect of *stimulus* ($p = 0.091$) was close to reaching significance. Interestingly, the control group initially produced high F2 values in minimal pairs, gradually approaching native English values over time. However, they consistently maintained high F2 values in sentence production. On the other hand, the experimental group started with lower-than-expected F2 values in minimal pairs, which increased in the posttest. And in sentence production, they began with high F2 values and produced values closer to the native ones in the posttest.

4.1.2.3. Low Front Vowel Production Discussion

Examining the low front vowel /æ/ results, both groups demonstrated stability in their F1 production throughout the testing period. Nonetheless, it is worth noting that there were several scattered data points, particularly in the minimal pairs (M = 754.115, SD = 106.713) and in the non-native production (M = 732.646, SD = 128.753). The non-native mean suggests the students produced a vowel sound closer to the Catalan /a/ (F1 = 730).

Regarding F2 production, which yielded a significant difference in the main effect of *nativeness* ($p < .001$), the expected native F2 values for the English /æ/ sound (M = 1800) sharply contrasted with what the Spanish-Catalan students produced (M = 1503.719). The control and experimental groups produced lower F2 values than expected, indicating a sound further back in the mouth resembling the low central vowel sound /a/ from Catalan (F2 = 1358) and Spanish (F2 = 1471). Alternatively, it could be seen as an intermediate sound between the Catalan sounds /ɛ/ (F2 = 1700) and /a/ (Rallo Fabra & Romero, 2012).

4.1.2.4. Low Central Vowel Production Discussion

The production results for the low central vowel were comparable in height and frontness to those for the low front vowel. No significant differences were observed in F1 production for any effects. However, the non-native groups exhibited considerable data dispersion, especially in the minimal pairs pretest. A highlight was that the control group approximated native-like F1 production for minimal pairs in the posttest. Both non-native groups (M = 676.184) increased their F1 values over time, moving closer to the native F1 value (M = 700).

In contrast, there was a significant difference in F2 production of /ʌ/ due to the effect of *nativeness* ($p < .001$), just like in the vowel sound /æ/. Additionally, the interaction between *group* and *stimulus* approached significance ($p = 0.070$). In sentence production, the control group moved closer to native-like F2 production in the posttest, whereas the experimental group produced much higher values. This discrepancy may be attributed to the lack of an equivalent sound in Catalan or Spanish, as Rallo Fabra and Romero (2012) suggested. Instead of achieving the native F2 value ($M = 1250$), the non-native groups produced a higher F2 ($M = 1338.299$), resembling the F2 values of the Catalan /a/ sound (F2 = 1358).

4.1.3. Additional Findings: Post-Questionnaire

While not directly tied to the initial research questions, a post-questionnaire was administered to gauge participants' perspectives on EFL pronunciation teaching and learning. The answers indicated that they derived a sense of enjoyment from learning pronunciation within the EFL classroom environment, as indicated by 52.2% strongly agreeing and 43.5% agreeing. However, divergent opinions emerged concerning the impact of the classes on their English pronunciation, with 43.4% agreeing, 43.5% adopting a neutral stance, and 4.3% each strongly agreeing, disagreeing, and strongly disagreeing. Regarding the classroom activities, a significant proportion of participants (52.2% agreed and 34.8% strongly agreed) felt that activities using traditional methods or the Speakometer app benefited their learning experience. Additionally, 47.8% agreed that their ability to discern vowel sounds had improved thanks to the intervention.

Despite lacking prior English pronunciation instruction, a notable percentage (30.4% and 26.1%) found learning pronunciation not to be a difficult task, while 21.7% disagreed. Most participants (69.6%) acknowledged the importance of pronunciation learning in SLA, with half expressing a desire to learn more about English pronunciation. Moreover, 60.9% strongly agreed that regular practice enhances pronunciation skills. Regarding AI tools, 9.1% agreed, and 34.8% strongly agreed they could facilitate pronunciation practice and improvement. However, it is crucial to remember that only one group underwent practice with AI, limiting the other group's ability to assess the effectiveness of this tool.

Students' positive attitudes towards pronunciation instruction and their recognition of its importance in EFL learning indicate a strong desire to enhance their oral communicative abilities in the L2. As Szpyra-Kozłowska (2014) remarked, acquiring pronunciation skills can greatly improve students' proficiency and confidence in speaking the L2. To achieve this, EFL teachers need to receive training in teaching pronunciation. Once equipped with this knowledge, EFL classrooms can adopt the framework proposed by Celce-Murcia et al. (2010), focusing on achieving intelligibility and comprehensibility while also raising awareness of the consequences that phonetic errors can have (Alharthi, 2024; Derwing & Munro, 2005; Szpyra-Kozłowska, 2014).

4.2. Addressing the Research Questions

4.2.1. Research Question 1

The first research question aimed to evaluate the efficacy of AI in improving EFL learners' pronunciation. The results from the perception analysis, which indicated no significant changes in the experimental group's test scores over time following pronunciation training with the Speakometer app, combined with the production analysis, which found no meaningful differences between the control and experimental groups, lead to the rejection of the initial hypothesis statement.

While it is indeed true that the students in the experimental group manifested greater motivation and interest in learning with the CAPT tool, corroborating the assertions of authors like Aini et al. (2023), Kholis (2021), Rosell-Aguilar (2017), and Shortt et al. (2021), the absence of significant differences between groups suggests that the incorporation of AI does not always necessarily yield discernible improvements in their English pronunciation skills. Therefore, the convenience, user-friendly nature, and attractiveness of AI-powered mobile apps should not be considered indicative of their effectiveness in L2 pronunciation training.

4.2.2. Research Question 2

Regarding how the nature of stimuli influences AI comprehension, which was the concern of the second research question, the perception results revealed that Siri’s recognition of speech varied significantly based on the type of stimulus provided by EFL learners. Siri’s ASR exhibited significantly higher comprehension rates for sentences (M = 75.781) than for minimal pairs (M = 27.214). This discrepancy can be attributed to Alharthi's (2024) and Molden's (2015) observations that AI-powered systems rely on contextual cues to understand speech. With enough context, these systems can deduce the speaker's intent, even when mispronunciations arise. However, when students worked with minimal pairs, the lack of contextual clues troubled Siri's understanding.

During the tests, Siri hardly ever understood both words in the minimal pairs, and even when it did, it would replace them with commands that made more sense within its designed functionality. For example, with the minimal pair *pull-pool*, Siri would interpret the input as *Call Paula*. Alternatively, depending on the students’ pronunciation and Siri’s comprehension, it would create calendar events, play songs, or open specific apps. This last example happened with the minimal pair *feel-fill*, where Siri usually opened the Health app, inferring the user wanted to report how they were feeling. In extreme cases, the IPA would stop functioning and close itself if it could not perceive any clear command due to low voice, background noise, or significant mispronunciations.

These observations highlight Siri's limited ability to understand non-native speakers, as Dizon (2021) noted. They also underscore the need for further improvements to ensure Siri's suitability in SLA contexts to assess non-native speakers' pronunciation accurately.

4.2.3. Research Question 3

As for the third research question, which explored whether the use of different vowels affected the pronunciation accuracy of EFL students and, consequently, AI understanding, the findings suggest that such an impact does exist. The production results showed that students' F1 and F2 values did not always resemble those of native American English speakers. The limited vowel inventory of Spanish and Catalan likely exacerbates difficulties in accurately pronouncing English vowels, as learners tend to assimilate new sounds with those they perceive as articulatorily similar (Best & Tyler, 2007). As a result, these mispronunciations can pose challenges for AI systems in understanding the spoken language of non-native speakers.

The participants' production of the tense-lax vowel pair /i/-/ɪ/ aligns with the observations of Rallo Fabra and Romero (2012) on the cross-linguistic similarity of English vowels to Catalan vowels and with the PAM-L2 (Cebrian, 2006; Rallo Fabra & Tyler, 2021; Yavaş, 2011). Since neither Spanish nor Catalan feature the property of tenseness, participants from both groups assimilated the high front lax vowel sound from English to a sound from their L1 vowel inventory (i.e., Catalan /e/). A similar pattern was observed with the English vowel sounds /æ/ and /ʌ/. The former sound is intermediate between the Catalan /ɛ/ and /a/ sounds, while the latter lacks a Catalan equivalent, leading non-native speakers to substitute it with the central Catalan vowel /a/, which they perceived as the closest match (Rallo Fabra & Romero, 2012).

Moreover, some students mispronounced the word *since* using the /aɪ/ sound and articulated the vowels from *least, much,* and *cup* as if they were reading them in Spanish or Catalan. This phenomenon was probably due to heightened self-awareness during the recordings, leading to nervousness and overcorrection to avoid errors. Additionally, the differences in orthographic transparency between English and Catalan or Spanish may also play a significant role. As Rallo Fabra (2022) explains, these differences can pose substantial challenges for EFL learners when pronouncing English words.

4.2.4. Research Question 4

The fourth research question aimed to evaluate the extent to which AI aids EFL students in achieving more native-like pronunciation. The production data analysis revealed no statistically significant differences between the control and experimental groups across the four vowel sounds from the pretest to the posttest. Although the integration of AI through the Speakometer app increased students' motivation and willingness to participate in the activities, it did not effectively assist Spanish-Catalan learners in achieving a pronunciation comparable to that of native speakers.

This finding aligns with the conclusions drawn by Neri, Cucchiarini, et al. (2008), who found that L2 Dutch learners did not experience substantial improvements in pronunciation when utilizing CAPT tools, in contrast to the control group that employed conventional instructional methods. Conversely, it sharply contrasts with the favorable outcomes reported in the study by Liakin et al. (2014).

The precision of Speakometer in providing pronunciation instruction and feedback did not meet expectations. Several errors in the app's phonetic transcriptions were identified while monitoring student usage (see *Appendix I).* Notably, Speakometer misplaces stress marks before vowel sounds and confuses British and American phonetic transcriptions by incorporating length marks and distinguishing between short and long vowels, which, according to Kenyon (1997), is not a distinctive feature of American English. Instances of the app's errors include the word *tour,* transcribed as /tˈɔː/ instead of /'tʊr/, and identical phonetic transcriptions for *saw* and *sore* as /sˈɔː/, despite the correct American English pronunciations being /ˈsɔ/ and /ˈsɔr/, respectively. Such inaccuracies can confuse learners and undermine their progress in EFL pronunciation learning.

Additionally, Speakometer often failed to offer clear feedback or explanations on how to pronounce vowel sounds properly, merely marking errors in red, telling students that they "could do better" or should "try again." This limitation further underscores the value of human teachers in guiding students and providing corrective feedback. Even if AI-powered tools are utilized in the classroom, skilled language instructors must supervise students' progress and assist them whenever needed, aligning with the views of Neri, Mich, et al. (2008) and Rogerson-Revell (2021).

4.3. Limitations of the Study and Suggestions for Further Research

As with the majority of studies, the design of the current study faced some constraints. Firstly, the number of teaching sessions was limited, with each session being of short duration. This may have prevented the students from making meaningful progress because it limited their exposure to and practice with American English vowel sounds. Secondly, due to the complexity of the analysis, the extensive production data, and the deadlines, the focus had to be reduced to a subset of vowel sounds. As for the AI tools used, the anticipated efficacy of Siri and Speakometer did not materialize as expected.

Siri exhibited shortcomings in comprehension, while Speakometer failed to deliver accurate pronunciation guidance, which likely influenced the study's outcomes.

To achieve more consistent and robust results, future research should ideally involve a greater number of participants and practice the first eight vowel sounds suggested in this thesis over a longer time frame—ideally several months or even years. The difficulties in differentiating the vowel pair /ɝ/-/ɔr/ should also be examined more in-depth. As for IPAs and CAPT tools, more investigation is required to ascertain whether AI can genuinely assist with L2 pronunciation or if it unintentionally causes students to acquire incorrect pronunciation patterns.

5. Conclusion

Despite the remarkable strides in machine learning and deep learning, propelling advancements in AI, ASR, and NLP, this study has demonstrated that there is still a long journey ahead before AI-powered tools can fully optimize EFL pronunciation learning and practice. The integration of AI did not appear to enhance pronunciation learning beyond the capabilities of traditional methodologies, nor did it effectively bring students' pronunciation closer to native-like articulation or improve intelligibility. It has been found that, as of today, IPAs' comprehension is notably limited by the type of stimuli and the vowel sounds produced by students. Likewise, CAPT tools do not truly help L2 speakers refine their English pronunciation skills, largely due to the high error rate and the reduced practical exercises they allow for free.

The incorporation of AI in language learning is still in its nascent stages, and until substantial improvements in AI comprehension and interaction with non-native speakers are seen, its pedagogical application remains a subject of scrutiny. Language learners necessitate the guidance of human instructors who can effectively teach English phonetics and phonology and support them in achieving intelligibility and comprehensibility in oral communication. These instructors are crucial in offering constructive feedback and helping learners understand their errors effectively.

It is paramount to familiarize students with English speaking and pronunciation from the moment they start learning the language. Doing so could help L2 learners to clearly differentiate tense and lax vowel sounds and prevent the assimilation of unfamiliar sounds to those of their L1 phonological inventory. Likewise, it could alleviate the anxiety related to SLA and increase their willingness to communicate. To achieve this, EFL teachers require appropriate training and resources to teach pronunciation and give it the importance it deserves within the classroom environment.

While AI tools may eventually evolve and become proficient enough to find a place within L2 pronunciation instruction, learners should never underestimate the irreplaceable value of the expertise and human touch that EFL teachers bring to the classroom—a quality that machines have not yet been able to match. This particular facet of classroom interaction is still unmatched by technology, highlighting the continued importance of human teachers in the process of teaching and learning languages.

References

Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning With Applications*, *2*, 100006. https://doi.org/10.1016/j.mlwa.2020.100006

Aini, D. N., Setiawan, S., Anam, S., Masyitoh, R., Ibad, W., & DJ, A. W. (2023). Artificial intelligence Learning Platform for Promoting Speaking Competence: A Suggested scaffolding for Autonomous Learning in Pesantren. In *Advances in Social Science, Education and Humanities Research/Advances in social science, education and humanities research* (pp. 131–144). https://doi.org/10.2991/978-2-38476-120-3_13

Alharthi, S. M. (2024). Siri as an interactive pronunciation coach: its impact on EFL learners. *Cogent Education*, *11*(1). https://doi.org/10.1080/2331186x.2024.2304245

Apple. (n.d.). *Siri*. https://www.apple.com/siri/

Avery, P., & Ehrlich, S. (2013). *Teaching American English Pronunciation*. Oxford University Press.

Baran-Łucarz, M. (2014). The link between pronunciation anxiety and willingness to communicate in the foreign-language classroom: The Polish EFL context. *Canadian Modern Language Review*, *70*(4), 445–473. https://doi.org/10.3138/cmlr.2666

Barber, C., Barber, C. L., Beal, J., & Shaw, P. (2012). *The English language*. Cambridge University Press.

Bellegarda, J. R. (2013). Large-scale personal assistant technology deployment: The siri experience. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (pp. 2029–2033). International

Speech and Communication Association. https://doi.org/10.21437/interspeech.2013-481

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model .In C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception* (pp. 167–224). Cambridge, MA: The MIT Press.

Best, C. T. & Tyler, M. D. (2007). Nonnative and second-language speech perception. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James E. Flege* (pp. 13–34).Amsterdam: John Benjamins.

Bliss, H., Johnson, K. A., Burton, S., Yamane, N., & Gick, B. (2017). Using multimedia resources to integrate ultrasound visualization for pronunciation instruction into postsecondary language classes. *Journal of Linguistics and Language Teaching*, *8*(2), 47–62

Boersma, P., & Weenink, D (2023). Praat: doing phonetics by computer (Version 6.3.17) [Software]. Available from https://www.praat.org.

Bringsjord, S., & Govindarajulu, N. S. (2018). *Artificial Intelligence*. Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/artificial-intelligence/#WhatExacAI

Cavalcanti, J. C., Eriksson, A., & Barbosa, P. A. (2021). Acoustic analysis of vowel formant frequencies in genetically-related and non-genetically related speakers with implications for forensic speaker comparison. *PloS one*, *16*(2), e0246645. https://doi.org/10.1371/journal.pone.0246645

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics,* *34*(3), 372–387.

https://doi.org/10.1016/j.wocn.2005.08.003

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

Chen, H. H., Yang, C. T., & Lai, K. K. (2020). Investigating college EFL learners' perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments, 31*(3), 1335–1350. https://doi.org/10.1080/10494820.2020.1833043

Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Peter Campbell, J. (2020). Introduction to machine learning, neural networks, and deep learning. *Translational Vision Science and Technology*, *9*(2). https://doi.org/10.1167/tvst.9.2.14

Clark, L., Doyle, P. R., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., Aylett, M. P., Cabral, J. P., Munteanu, C., Edwards, J., & Cowan, B. R. (2019). The State of Speech in HCI: Trends, Themes and challenges. *Interacting With Computers*, *31*(4), 349–371. https://doi.org/10.1093/iwc/iwz016

Coe, N. (2001). Speakers of Spanish and Catalan. In Swan, M. & Smith, B (Eds.), *Learner English: A Teacher's Guide to Interference and Other Problems* (2nd ed., pp. 90–112). Cambridge University Press.

Coheur, L. (2020). From Eliza to Siri and Beyond. In *Communications in Computer and Information Science* (Vol. 1237 CCIS, pp. 29–41). Springer. https://doi.org/10.1007/978-3-030-50146-4_3

CSU Global. (2021). How does AI actually work? https://csuglobal.edu/blog/how-does-ai-actually-work

Daniels, P., & Iwago, K. (2017). The suitability of cloud-based speech recognition engines for language learning. *The JALT CALL Journal*, *13*(3), 211–221. https://doi.org/10.29140/jaltcall.v13n3.220

Dakakni, D., & Safa, N. (2023). Artificial intelligence in the L2 classroom: Implications and challenges on ethics and equity in higher education: A 21st century Pandora's box. *Computers and Education. Artificial Intelligence*, *5*, 100179. https://doi.org/10.1016/j.caeai.2023.100179

Darcy, I. (2018). Powerful and Effective Pronunciation Instruction: How Can We Achieve It? *Catesol Journal*, *30*(1), 13-45. https://files.eric.ed.gov/fulltext/EJ1174218.pdf

Derwing, T. M., & Munro, M. J. (2005). Second Language Accent and Pronunciation Teaching: A Research-Based Approach. *TESOL Quarterly*, *39*(3), 379. https://doi.org/10.2307/3588486

Dizon, G. (2017). Using Intelligent Personal Assistants for Second Language Learning: A Case Study of Alexa. *TESOL Journal*, *8*(4), 811–830. https://doi.org/10.1002/tesj.353

Dizon, G. (2021). Affordances and Constraints of Intelligent Personal Assistants for Second-Language Learning. *RELC Journal*, *54*(3), 848–855. https://doi.org/10.1177/00336882211020548

El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, *38*(3). https://doi.org/10.1111/exsy.12667

Flege, J. E., & Wayland, R. (2019). The role of input in native Spanish Late learners' production and perception of English phonetic segments. *Journal of Second Language Studies*, *2*(1), 1–44. https://doi.org/10.1075/jsls.00004.fle

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

Gardner, H. (1983). *Frames of Mind. The Theory of Multiple Intelligences*. Basic Books.

Garita Sánchez, M. D. R., González Lutz, M. I., & Solís Pérez, N. (2019). English vowel sounds: Pronunciation issues and student and faculty perceptions. *Revista Actualidades Investigativas En Educación*, *19*(3), 1–32. https://doi.org/10.15517/aie.v19i3.38629

Gick, B., Bernhardt, B., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In M. L. Zampini & J. G. Hansen Edwards (Eds.), *Phonology and Second Language Acquisition* (pp. 309–322). John Benjamins Publishing Company.

Gómez González, M. D. L. Á., & Sánchez Roura, T. S. (2016). *English Pronunciation for Speakers of Spanish: From Theory to Practice*. De Gruyter Mouton.

Hillenbrand, J., Getty, L. A., Wheeler, K., & Clark, M. J. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *95*(5), 3099–3111. https://doi.org/10.1121/1.409456

Hirai, A., & Kovalyova, A. (2023). Using Speech-to-Text Applications for Assessing English Language Learners' Pronunciation: A Comparison with Human Raters. In *English language education* (pp. 337–355). https://doi.org/10.1007/978-3-031-27825-9_17

Holt, E. (2018). Teaching tip: The use of MRI and ultrasound technology in teaching about Spanish (and general) phonetics and pronunciation. In J. Levis (Ed.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference* (pp. 252–262). Iowa State University. https://www.researchgate.net/publication/327867809

Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice

Assistants. *Medical Reference Services Quarterly*, *37*(1), 81–88. https://doi.org/10.1080/02763869.2018.1404391

Huckvale, M. (2017). *Introduction to Speech Science. Vowels*. Speech, Hearing & Phonetic Science. UCL Division of Psychology and Language Sciences. https://www.phon.ucl.ac.uk/courses/spsci/iss/week5.php

Humardhiana, A. (2022). AI-POWERED APPS TO ENHANCE NOVICE NEWSREADERS' ENGLISH PRONUNCIATION. *ELT Echo : The Journal of English Language Teaching in Foreign Language Context*, *7*(2). https://doi.org/10.24235/eltecho.v7i2.12503

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Jarosz, A. (2019). English pronunciation in l2 instruction: The case of secondary school learners. *Second Language Learning and Teaching*. Springer Verlag.

Kenyon, J. S. (1997). *American Pronunciation* (D. M. Lance & S. A. Kingsbury, Eds.; 12th ed.). George Wahr.

Kholis, A. (2021). Elsa Speak App: Automatic Speech Recognition (ASR) for supplementing English pronunciation skills. *Pedagogy : Journal of English Language Teaching/Pedagogy : Journal of English Language Teaching*, *9*(1), 01. https://doi.org/10.32332/joelt.v9i1.2723

Kurt, Ş. (2022). Speakometer: English Pronunciation Coach. *TESL-EJ*, *26*(101). https://doi.org/10.55593/ej.26101m2

Ladefoged, P., & Johnson, K. (2014). *A course in phonetics*. Cengage Learning.

Lawson, E., Stuart-Smith, J., Scobbie, J. M., & Nakai, S. (2018a). *Dynamic Dialects: An articulatory web resource for the study of accents*. University of Glasgow.

https://www.dynamicdialects.ac.uk/

Lawson, E., Stuart-Smith, J., Scobbie, J. M., & Nakai, S. (2018b). *Seeing Speech: an articulatory web resource for the study of Phonetics.* University of Glasgow. https://seeingspeech.ac.uk/

Learn Academic English. (2020). *NATIVE ENGLISH PRONUNCIATION: How to Make the /æ/ Sound in American English Pronunciation* [Video]. YouTube. https://www.youtube.com/watch?v=U6g10UG5pu0

Levis, J. M., & Grant, L. (2003). Integrating Pronunciation Into ESL/EFL Classrooms. *TESOL Journal*, *12*(2), 13–19. https://doi.org/10.1002/j.1949-3533.2003.tb00125.x

Levis, J. & Suvorov, R. (2012). Automatic speech recognition. In C. Chapelle (Ed.) *The encyclopedia of applied linguistics*.

Liakin, D., Cardoso, W., & Liakina, N. (2014). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, *32*(1), 1–25. https://doi.org/10.1558/cj.v32i1.25962

Lippmann, R. (2023). Understanding and Applying Deep Learning. *Neural Computation*, *35*(3), 287–308. https://doi.org/10.1162/neco_a_01518

Liu, B. (2021). “Weak AI” is Likely to Never Become “Strong AI”, So What is its Greatest Value for us? *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2103.15294

Liu, M. (2006). Anxiety in EFL Classrooms: Causes and Consequences. *TESL Reporter*, *39*(1), 13–32. https://lir.byuh.edu/index.php/Issue1/article/view/1189

Macdonald, S. (2002). Pronunciation – views and practices of reluctant teachers. *Prospect*, *17*(3), 3–18. https://www.researchgate.net/publication/285517759_Pronunciation-views_and_practices_of_reluctant_teachers

Martínez Celdrán, E. (1995). En torno a las vocales del español: análisis y reconocimiento. *Journal of Experimental Phonetics*, *7,* 197–218. Retrieved from https://revistes.ub.edu/index.php/experimentalphonetics/article/view/44617

Martínez Celdrán, E., & Elvira-García, W. (2019). Description of Spanish Vowels and Guidelines for Teaching Them. In R. Rao (Ed.), *Key Issues in the Teaching of Spanish Pronunciation* (pp. 17–39). Routledge.

Martinez, R. (2019). Artificial Intelligence: Distinguishing Between Types & Definitions. *Nevada Law Journal, 19*(3), 1015–1042. https://scholars.law.unlv.edu/cgi/viewcontent.cgi?article=1799&context=nlj

McCarthy, J. (2007). What is artificial intelligence? Computer Science Department, Stanford University. https://www-formal.stanford.edu/jmc/whatisai.pdf

Molden, M. (2015). Employing Apple's Siri to practice pronunciation: A preliminary study on Arabic speakers. Hawaii Pacific University TESOL Working Paper Series, 13, 2-17. https://www.hpu.edu/research-publications/tesol-working-papers/2015/02Molden2015Siri.pdf

Montesinos López, O.A., Montesinos López, A., Crossa, J. (2022). Fundamentals of Artificial Neural Networks and Deep Learning. In *Multivariate Statistical Machine Learning Methods for Genomic Prediction* (pp. 379–425) Springer, Cham. https://doi.org/10.1007/978-3-030-89010-0_10

Morandín-Ahuerma, F. (2022). What is Artificial Intelligence? *International Journal of Research Publication and Reviews*, *3*(12), 1947–1951. https://ijrpr.com/uploads/V3ISSUE12/IJRPR8827.pdf

Moussalli, S., & Cardoso, W. (2016). Are commercial 'personal robots' ready for language learning? Focus on second language speech. In S. Papadima-

Sophocleous, L. Bradley & S. Thouësny (Eds), *CALL communities and culture – short papers from EUROCALL 2016* (pp. 325-329). https://doi.org/10.14705/rpnet.2016.eurocall2016.583

Moussalli, S., & Cardoso, W. (2019). Intelligent personal assistants: can they understand and be understood by accented L2 learners? *Computer Assisted Language Learning, 33*(8), 865–890. https://doi.org/10.1080/09588221.2019.1595664

Neri, A., Cucchiarini, C., & Strik, H. (2008). The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. *ReCALL*, 20, 225–243. https://doi.org/10.1017/S0958344008000724

Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, *21*(5), 393–408. https://doi.org/10.1080/09588220802447651

Odden, D. (2013). *Introducing Phonology* (2nd ed.). Cambridge University Press.

Ogden, R. (2009). *Introduction to English phonetics*. Edinburgh University Press.

Park, M. S. (2015). Second Language Pronunciation Teaching: Insights from Research. *Academic Commons*. https://doi.org/10.7916/D83V0V38

Paschen, U., Pitt, C., & Kietzmann, J. (2020). Artificial intelligence: Building blocks and an innovation typology. Business Horizons, 63(2), 147–155. https://doi.org/10.1016/j.bushor.2019.10.004

Pieraccini, R. (2012). *The voice in the machine: Building Computers That Understand Speech*. MIT Press.

Rallo Fabra, L. (2022). Factors affecting pronunciation accuracy in English as a foreign language: the case of Spanish-Catalan intermediate learners. *Atlantis, 44*(2), 45–64. https://doi.org/10.28914/atlantis-2022-44.2.03

Rallo Fabra, L., & Romero, J. (2012). Native Catalan learners’ perception and production of English vowels. *Journal of Phonetics*, *40*(3), 491–508. https://doi.org/10.1016/j.wocn.2012.01.001

Rallo Fabra, L., & Tyler, M. D. (2021). Discrimination of Californian central vowel contrasts by Spanish-Catalan EFL learners. In *Proc. 3rd International Symposium on Applied Phonetics (ISAPh 2021)* (pp. 59–62). https://doi.org/10.21437/isaph.2021-10

Raphael, L. & Bell-Berti, F. (1975). Tongue Musculature and the Feature of Tension in English Vowels. *Phonetica*, *32*(1), 61-73. https://doi.org/10.1159/000259686

Recasens, Daniel (2014). *Fonètica i Fonologia Experimentals del Català. Vocals i Consonants*. Biblioteca Filològica, Institut d’Estudis Catalans.

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

Romero, J. (n.d.-a) *English vowels. Vowel length* [PowerPoint slides]. Department of English and German Studies, Universitat Rovira i Virgili.

Romero, J. (n.d.-b). *The sound system of English. Consonants and vowels* [PowerPoint slides]. Department of English and German Studies, Universitat Rovira i Virgili. https://campusvirtual.urv.cat/pluginfile.php/5238644/mod_resource/content/5/UNIT%202.3%20Consonants%20and%20vowels.pdf

Rosell-Aguilar, F. (2017). State of the App: A taxonomy and framework for evaluating language learning mobile applications. *CALICO Journal*, *34*(2). https://doi.org/10.1558/cj.27623

Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: A Modern Approach* (3rd ed.). Prentice Hall.

Saito, K., Trofimovich, P., & Isaacs, T. (2015). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, *37*(2), 217–240. https://doi.org/10.1017/S0142716414000502

San Diego Voice and Accent. (2020). *Learn the American accent! How to Pronounce the UH /ʌ/ Vowel* [Video]. YouTube. https://www.youtube.com/watch?v=ogwHrVH2Paw

Setter, J., & Jenkins, J. (2005). State-of-the-Art Review Article. *Language Teaching*, *38*(1), 1–17. https://doi.org/10.1017/s026144480500251x

Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., & Akinkuolie, B. (2021). Gamification in mobile-assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020. *Computer Assisted Language Learning*, *36*(3), 517–554. https://doi.org/10.1080/09588221.2021.1933540

Sofia, S., Montserrat, L. C., Emilia, G. G., Giuditta, D. P., Fernando, M., & Blagoj, D. (2020). AI WATCH. Defining Artificial Intelligence. *Publications Office of the European Union*, JRC118163. https://doi.org/10.2760/382730

Sounds American. (2017). *R-Colored Vowel Sound / ɝ / as in "first" - American English Pronunciation* [Video]. YouTube. https://www.youtube.com/watch?v=6ppOrwjvslc

Sounds American. (2018). *R-Colored Vowel Sound / ɔr / as in "sport" – American English Pronunciation* [Video]. YouTube. https://www.youtube.com/watch?v=ZbDrxmP4_S4

Speakometer. (n.d.). *Speakometer: Free American/British Accent training app | Learn English Pronunciation*. https://www.speakometer.net/

Svenningsson, N., & Faraon, M. (2019). Artificial Intelligence in Conversational Agents: A Study of Factors Related to Perceived Humanness in Chatbots. In *ACM International Conference Proceeding Series* (pp. 151–161). Association for Computing Machinery. https://doi.org/10.1145/3375959.3375973

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

Toutios, A., Lingala, S. G., Vaz, C., Kim, J., Esling, J., Keating, P., Gordon, M., Byrd, D., Goldstein, L., Nayak, K. Narayanan, S. (2016). Illustrating the production of the international phonetic alphabet sounds using fast real-time magnetic resonance imaging. In *Proceedings of the Annual Conference of the International Speech Communication Association*, *INTERSPEECH*, *8*(12), 2428–2432. International Speech and Communication Association. http://dx.doi.org/10.21437/Interspeech.2016-605

Wayland, R. (2018). Phonetics: A practical introduction. Cambridge University Press. https://doi.org/https://doi.org/10.1017/9781108289849

Yavaş, M. (2011). *Applied English Phonology* (2nd ed.). John Wiley & Sons.

Appendix A

Consent Form

UNIVERSITAT ROVIRA i VIRGILI

Formulario de Consentimiento

Título del Trabajo de Fin de Máster (TFM): *Talking with Machines: Exploring AI as a Tool to Improve the English Pronunciation of Spanish/Catalan Teenagers* [Hablando con Máquinas: Explorando la IA como una Herramienta para Mejorar la Pronunciación del Inglés de los Adolescentes Españoles/Catalanes].

Número de referencia: CEIPSA-2024-TFM-0009.

Datos de contacto del investigador: Marina Palomo Martín – marina.palomo@urv.cat – Teléfono: 977559755 – Dirección postal: Departamento de Estudios Ingleses y Alemanes; Campus Catalunya, Av. Catalunya, 35, 43002, Tarragona.

INFORMACIÓN SOBRE PROTECCIÓN DE DATOS

Controlador de Datos: El controlador de datos es la Universitat Rovira i Virgili con Número de Identificación Fiscal Q9350003A y con sede en Carrer de l'Escorxador, s/n, 43003, Tarragona.

Propósito: Participar en la Tesis de Máster bajo los términos descritos en la hoja de información para el participante. Si el estudio tiene la intención de publicar, difundir y reutilizar los resultados obtenidos, incluidos datos personales, los datos personales se utilizarán para estos fines siempre que la parte interesada haya dado su consentimiento.

Derechos: Los interesados pueden ejercer su derecho de acceso, rectificación, supresión, circulación, limitación u oposición al tratamiento de sus datos dirigiéndose por escrito al Registro General de la URV en la misma dirección de la URV, o personalmente en el Registro General de la URV o telemáticamente de acuerdo con las instrucciones en https://seuelectronica.urv.cat/registre.html.

Más información: Las personas pueden encontrar información adicional sobre el tratamiento de datos personales en la *Tesis de Máster en la URV* y sobre sus derechos en el Registro de Tratamientos de la URV, que se publica en https://seuelectronica.urv.cat/rgpd, donde también encontrarán la Política de Privacidad de la URV. También pueden encontrar esta información en la Hoja de Información a la Persona Participante sobre el estudio. Además, pueden hacer preguntas a nuestros delegados de protección de datos sobre la protección de datos personales enviando un correo electrónico a dpd@urv.cat.

* Required

1. Apellidos, Nombre: *

2. DNI/NIE: *

3. - He leído la copia que he recibido del documento de información al participante sobre el estudio.
- He podido preguntar y he recibido respuesta a mis dudas personales sobre el estudio y mi participación en el mismo.
- Entiendo que estoy participando en este estudio de acuerdo con las especificaciones del documento de información del participante y de acuerdo con las respuestas que he recibido a mis preguntas y entiendo los riesgos y beneficios que esto conlleva.
- Acepto que mi participación es voluntaria y acepto libremente participar en el estudio.
- Entiendo que puedo retirarme en cualquier momento de participar en el estudio y que mi retiro no me afectará negativamente de ninguna manera.
- He sido informado sobre cómo se procesarán mis datos personales.
- Doy mi consentimiento para que mis datos sean accedidos y utilizados en las condiciones especificadas en el documento que contiene información sobre el estudio dirigido al participante. *

◯ Sí

◯ No

4. Una vez finalizada la investigación, los datos obtenidos pueden ser de interés para otros estudios relacionados. En este sentido, se ofrecen las siguientes opciones: *

◯ NO AUTORIZAR el uso de los datos en otros proyectos de investigación relacionados.

◯ AUTORIZAR el uso de los datos en otros proyectos de investigación relacionados.

5. Fecha y ubicación (ciudad). *

6. Al escribir su nombre y enviar este formulario, acepta participar en este estudio. *

Appendix B

Students' Sheet for the Pretest and Posttest

1	feel	fill
2	match	much
3	pull	pool
4	warm	worm
5	least	list
6	ankle	uncle
7	fool	full
8	burn	born

1	The blanket keeps Jill warm.
2	Since when does Luke work so much?
3	I saw four furry cats at school.
4	I took an English class this morning.
5	Uncle Max hurt his foot.
6	John reads his new book at the beach.
7	Can you fill my cup of tea with sugar?
8	The young girl was born in June.

Appendix C

Pretest and Posttest: Sheets for Grading Siri's Understanding

MINIMAL PAIRS

<table>
<tr><td colspan="3">1) FEEL - FILL
/i/ - /ɪ/</td><td colspan="3">2) MATCH - MUCH
/æ/ - /ʌ/</td><td colspan="3">3) PULL - POOL
/ʊ/ - /u/</td><td colspan="3">4) WARM - WORM
/ɔr/ - /ɝ/</td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td></tr>
<tr><td colspan="3">5) LEAST - LIST
/i/ - /ɪ/</td><td colspan="3">6) ANKLE - UNCLE
/æ/ - /ʌ/</td><td colspan="3">7) FOOL - FULL
/u/ - /ʊ/</td><td colspan="3">8) BURN - BORN
/ɝ/ - /ɔr/</td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td><td>0</td><td>1</td><td>2</td></tr>
</table>

SENTENCES

<table>
<tr><td colspan="5">1) The blanket keeps Jill warm.
/æ/ /i/ /ɪ/ /ɔr/</td><td colspan="5">2) Since when does Luke work so much?
/ɪ/ /u/ /ɝ/ /ʌ/</td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td colspan="5">3) I saw four furry cats at school.
/ɔr/ /ɝ/ /æ/ /u/</td><td colspan="5">4) I took an English class this morning.
/ʊ/ /ɪ/ /æ/ /ɔr/</td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td colspan="5">5) Uncle Max hurt his foot.
/ʌ/ /æ/ /ɝ/ /ʊ/</td><td colspan="5">6) John reads his new book at the beach.
/i/ /u/ /ʊ/ /i/</td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td></tr>
<tr><td colspan="5">7) Can you fill my cup of tea with sugar?
/ɪ/ /ʌ/ /i/ /ʊ/</td><td colspan="5">8) The young girl was born in June.
/ʌ/ /ɝ/ /ɔr/ /u/</td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td></tr>
</table>

Appendix D

Pretest: Background Information Questions and their Translation

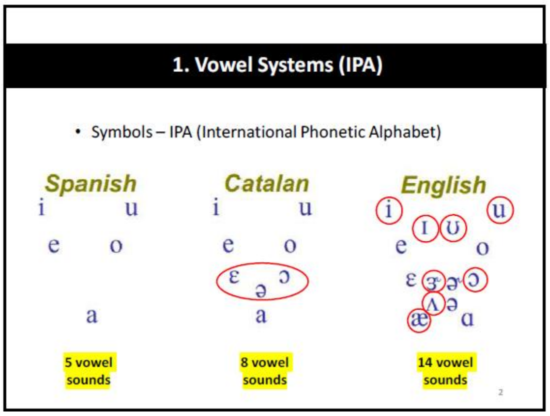
Table 17. *Pretest: Background Information Questions and Their Translation*

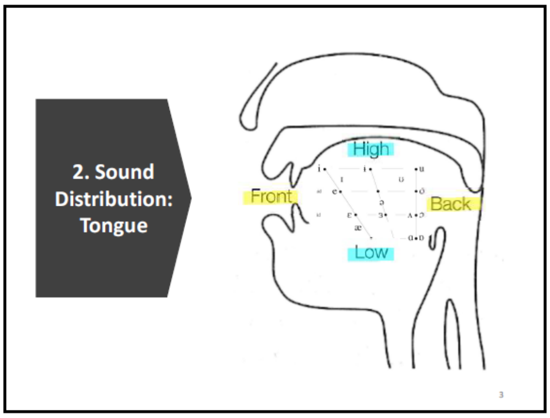
Questions	Translation
1. ¿Dónde naciste?	1. Where were you born?
2. ¿Qué idiomas hablas?	2. What languages do you speak
3. ¿Vas a clases de inglés fuera del horario escolar?	3. Do you go to English classes outside of school?
4. ¿A qué edad empezaste a aprender inglés?	4. At what age did you start learning English?

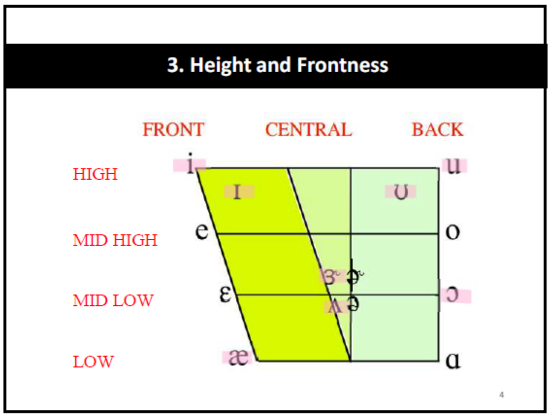
Appendix E

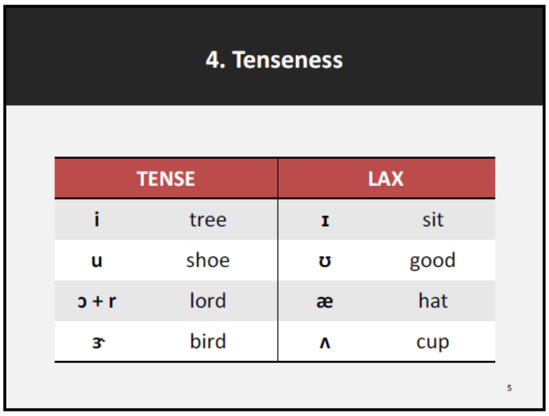
PowerPoint Presentations for the Intervention

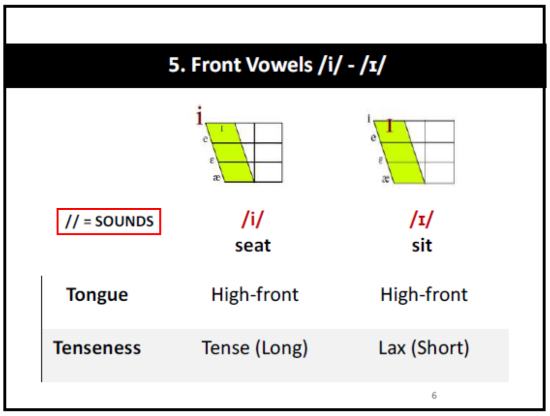
PowerPoint Presentation: Class 1











6. MRI of Vowels

[/i/]

[/ɪ/]

8. Sound Distribution and Spelling

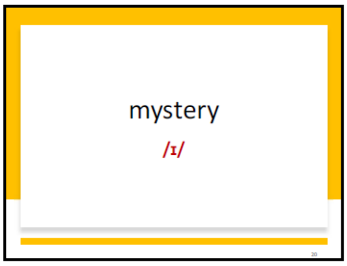
	/i/				/ɪ/	
SOUND	Beginning Middle End + /r/				Beginning Middle	
SPELLING	ee	keep, beer	ie	cookie	i	big, English
	ea	beach	ei	receive	y	mystery
	e	be	ey	key	e	English
	y	happy	eo	people		

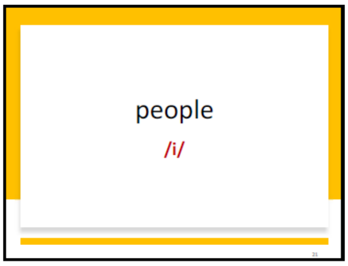
9. Practice

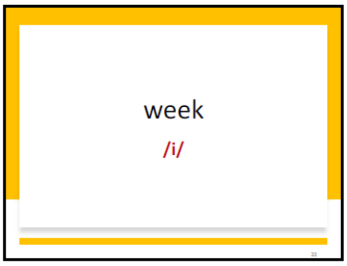
Speakometer

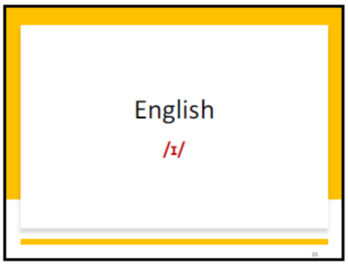
1. Download app on Apple Store / Play Store (in pairs)
2. Native language: Catalan/Spanish + Practice accent: American
3. Go to LEARN → Short and Long vowels → ɪ - i
4. Click on ⓘ or ⓒ → Speakometer (10 words)
5. Listen to words and record yourself (microphone)
6. Allow app to record
7. At the end, tell me your score

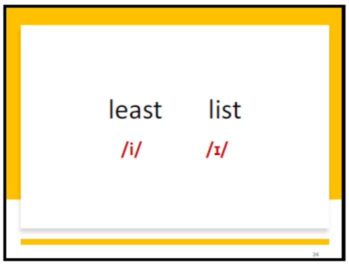
dream /i/	hit /ɪ/ heat /i/	since /ɪ/
think /ɪ/	beach /i/	sheep /i/ ship /ɪ/
green /i/	fill /ɪ/ feel /i/	big /ɪ/

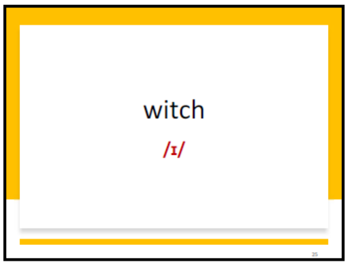


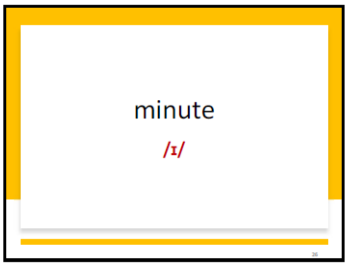




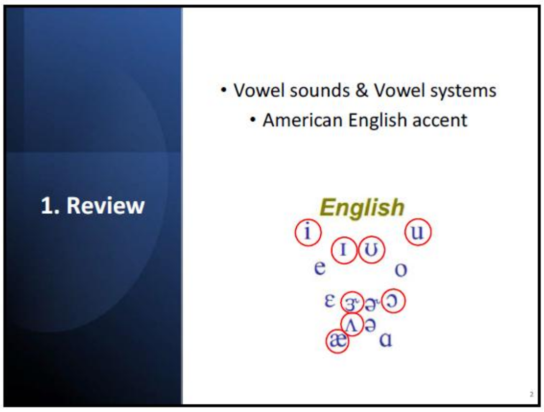


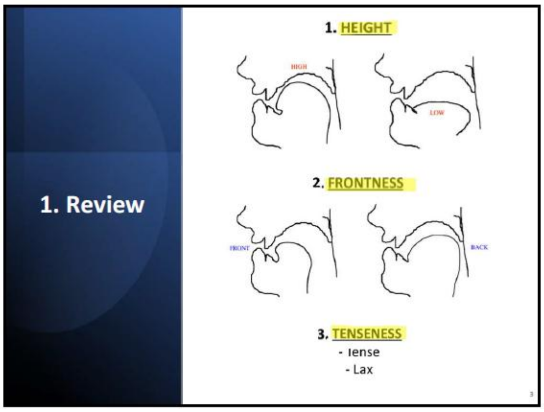


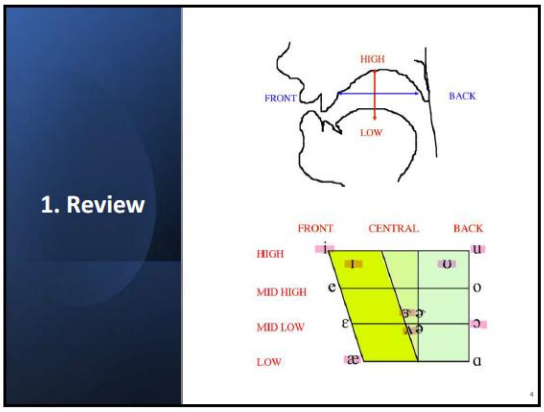


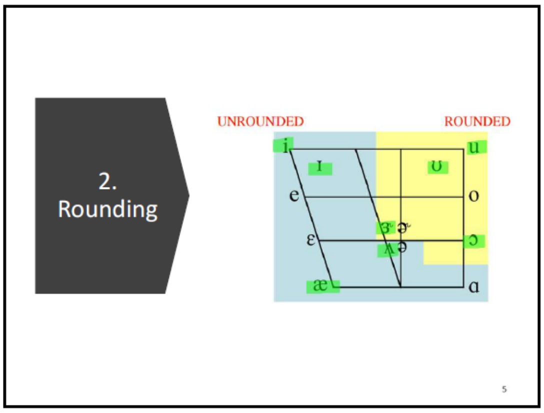


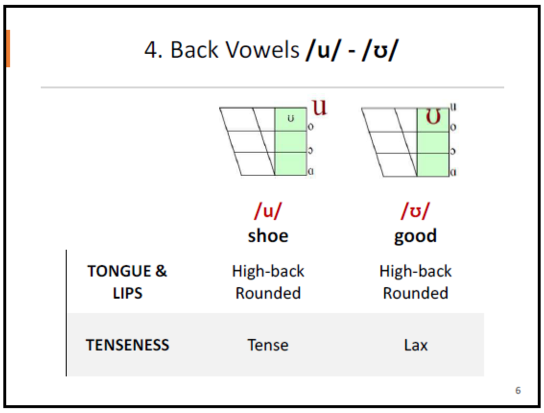
PowerPoint Presentation: Class 2



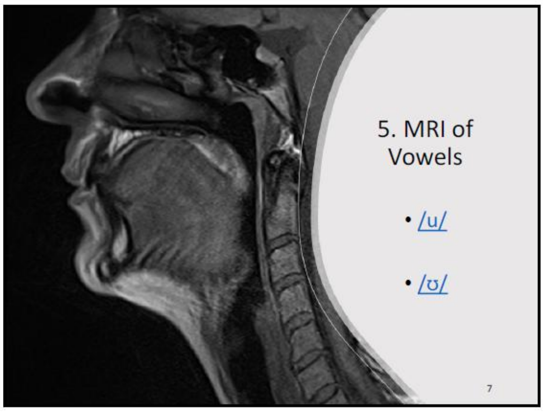


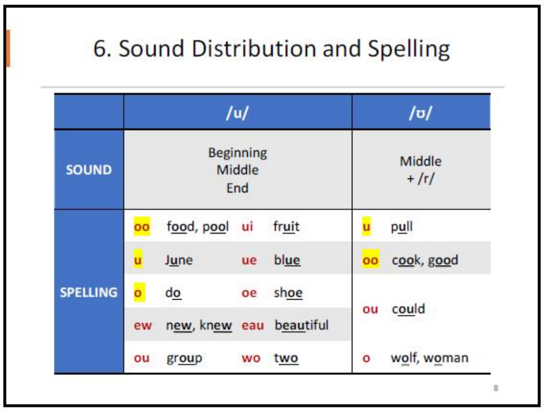






	/u/ shoe	/ʊ/ good
TONGUE & LIPS	High-back Rounded	High-back Rounded
TENSENESS	Tense	Lax





	/u/				/ʊ/	
SOUND	Beginning Middle End				Middle + /r/	
SPELLING	oo	food, pool	ui	fruit	u	pull
	u	June	ue	blue	oo	cook, good
	o	do	oe	shoe	ou	could
	ew	new, knew	eau	beautiful		
	ou	group	wo	two	o	wolf, woman

push
/ʊ/

cool
/u/

music
/u/

sugar
/ʊ/

drew
/u/

tour
/ʊ/

suit
/u/

pull /ʊ/ pool /u/

lose
/u/

full /ʊ/ fool /u/

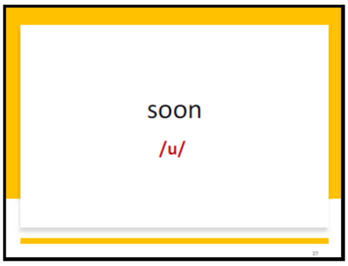
could
/ʊ/

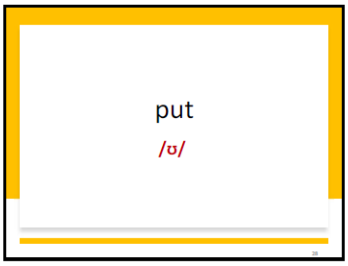
Luke /u/ look /ʊ/

school
/u/

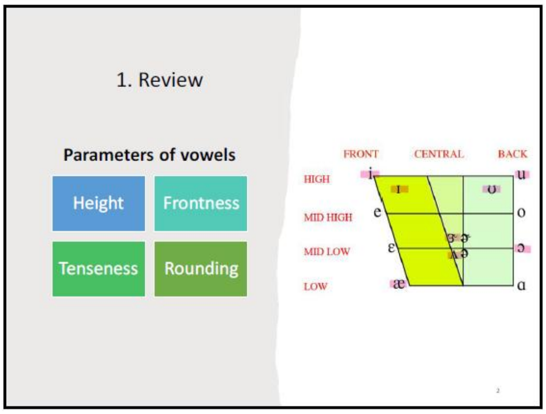
woman
/ʊ/

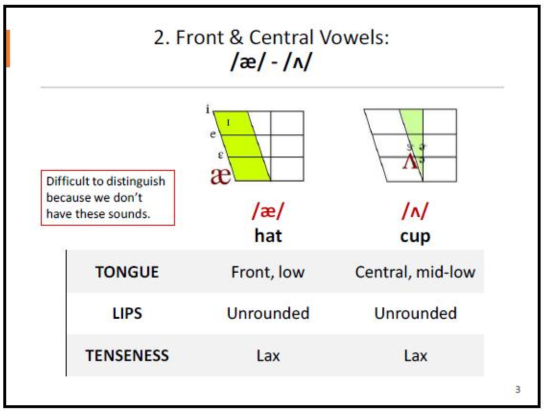
blue book
/u/ /ʊ/

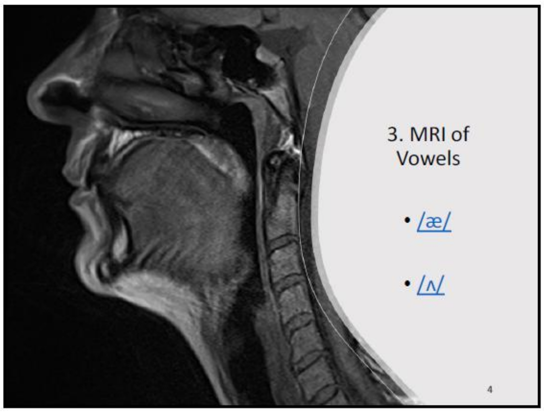


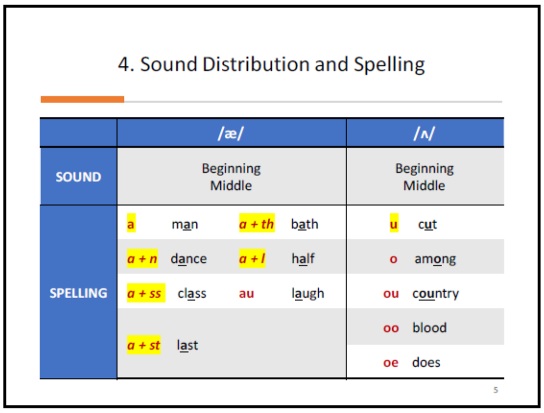


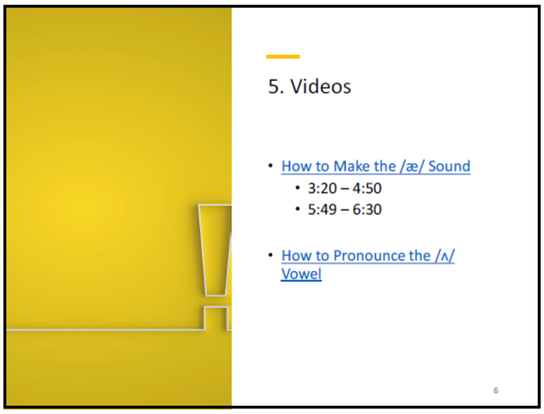
PowerPoint Presentation: Class 3











Speakometer

1. Go to LEARN → Short and Long vowels → æ - ʌ
2. Click on or → Speakometer (10 words)
3. Listen to words and record yourself
4. At the end, tell me your score

hand
/æ/

under
/ʌ/

back /æ/ — buck /ʌ/

aunt
/æ/

none /ʌ/ — nun /ʌ/

unkle /ʌ/ — ankle /æ/

flood
/ʌ/

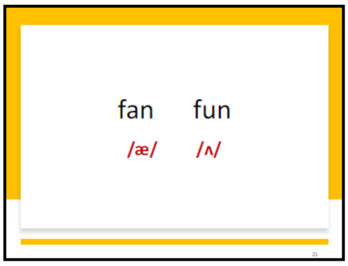
money
/ʌ/

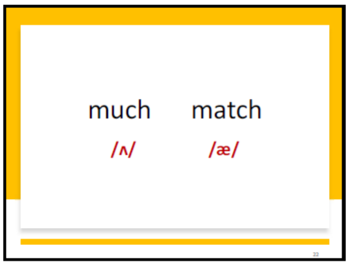
drank /æ/ — drunk /ʌ/

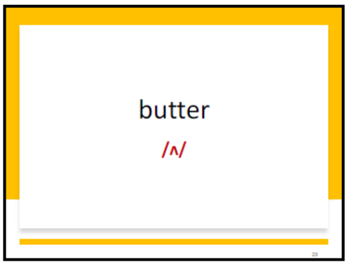
salmon
/æ/

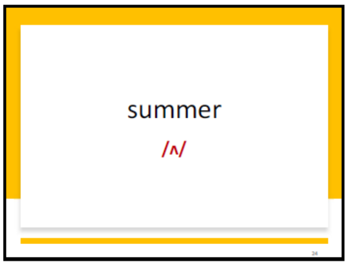
example
/æ/

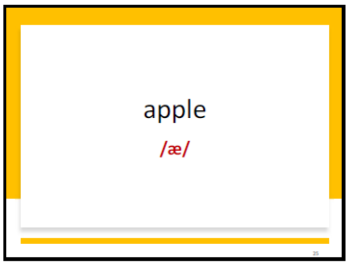
young
/ʌ/

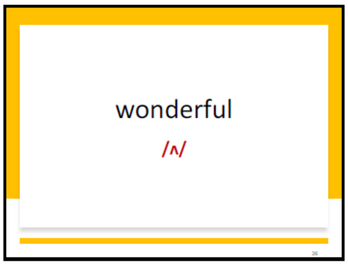




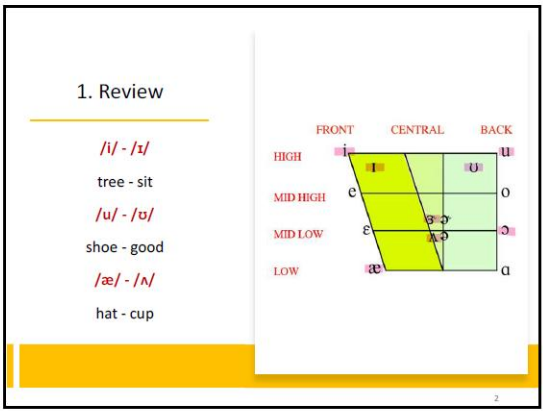


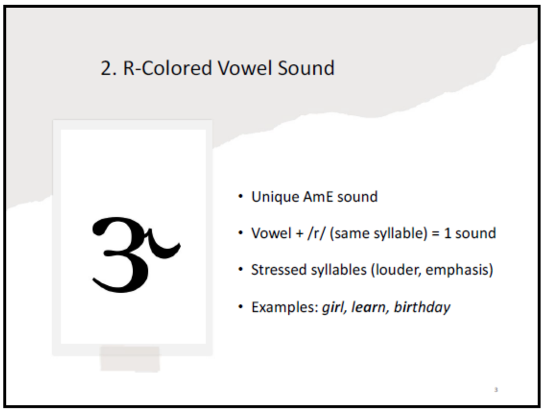


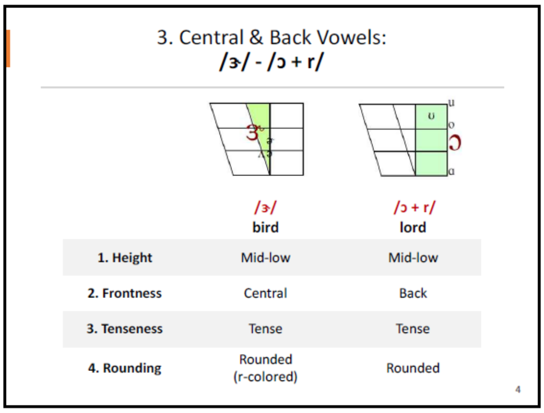


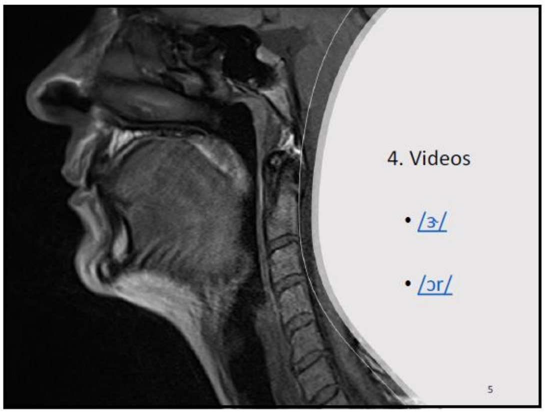


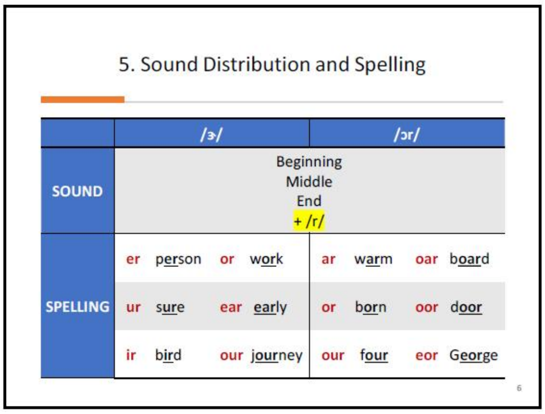
PowerPoint Presentation: Class 4



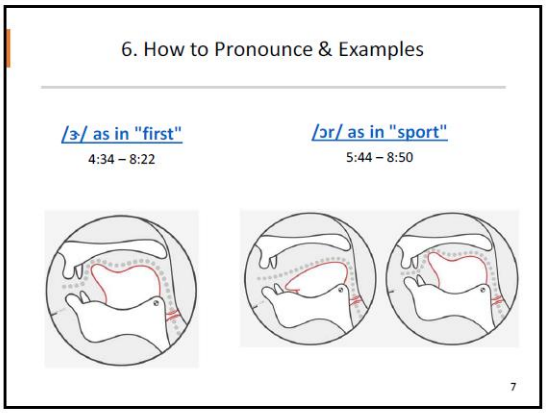


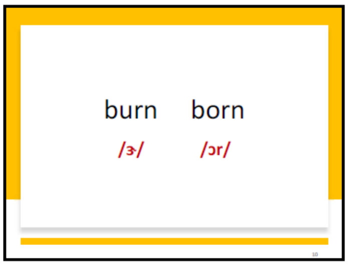


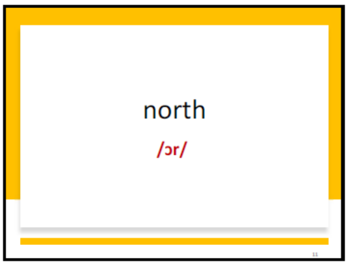


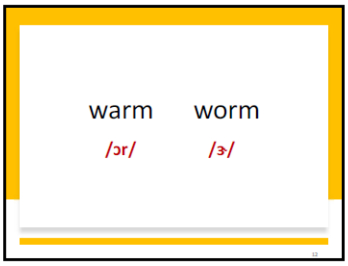


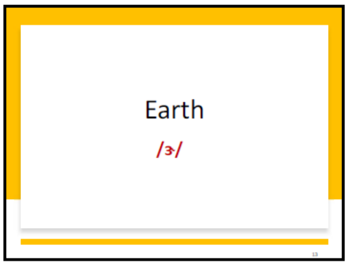
	/ɜ/				/ɔr/			
SOUND	Beginning Middle End + /r/							
SPELLING	er	person	or	work	ar	warm	oar	board
	ur	sure	ear	early	or	born	oor	door
	ir	bird	our	journey	our	four	eor	George

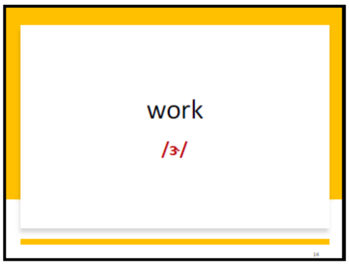


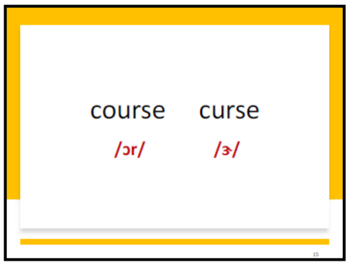


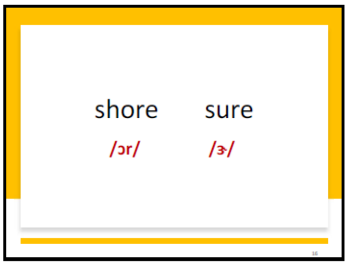


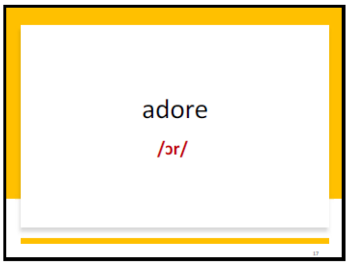


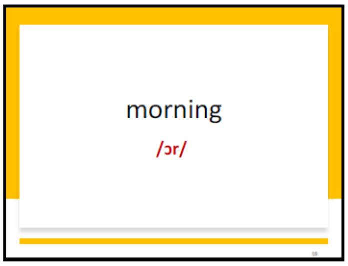

Earth
/ɝ/

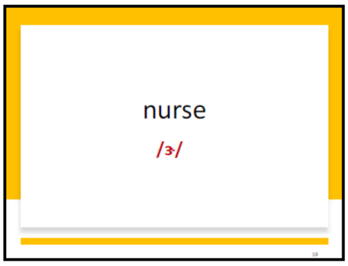

work
/ɝ/

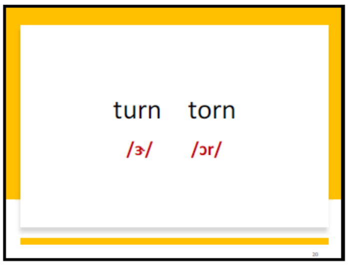

course curse
/ɔr/ /ɝ/

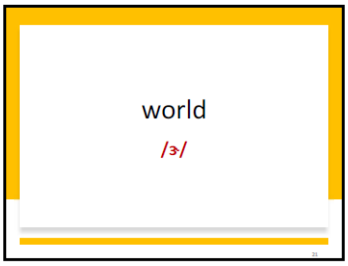

shore sure
/ɔr/ /ɝ/

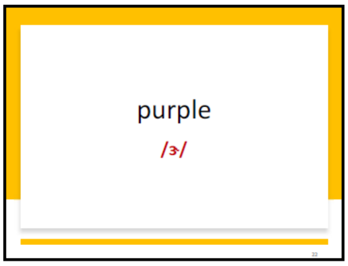

adore
/ɔr/

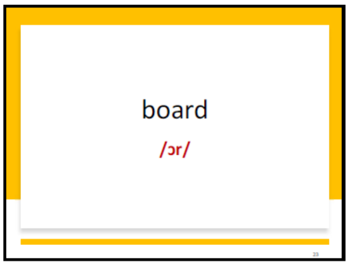

morning
/ɔr/

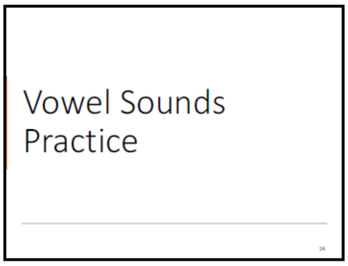

nurse
/ɝ/

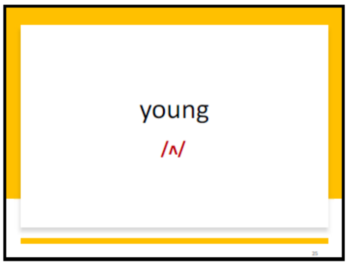

turn torn
/ɝ/ /ɔr/

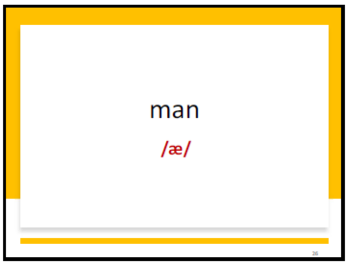

world
/ɝ/

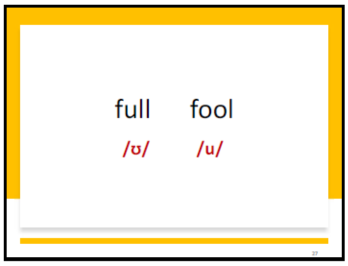

purple
/ɝ/


board
/ɔr/


Vowel Sounds Practice


young
/ʌ/


man
/æ/


full fool
/ʊ/ /u/

bit beat
/ɪ/ /i/

woman
/ʊ/

first
/ɝ/

forth
/ɔr/

book buck
/ʊ/ /ʌ/

lift
/ɪ/

court
/ɔr/

foot food
/ʊ/ /u/

sheep ship
/i/ /ɪ/

horse
/ɔr/

much match
/ʌ/ /æ/

street
/i/

tour
/ʊ/

fast
/æ/

short shirt
/ɔr/ /ɝ/

Appendix F

***Card Game*[3]**

i	ɪ
æ	ʌ
u	ʊ
ɝ	ɔr

[3] Each student received two cards during each class to practice the two vowel sounds taught in that session.

Appendix G

Post-Questionnaire: MS Forms (Original Version in Spanish)

Este breve cuestionario tiene como objetivo conocer vuestra experiencia y opinión personal tras haber recibido clases de pronunciación en inglés. Solo tardaréis unos 5 minutos en completarlo y agradecería vuestra sinceridad en cada respuesta. ¡Gracias por vuestra colaboración!

Lee atentamente cada afirmación y selecciona la respuesta correspondiente en función de si estás totalmente en desacuerdo (1), en desacuerdo (2), ni de acuerdo ni en desacuerdo (3), de acuerdo (4), o totalmente de acuerdo (5).

Table 18. *Post-Questionnaire (Original Version)*

	1	2	3	4	5
1. Me ha gustado aprender pronunciación en inglés.					
2. Creo que mi pronunciación ha mejorado gracias a las clases que hemos hecho.					
3. Aprender pronunciación es difícil.					
4. Es importante aprender pronunciación en las clases de inglés.					
5. Las nuevas tecnologías y la inteligencia artificial son herramientas útiles para practicar la pronunciación en inglés.					
6. La pronunciación del inglés es importante para una comunicación efectiva.					
7. Me gustaría aprender más sobre la pronunciación en inglés.					
8. Las actividades que hemos hecho en clase han sido útiles para mi aprendizaje.					
9. Mi comprensión auditiva del inglés ha mejorado gracias a las clases de pronunciación.					
10. La práctica regular de pronunciación es esencial para mejorar.					

Appendix H

Post-Questionnaire: MS Forms (English Translation)

This brief questionnaire aims to learn about your personal experience and opinion after taking English pronunciation classes. It will only take you about 5 minutes to complete it, and I would appreciate your sincerity in each answer. Thank you for your cooperation!

Read each statement carefully and select the corresponding response based on whether you strongly disagree (1), disagree (2), neither agree nor disagree (3), agree (4), or strongly agree (5).

Table 19. *Post-Questionnaire (English Translation)*

	1	**2**	**3**	**4**	**5**
1. I have enjoyed learning English pronunciation.					
2. I believe my pronunciation has improved thanks to the classes we have had.					
3. Learning pronunciation is difficult.					
4. It is important to learn pronunciation in English classes.					
5. New technologies and artificial intelligence are useful tools for practicing English pronunciation.					
6. English pronunciation is important for effective communication.					
7. I would like to learn more about English pronunciation.					
8. The activities we have done in class have been helpful for my learning.					
9. My listening comprehension of English has improved thanks to pronunciation classes.					
10. Regular pronunciation practice is essential for improvement.					

Appendix I

Examples of Mistakes Produced by Speakometer