Exploring High-Variability Phonetic Training through Text-To-Speech Technology in ESL Pronunciation Pedagogy

Forcan Al-Shami

A Thesis

in

The Department

of

Education

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Arts (Applied Linguistics) at

Concordia University

Montreal, Quebec, Canada

September, 2024

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Forcan Al-Shami

Entitled: Exploring High-Variability Phonetic Training through Text-To-Speech Technology in ESL Pronunciation Pedagogy

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Applied Linguistics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Examiner
Dr. Paul John

_____ Thesis Supervisor
Dr. Walcir Cardoso

Approved by

_____
Dr. Walcir Cardoso Graduate Program Director or Chair of Department

_____
Dr. Pascale Sicotte Dean of Faculty

ABSTRACT

Exploring High-Variability Phonetic Training through Text-To-Speech Technology in ESL Pronunciation Pedagogy

Forcan Al-Shami

Traditional language classrooms often face constraints that limit ESL learners' exposure to varied speech sounds, hindering their pronunciation development. High-Variability Phonetic Training (HVPT) offers a potential solution by exposing learners in diverse speech sounds and accents to improve pronunciation skills. However, the practical implementation of HVPT in everyday language classrooms is still underexplored. This study examines the integration of Text-To-Speech (TTS) technology with HVPT to enhance pronunciation skills among English as a Second Language (ESL) learners beyond traditional classroom settings.

Using a mixed-method design with pretest-posttest evaluations, this study assessed morphophonemic features (*past -ed* allomorphs) and holistic aspects of pronunciation (i.e., comprehensibility, accentedness). Thirty adult university-level ESL learners from Kuwait were assigned to either a Treatment group (TTS with varied voices) or a Control group (TTS with a single voice). Participants engaged in self-paced sessions over a period of four weeks.

The results revealed significant improvements in phonological awareness of past *-ed* allomorphy for both groups, with no notable differences observed between them. In terms of the holistic measures of pronunciation development, evaluated by a panel of eleven raters, the Treatment group achieved statistically significant improvements in both comprehensibility and accentedness when compared to the Control group. These findings indicate that while TTS technology enhances phonological awareness irrespective of HVPT implementation, TTS-based

HVPT further enhances pronunciation development considering the two holistic measures adopted.

This study highlights the potential of TTS technology to provide variable aural input and improve ESL pronunciation training, offering valuable insights for developing effective and accessible language learning resources.

Acknowledgements

I would like to express my most sincere appreciation to my supervisor, Dr. Walcir Cardoso. Throughout this whole process, you have been incredibly supportive and a great mentor. You guided me through every step and offered insightful feedback that has taken my research and academic writing skills to the next level. You are a true inspiration as you have given me the courage and confidence I needed to pursue my academic goals.

Special thanks also extend to Dr. Paul John, who generously dedicated his time to review my proposal and thesis. Your constructive feedback and comments have enriched my work, and I am grateful for your invaluable contribution. I am also thankful to Damian Diaz for his assistance with the statistical analysis.

I am deeply indebted to my beloved husband, Ahmed, who has always been my main source of support and comfort. Despite the miles that separated us, your strong faith in my abilities has carried me through every challenge. Your unconditional love and encouragement have been my guiding light, and I am forever grateful to have you by my side.

To my family, especially my brothers and in-laws, I owe you my deepest gratitude for your ongoing support and prayers. Your belief in my aspirations has been a constant source of strength throughout this journey.

Lastly, but certainly not least, I pay tribute to my late mother, whose love and sacrifices have left an indelible mark on my life. Your guidance and solid faith in my dreams laid the foundation for everything I have achieved. Although you are no longer with me physically, your spirit and presence continue to guide me each day. I cherish the memories we shared and strive to honor your legacy with every step I take. Thank you for everything, Mama. I miss you dearly, and I hope to continue making you proud.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Chapter One

In 1998, my family and I moved from Canada to Kuwait. At just nine years old, I found myself navigating a world where English was considered a foreign language, primarily taught in schools. Our English classes were brief, occurring once a day for 45 minutes, five times a week. This transition posed significant challenges, particularly in how pronunciation was taught in my language classes. Due to the limited class time, pronunciation instruction was often ignored and poorly addressed. Teachers focused primarily on vocabulary and grammar, as well as other linguistic skills such as reading and writing, while neglecting pronunciation practice and conversational skills to cover other essential parts of the syllabus (see Farhat & Dzakiria, 2017 and Jing, 2010 for similar claims in Pakistan and China, respectively).

To make matters more challenging, there were insufficient resources for practicing English pronunciation. We relied solely on the teacher's input, which was often heavily accented and highly unintelligible. We would also listen to traditional audio cassettes that accompanied the required language textbooks, but these cassettes featured a single British variety of English with pre-recorded dialogues, limiting our exposure to other varieties and stifling our ability to explore the language creatively or in a way that reflects real-life experiences.

This lack of resources greatly affected the quality and quantity of language input I received and the output I could produce. The contrast between learning English in Canada and Kuwait was stark. In Canada, I was surrounded by a diverse group of English speakers (including native and non-native speakers), which enriched my language learning experience. In the context of Kuwait, I later realized that the limited input and practice opportunities available to Arabic speakers were detrimental to the development of their pronunciation skills.

To maintain my speaking skills, I dedicated countless hours outside the classroom to self-study. I watched TV shows and movies with English subtitles, frequently pausing to read and repeat the dialogue. When personal computers became prevalent in the early 2000s, I started seeking for programs that offered precise and practical language practice, such as "Talking Dictionary" applications, although these were limited to isolated words and a single variety of English.

These challenges fueled my passion for pursuing an education degree and becoming an English teacher, with the aim of enhancing second language acquisition (SLA) among Arabic speakers. However, I soon encountered a significant limitation: lack of time for pronunciation instruction. As a teacher, I could only allocate a small portion of class time to pronunciation due to a syllabus filled with other essential skills like reading and writing and subskills such as grammar and vocabulary. Consequently, I often had to rush through the basics of pronunciation without knowing if the students had fully acquired the target features.

This experience highlighted a common challenge in English as a Second/Foreign Language (ESL/EFL) settings: limited classroom time (Collins & Muñoz, 2016; Moghaddam et al., 2012; Morin, 2007). Such a constraint deprives students of sufficient, high-quality aural input and the opportunity to produce enough output – both of which are essential for effective learning (e.g., Gass & Varonis, 1994). Pronunciation learning, which requires extensive input and output practice (Cardoso, 2018; Everly, 2019), is particularly vulnerable to these time constraints. Such limitations further limit students' exposure to different speech sounds and accents, causing them to depend extensively on their teachers' input (Bione et al., 2016). This dependency can hinder their communication skills and decrease their confidence in using the target language (Kachru, 1992).

High Variability Phonetic Training

High Variability Phonetic Training (HVPT) emerges as a promising solution to these challenges. HVPT involves exposing learners to a wide range of phonetic contrasts from multiple speakers, accents, and pitch variations. This method enhances learners' aural perception skills, leading to improved pronunciation outcomes (Thomson, 2018). However, implementing HVPT in regular classroom settings remains challenging due to time constraints and limited digital resources (Barriuso & Hayes-Harb, 2018; Thomson, 2018). To address these challenges, we explore below an alternative approach that could help mitigate some of the time and accessibility constraints observed: the pedagogical use of text-to-speech synthesizers.

Text-To-Speech Synthesis

In recent years, the rapid growth of the internet and the emergence of new technologies have radically transformed every aspect of our lives. Thankfully, these technologies are often free and user-friendly, making them highly accessible. As a result, they hold significant potential to address the time constraints faced by language teachers by extending classroom learning and promoting autonomous language practice beyond formal classroom settings (Nunan & Richards, 2015). Among these technologies, Text-To-Speech Synthesizers stand out as a particularly promising pedagogical tool for this purpose.

Text-To-Speech (TTS) technology converts written text into spoken output. It provides language learners with extensive and varied aural input essential for HVPT and with speech quality that resembles human speakers (e.g., Bione & Cardoso, 2020; John & Cardoso, 2017). Research has shown that this technology can be pedagogically beneficial for pronunciation learning (Bione et al., 2016). By enabling learners to practice pronunciation outside the classroom, TTS addresses time constraints and facilitates repeated exposure to diverse accents and speech patterns (Cardoso,

2018). Integrating TTS into HVPT enables learners to access a broader range of linguistic input at their convenience, significantly enhancing their overall language learning experience.

This study: Exploring a TTS-Assisted HVPT Approach for Enhancing L2 Pronunciation

This study explores the pedagogical use of TTS technology, particularly its ability to provide varied linguistic input, a crucial element for effective phonetic training within the HVPT approach to language learning. As such, the study investigates how HVPT can help learners improve their pronunciation abilities and how this training can generalize to their overall communicative competence. More specifically, this study focuses on two types of pronunciation analysis: discrete and holistic. The discrete analysis assesses learners' phonological awareness of regular past tense marking in English (e.g., walk/t/, play/d/, and visit/id/). This analysis follows the framework for pronunciation development proposed by Celce-Murcia et al. (2010), which posits that pronunciation learning follows a developmental hierarchy that begins with phonological awareness, progresses through aural perception and controlled oral production, and culminates in communicative oral skills. The holistic analysis, on the other hand, evaluates two broader aspects of pronunciation: comprehensibility and accentedness, based on insights by Munro and Derwing (1995).

This research contributes to the field of applied linguistics by investigating the practical application of HVPT in real-world ESL classrooms (not in laboratory settings), an area that remains underexplored (Barriuso & Hayes-Harb, 2018; Thomson, 2018). By translating the theoretical potential of HVPT into practical applications within technology-assisted ESL/EFL settings, this study aims to offer valuable insights into pronunciation pedagogy. Addressing the significant issue of limited exposure to diverse language input and inadequate pronunciation training experienced by EFL/ESL learners, the ultimate goal of this study is to contribute to the

field by proposing an alternative method to enhance the quantity and quality of L2 pronunciation instruction.

As per the guidelines for a manuscript-based MA thesis, the next section constitutes “a full submittable draft of a manuscript” that presents the full literature review, methodology, results, and discussion of the abovementioned research.

Chapter Two

English is expanding rapidly, introducing diverse varieties and accents around the globe (Hamm, 2020). Nevertheless, learners of English as a Second/Foreign Language (ESL/EFL) often receive limited exposure to varied forms of language input within traditional learning settings, primarily relying on the teacher's input alone (Bione & Cardoso, 2020). This limitation hinders the development of pronunciation, a fundamental skill for effective oral communication (Fraser, 2000; Levis & Grant, 2003), potentially leading to communication breakdowns and misunderstandings that could impair mutual intelligibility (Kachru, 1992; Sifakis & Sougari, 2005). Therefore, ensuring adequate pronunciation instruction becomes crucial for fostering successful intelligible speech, an essential aspect of communicative competence (Morley, 1991; Prashant, 2018).

Pronunciation acquisition is intricately linked to the richness of input, including both its quantity and quality (Flege, 1999; Moyer, 2009). The literature highlights that optimal learning occurs when learners are exposed to input that is both comprehensible (Krashen, 1985) and acoustically varied (Barcroft & Sommers, 2005). Consequently, ESL/EFL learners must receive substantial exposure to second/foreign-language (L2) input to facilitate acquisition (Bione et al., 2016; Bione & Cardoso, 2020).

One way of enriching linguistic input is via High Variability Phonetic Training (HVPT). This pedagogical approach involves exposing learners to diverse speech sounds, incorporating acoustic properties such as different accents, voice types, and pitch variations (Levis, 2016). Research indicates that HVPT can potentially enhance phonetic perception skills (Bradlow & Bent, 2008), leading to subsequent improvements in pronunciation proficiency (Logan et al., 1991).

While HVPT has consistently demonstrated improvements in L2 pronunciation skills, its application has primarily been studied in controlled lab settings (Barriuso & Hayes-Harb, 2018), raising concerns about its practical validity and applicability in regular learning environments (Thomson, 2018). Moreover, implementing HVPT in classrooms faces challenges due to the limited time allocated for language instruction, as highlighted by Collins and Muñoz (2016). Such time constraint deprives learners of adequate pronunciation training (e.g., using HVPT techniques), resulting in insufficient exposure to linguistic input.

To tackle these challenges, this study explores the integration of Text-To-Speech (TTS) technology as a practical solution for HVPT. TTS, which converts written text into spoken output, provides learners with varied linguistic input essential for HVPT (Bione & Cardoso, 2020). While several studies reveal TTS's potential in enhancing L2 pronunciation and learner autonomy (e.g., Cardoso, 2018, 2022; Kiliçkaya, 2008), its application within HVPT remains underexplored.

This study aims to examine the impact of HVPT afforded by TTS on ESL learners' pronunciation abilities and the generalizability of this training to their overall communicative competence. The assessment includes two measures: a discrete phonological analysis and a holistic evaluation of pronunciation. The discrete analysis assesses phonological awareness of Regular Past Tense marking in English (i.e., *past -ed*; past-inflected forms of *-ed* as found in walk/t/, play/d/ and visit/id/). The analysis in based on Celce-Murcia et al.'s (2010) framework for pronunciation development, which presupposes that phonological acquisition progresses hierarchically from phonological awareness through aural perception and controlled oral production, ultimately leading to communicative oral skills. The holistic analysis, on the other hand, examines two broader aspects of pronunciation: comprehensibility and accentedness, following insights from Munro and Derwing (1995).

The present study seeks to bridge the gap between the theoretical promise of HVPT and its practical application in technology-assisted ESL settings. By utilizing TTS, we aim to contribute valuable insights to pronunciation pedagogy, addressing issues such as limited classroom time and lack of practice. Through TTS, learners can access a broader range of linguistic input at their convenience, optimizing their language experiences. The study anticipates positive outcomes in language learning, foreseeing improvements in pronunciation proficiency within an increasingly interconnected world.

Background

Input in Second Language Acquisition

Second language acquisition (SLA) is a complex process that relies heavily on learners' exposure to language input. 'Input' in language learning goes beyond mere exposure to raw linguistic data; it encompasses meaningful interactions that learners can understand, as reflected in Krashen's (1985) Input Hypothesis. This hypothesis posits that effective L2 learning requires comprehensible input slightly above the learner's current level of competence (i.e., i+1). It emphasizes that mere language exposure is insufficient, as the input must be comprehensible to facilitate learning. As a result, educators and researchers strive for linguistically rich environments that offer not only the quantity but also the quality of input necessary for successful acquisition.

Empirical studies consistently recognize the critical role of both quantity and quality of input for L2 mastery. Flege (1995, 1999) demonstrated a direct relationship between the amount of input received from native speakers and the degree of foreign accent in L2 speech. Moyer (2009) extended this perspective by showing that high-quality interactions with native speakers can lead to advanced fluency and improved pronunciation. A key aspect in this dynamic process is 'input richness' (Barcroft & Sommers, 2005), which, when both comprehensible and acoustically varied,

is likely to enhance learning, as noted by Bione et al. (2016). This involves exposure to diverse accents, speaking rates, and intonation patterns, facilitating a more nuanced language understanding and the ability to generalize learned material to new contexts. Iverson and Evans (2009) and Thomson (2012) found that exposure to ample and varied input improved the perception and production of non-native phonemes in L2 learners. These studies collectively confirm that, for input to be truly beneficial for language acquisition, it must be comprehensible to the learner and sufficiently diverse to reflect the complex nature of the target language.

High Variability Phonetic Training (HVPT) as Enhanced Input

As rich input forms the basis of SLA, HVPT emerges as a compelling pedagogical approach designed to optimize linguistic input for language learners. HVPT exposes learners to a wide range of target phonetic contrasts characterized by varied acoustic properties from multiple talkers, accents, types of voices, and pitch variations (Ingvalson et al., 2014; Levis, 2016). This approach integrates both *inter-speaker* variation—differences among speakers such as regional accents and individual speech styles—and *intra-speaker* variation—changes within a single speaker's speech influenced by context or emotional state (Honeybone, 2011). By exposing learners to this broad range of speech sounds rather than a limited set, HVPT researchers hypothesize that this diversity enhances phonetic perception skills (Bradlow & Bent, 2008), thereby improving pronunciation proficiency (Logan et al., 1991).

To achieve variability in the input, HVPT employs a perceptual training method that utilizes audio recordings from different speakers to present stimuli to learners. During training, learners identify specific sounds within these stimuli and receive immediate feedback on their responses. McCandliss et al. (2002) explain that providing immediate feedback on response correctness significantly enhances learning by directing learners' attention to the crucial acoustic

properties for forming L2 phonemic categories, such as using formants to identify specific vowel sounds (Thomson & Derwing, 2016). Learners are trained to focus on these properties while disregarding other variable properties that differ among speakers. This aspect is essential because, as Schmidt (1990) suggests, 'noticing' is necessary for converting input into 'intake,' where linguistic information is not only received but also processed and integrated into the learner's interlanguage system.

The robust evidence supporting the efficacy of HVPT reveals consistent improvements in pronunciation skills among learners undergoing this type of training. In a study by Lively et al. (1993), Japanese learners were trained to discriminate the English /l/-/r/ contrast using either multiple talkers (HVPT) or a single talker (non-HVPT). While both groups improved during training, only the HVPT participants were able to generalize their knowledge to a novel talker. This generalization ability, replicated in studies involving American English learners of Mandarin tone (Perrachione et al., 2011) and Cantonese learners of English /e/-/æ/ sounds (Wong, 2014), signifies a crucial progression for L2 learners—from simple imitation to a more profound cognitive understanding and processing of phonetic input.

These findings align with Bradlow et al. (1997), who demonstrated that Japanese learners of English significantly improved their perception of English /r/ and /l/ sounds following HVPT, and these improvements generalized to novel items spoken by new talkers. Following perceptual training, they also observed improvement in the Japanese trainees' production of /r/–/l/, highlighting the intricate relationship between speech perception and production. Long-term retention studies by Bradlow et al. (1999) indicated that the enhanced performance levels in both perception and production persisted three months post-training, indicating the enduring impact of HVPT on phonological perception and production. Moreover, Thomson's (2018) synthesis of 32

studies implementing HVPT in teaching L2 sound perception and production confirmed significant and lasting improvements in phonetic discrimination and production abilities.

Despite its potential benefits, HVPT remains relatively uncommon within the mainstream practices of L2 teachers. Thomson (2018) suggests that this lack of adoption likely stems from the complexities associated with integrating HVPT into traditional classroom settings. Furthermore, while HVPT has shown promising results in improving perception and production of target L2 phonemes, further research is essential to understand its potential in enhancing the comprehensibility and intelligibility of novel accents and speech varieties. This exploration is crucial for enabling L2 learners to communicate effectively across diverse speech communities.

Pronunciation Assessment

Understanding how learners' pronunciation skills develop is crucial in evaluating the effectiveness of HVPT. Pronunciation assessment falls into two broad categories: discrete phonological analysis and holistic pronunciation analysis.

Discrete Phonological Analysis: English Simple Past Tense

This approach evaluates specific phonological units, such as segments (consonants and vowels), stress, rhythm, and intonation, which are considered fundamental for pronunciation learning. Celce-Murcia et al. (2010) outline the stages of pronunciation development, emphasizing phonological awareness, aural perception, and oral production as critical stages.

Phonological awareness. This stage plays a crucial role in language acquisition, where learners consciously recognize the phonological structure of the target language. This awareness encompasses various aspects of language sounds, such as syllables (e.g., onsets, codas) and phonemes, which play a crucial role in both processing and producing language. As Layton et al.

(1998) outlined, the development of phonological awareness evolves from simple sound recognition to the ability to produce it (see also Anthony & Francis, 2005 for a similar claim).

The role of 'noticing' in language learning is integral to phonological awareness. Schmidt's (1990) 'noticing hypothesis' asserts that the conscious recognition of language elements is crucial for developing language proficiency. Schmidt and Frota (1986) stress the importance of being aware of the disparities between the input and the learner's existing interlanguage system for effective learning. According to Linebaugh and Roche (2015), pronunciation training becomes instrumental in elevating this awareness, aligning with the claims of Celce-Murcia et al. (2010).

Aural perception. This stage centers on the learner's ability to recognize and distinguish sounds (Soler-Urzúa,2011). The development of perceptual learning in L2 depends heavily on the quantity and quality of input from native L2 speakers (Flege & Liu, 2001; Jia & Aaronson, 2003). A vast body of literature indicates that exposure to diverse aural phonetic input significantly enhances perceptual knowledge in L2 learners (Bradlow & Bent, 2008; Shin & Iverson, 2013; Wang et al., 1999; Pruitt et al., 2006), reinforcing the value of HVPT in language pedagogy.

Furthermore, Flege's (1988) Speech Learning Model (SLM) suggests that adequate exposure to input facilitates improved L2 sound perception, a crucial element for achieving target-like production (Flege, 1995). This model has inspired extensive research, indicating that adults can enhance their perception of L2 sounds through proper training (Herd et al., 2013; Iverson & Evans, 2009). The symbiotic relationship between perception and production is further underscored by Sakai and Moorman (2018), who emphasize that training in L2 sound perception positively influences productive abilities. This supports the assumption that perception precedes oral production (Cardoso, 2011; Celce-Murcia et al., 2010; Flege, 1995), highlighting the necessity of developing aural perception as a basis for enhancing production skills.

Oral production. This final stage concerns the learner's ability to appropriately pronounce the target sound (Soler-Urzúa, 2011), reflecting the successful integration of phonological awareness and aural perception skills. According to SLM, speech production and perception converge, indicating that learners' output mirrors their perception (Flege, 1995). Hence, accurate perception is critical to developing accurate pronunciation (Crosby, 2020). This interplay highlights the importance of enriching aural input in L2, fostering improvements in oral production through exposure to varied speech (Soler-Urzúa, 2011; Liakin et al., 2017a; Flege, 1991).

Developing oral production skills in L2 learners extends beyond articulating individual phonemes to include aspects of prosody and fluency, such as intonation, rhythm, connected speech, and voice quality settings. According to Celce-Murcia et al. (2010), a comprehensive pronunciation training approach should address these elements, as they contribute to both speech intelligibility and naturalness, necessary for effective communication (Derwing & Munro, 2005).

In this study, we prioritize phonological awareness because of its essential role in the early stages of pronunciation development and its compatibility with TTS technology's emphasis on auditory input, as outlined in the introduction. TTS technology enriches pronunciation training by providing varied auditory input and diverse phonetic contrasts, which can potentially improve phonological awareness (see De Araújo Gomes et al., 2018 for similar claims). By focusing on phonological awareness, this study evaluates the impact of TTS technology on learners' initial pronunciation development, thereby setting the groundwork for future research into aural perception and oral production.

A notable feature in the field of phonological development, which is also the focus of this study, is the pronunciation of the English Regular Past Tense (RPT), formed by adding the morpheme *-ed* to the infinitive form of the verb (e.g., "cook" becomes "cooked"). This

morphophonemic phenomenon is characterized by three allomorphic variants: /t/, /d/, /id/ (as in 'asked' [askt], 'opened' [opend], and 'wanted' [wantid] respectively). While the pronunciation of past *-ed* is governed by well-defined phonological rules, seemingly rendering it easy to teach (Royani & Rahmi, 2019), it remains a persistent challenge for ESL learners to acquire (e.g., Cardoso, 2018; Collins et al., 2009).

Several factors can negatively impact RPT development in English. One such factor is the influence of L1 structures on L2 acquisition (Barros, 2003; Frese, 2006; Zimmer et al., 2009), impacting non-native speakers' pronunciation (Cook, 1992; Avery & Ehrlich, 1992). For instance, in L1 Arabic, coda clusters like /gd/ and /pt/, which result from RPT inflection, are not allowed. Consequently, Arabic speakers tend to insert an epenthetic vowel in illicit forms such as be/gid/ for 'begged' (Kharma & Hajjaj, 1997; Salim & Mohammed, 2023). Another factor is the impact of orthography on L2 phonology, where the congruency between grapheme-to-phoneme correspondence in the learner's L1 and L2 may impede the acquisition of L2 phonological contrasts (Jackson & Cardoso, 2022). For example, a single orthographic form *-ed* can represent three distinct pronunciations in English: /t, d, id/, depending on the final sound of the root verb. Finally, limited exposure to this morpho-phonological feature (i.e., English RPT) in classroom settings exacerbates these challenges, further hindering non-native speakers' mastery of RPT development in English (Collins et al., 2009).

Holistic Pronunciation Analysis: Comprehensibility and Accentedness

Holistic measures to assess pronunciation extend beyond mere discrete phonological analyses, offering a broader view on speech intelligibility and communicative effectiveness. Munro and Derwing (1995) proposed three key constructs for holistic assessment:

comprehensibility, intelligibility, and accentedness, each playing a vital role in evaluating the oral production of L2 speakers.

Comprehensibility refers to the listeners' perception of how well they understand the meaning conveyed in an utterance, while intelligibility assesses their ability to accurately understand the speaker's intended message. Accentedness, on the other hand, measures the degree to which an L2 speaker's pronunciation deviates from fluent or native speakers' norms. Although it may not necessarily hinder intelligibility or comprehensibility, accentedness acts as an indicator of non-nativeness in speech (Munro & Derwing, 1995).

Research involving pronunciation training has yielded positive results in improving these holistic measures. For instance, Couper (2006) demonstrated long-term improvements in accent features, indicating the potential for permanent alteration in learners' phonological representations. Kangatharan et al. (2021) found that high-variability training can enhance L2 learners' speech production, leading to greater intelligibility. Similarly, Bradlow et al. (1997) showed that perception training among Japanese speakers resulted in more intelligible production of English /r/-/l/ contrasts. Studies also suggest that exposure to phonetic variability leads to more robust representations and broader lexical categories (Lively et al., 1993; Rost & McMurray, 2009).

The current study primarily focuses on comprehensibility and accentedness as key measures for holistic pronunciation. Aligned with Kennedy and Trofimovich's (2019) findings, this choice recognizes that comprehensibility closely correlates with intelligibility in L2 speech research. Comprehensibility ratings, while not directly equivalent to intelligibility, often reflect similar patterns in listener comprehension. By prioritizing these measures, this study seeks to offer insights into practical pronunciation training and the development of communicative skills in diverse linguistic contexts.

A TTS-assisted HVPT approach to learning L2 pronunciation

Integrating technology into L2 learning, particularly through Computer-Assisted Language Learning (CALL), is recognized as an effective method to improve pedagogical practices and tackle learners' challenges (Reinders & White, 2010), offering opportunities for authentic language use beyond traditional classrooms (Richards, 2015). Among these technologies, TTS has emerged as a promising tool in L2 pedagogy, particularly for enhancing pronunciation instruction (Cardoso, 2018; 2022; Kiliçkaya, 2008; Soler-Urzúa, 2011). By allowing learners to practice pronunciation outside the classroom, TTS addresses time constraints and enables repeated exposure to diverse accents and speech patterns (Cardoso, 2018; Ekşi & Yeşilçınar, 2016; González, 2007; Kim, 2018; Moon, 2012).

When evaluating the implementation of TTS technology, crucial factors to consider are its pedagogical efficacy and its potential to enhance learning outcomes. Previous research indicates that TTS can significantly benefit language learning, particularly in improving L2 learners' pronunciation (Bione et al., 2016; Bione & Cardoso, 2020; Liakin et al., 2017b; Soler-Urzúa, 2011). For instance, Liakin et al. (2017b) found that the pedagogical use of a popular TTS application improved their participants' ability to produce L2 French liaison in comparison with a control group. Their findings suggest that TTS can serve as a tool in language learning settings, providing benefits such as extended access to language practice and enhanced learner control.

Further research advocates integrating HVPT into computer-assisted pronunciation training to reinforce classroom learning through additional practice (Thomson, 2018). Supporting this, Bione and Cardoso (2020) highlight the effectiveness of employing TTS in language learning for creating an HVPT environment where various voice characteristics (gender, age, accent, pitch) can be selected and manipulated. In line with these recommendations, this study examines the

pedagogical affordances of TTS by offering varied linguistic input—a crucial element for effective phonetic training within HVPT—and how it can help learners improve L2 pronunciation.

Current Study

Research findings on HVPT have demonstrated its potential in controlled settings. However, the practical implementation of HVPT in L2 classrooms remains limited (Barriuso & Hayes-Harb, 2018; Thomson, 2018). This gap between research and classroom practice underscores the need for innovative approaches to address the challenges faced by L2 learners in perceiving and producing non-native speech. The urgency of this issue is further intensified by the fact that the limited instructional time devoted to pronunciation in ESL settings often proves inadequate to meet learners' needs and aspirations (Barcomb & Cardoso, 2020).

This study explores TTS as a viable solution (Bione & Cardoso, 2020), providing learners with exposure to diverse English accents and varieties via its HVPT approach to L2 pronunciation pedagogy. The focus is to address both discrete and holistic aspects of pronunciation. This includes phonological awareness of the regular past tense *-ed* morpheme, as well as the impact of the treatment on comprehensibility and accentedness. To achieve these objectives, the study will explore the following two research questions:

1) Discrete Analysis of Pronunciation: Can the integration of HVPT through TTS technology improve ESL learners' pronunciation of English past *-ed* in terms of Phonological Awareness (i.e., awareness to the morphophonemic variation observed in past *-ed* inflection)?
2) Holistic Analysis of Pronunciation: Can the integration of HVPT through TTS technology voices improve ESL learners' pronunciation in terms of:
    a) Comprehensibility (a rater's perception of how easily an utterance can be understood)

b) Accentedness (degree to which an L2 speaker's pronunciation deviates from fluent or native speakers' norms)

Methods

Participants

Learners

30 adult university-level ESL learners (9 males and 21 females) living in Kuwait, aged 18 to 30, were selected for this study. They were divided equally into two groups: the treatment group, and the control group, each consisting of 15 individuals. The recruitment process involved a call for participation on social media platforms and outreach to English teachers within the local academic community. The participants were informed that the study is part of the researcher's master's degree requirements and aims to investigate development of English pronunciation among native Arabic speakers.

Demographic information, including age, gender, English proficiency level, and language background were collected from all participants. Participants were first-language (L1) speakers of Arabic, with low to intermediate proficiency in English and without any history of speech or hearing disorders. English proficiency was determined through multiple criteria, including pre-test results (participants scoring 50% or less were included), self-assessment responses from a background questionnaire, and the researcher's assessment of their language skills during pre-study interviews.

Raters

To assess the participants' oral production (holistic measures), a separate group of 11 raters was recruited as listeners to evaluate the non-native speech samples. This group included experienced ESL teachers and advanced ESL speakers. The expertise and linguistic backgrounds

of these raters were essential for enhancing the quality of the assessment process, given the global use of English as a lingua franca.

Design

This study employed a mixed-method research design with pre-test and post-test measurements, involving both control and treatment groups. The primary independent variable under examination was the application of HVPT through TTS technology. The study investigated the impact of this technology-enhanced approach to pronunciation training on a range of dependent variables, including participants' improvements in phonological awareness skills related to the English past *-ed* morpheme, as well as the changes in holistic pronunciation aspects: comprehensibility and accentedness.

As illustrated in Figure 1, the design of the study followed a structured series of steps. Initially, participants completed demographic surveys to collect essential information, such as age, linguistic knowledge, and familiarity with technology. Then, both groups completed a pre-test to evaluate their phonological awareness skills related to the English past *-ed* morphophonology, as well as their overall pronunciation proficiency in terms of comprehensibility and accentedness.

Following the pre-test, both groups watched an instructional video that introduced them to the *written* formation of regular past tense inflection (i.e., the addition of the orthographic *-ed*). Notably, the video did not provide any information about its pronunciation, which was the target of the study and was explored by the participants on their own. This approach is underpinned by Data-Driven Learning (DDL), a learner centred pedagogy (Johns, 1991). In the subsequent training activities, participants actively engaged with language data and used DDL techniques to independently discover how past *-ed* is pronounced in different linguistic contexts. As such, this study aimed to empower participants to actively explore these variations and develop a deeper

understanding of how pronunciation works, with some guidance to emulate the types of activities L2 learners engage in (e.g., when completing homework assignments). Additionally, participants watched a demonstration on how to use Speechify, the target TTS tool (see Materials section for more details).

Both groups were scheduled to undergo a learning phase of four weeks. The control group exclusively received input from a single standard (but synthesized) native English accent, while the treatment group was exposed to a variety of English accents delivered through five different voices generated by the TTS. The decision to use five different voices is grounded in both empirical findings and established HVPT practices. Empirical evidence, as found by Zhang et al. (2021) in their systematic review and meta-analysis, suggests that the multi-talker group typically receives input from an average of five talkers across various studies. Furthermore, the choice to incorporate five different talkers is in line with established studies in the field, including research conducted by Logan et al. (1991), Lively et al. (1993, 1994), Shinohara and Iverson (2013), Kartushina and Martin (2018), Kingston (2003), and Sadakata and McQueen (2014).

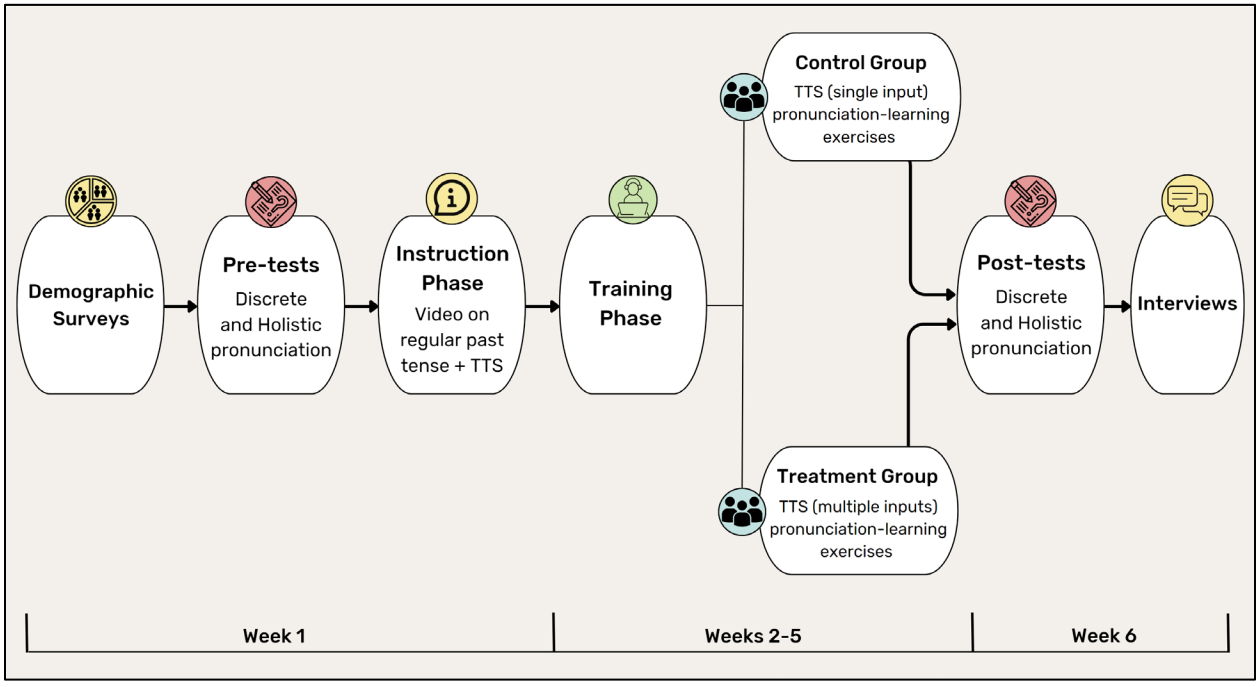
This study leverages TTS technology to introduce both inter-speaker and intra-speaker variation in pronunciation training, making both types of variation relevant to the current investigation. However, the study primarily employs inter-speaker variation by incorporating multiple TTS voices with different speech patterns, including North American, British, and Australian accents. While integrating intra-speaker variation—like modifications in speed within a single TTS voice—could be beneficial, this study prioritizes variation across different voices. Focusing on this type of variation supports learners' adaptation to various English pronunciations, reflecting the global nature of the language, particularly in regions such as the Middle East (e.g., Kuwait), where exposure to a wide range of English varieties is common. This approach also aligns

with the English as a Lingua Franca approach to L2 education (Jenkins, 2000), which recognizes the importance of preparing learners to communicate effectively with speakers from diverse linguistic and cultural backgrounds.

Throughout the duration of the experiment, participants engaged in twelve 30-minute self-paced training sessions, allowing them the freedom to choose when and where they complete pronunciation exercises. Following the training period, participants took post-tests, wherein all test items were shuffled randomly to reduce any potential testing effects. Similar to the pre-tests, these post-tests assessed both the discrete and holistic aspects of pronunciation.

Figure 1

Study Design



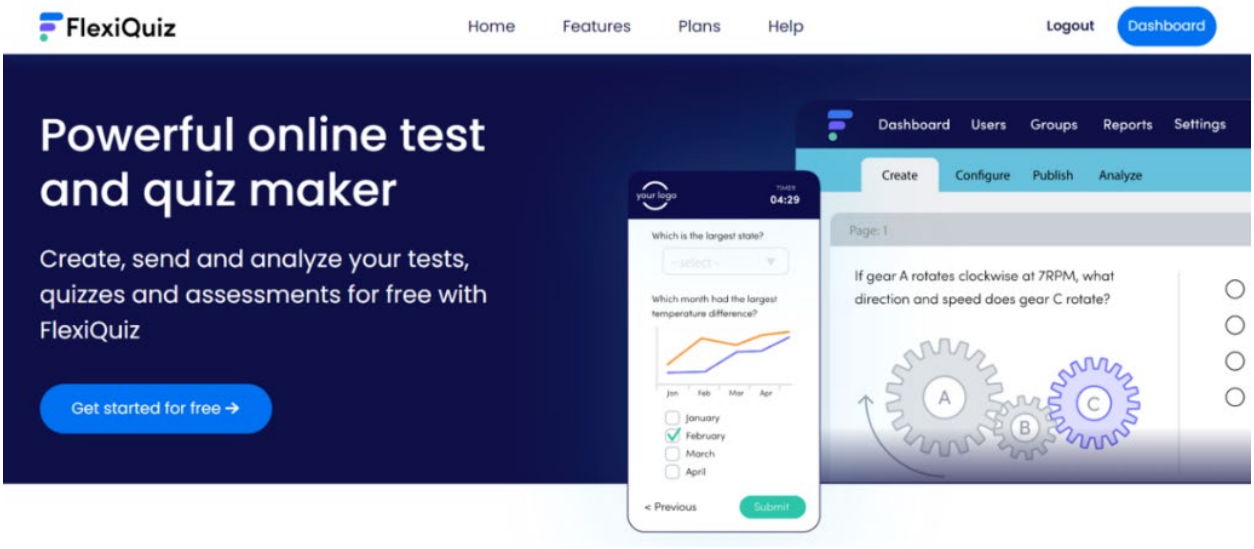
Procedures and Instruments

The entire study was conducted remotely. After verifying the quality of the microphones and recordings to ensure their suitability, the researcher met the participants via the video-conferencing service Zoom for the initial demographic surveys, as well as the discrete and holistic

assessments during both pre-tests and post-tests. For the training phase, we utilized FlexiQuiz (Figure 2), an online platform that enables users to create customized quizzes with various question types, including multiple-choice, file uploads, and short answer questions.

Figure 2

FlexiQuiz Homepage

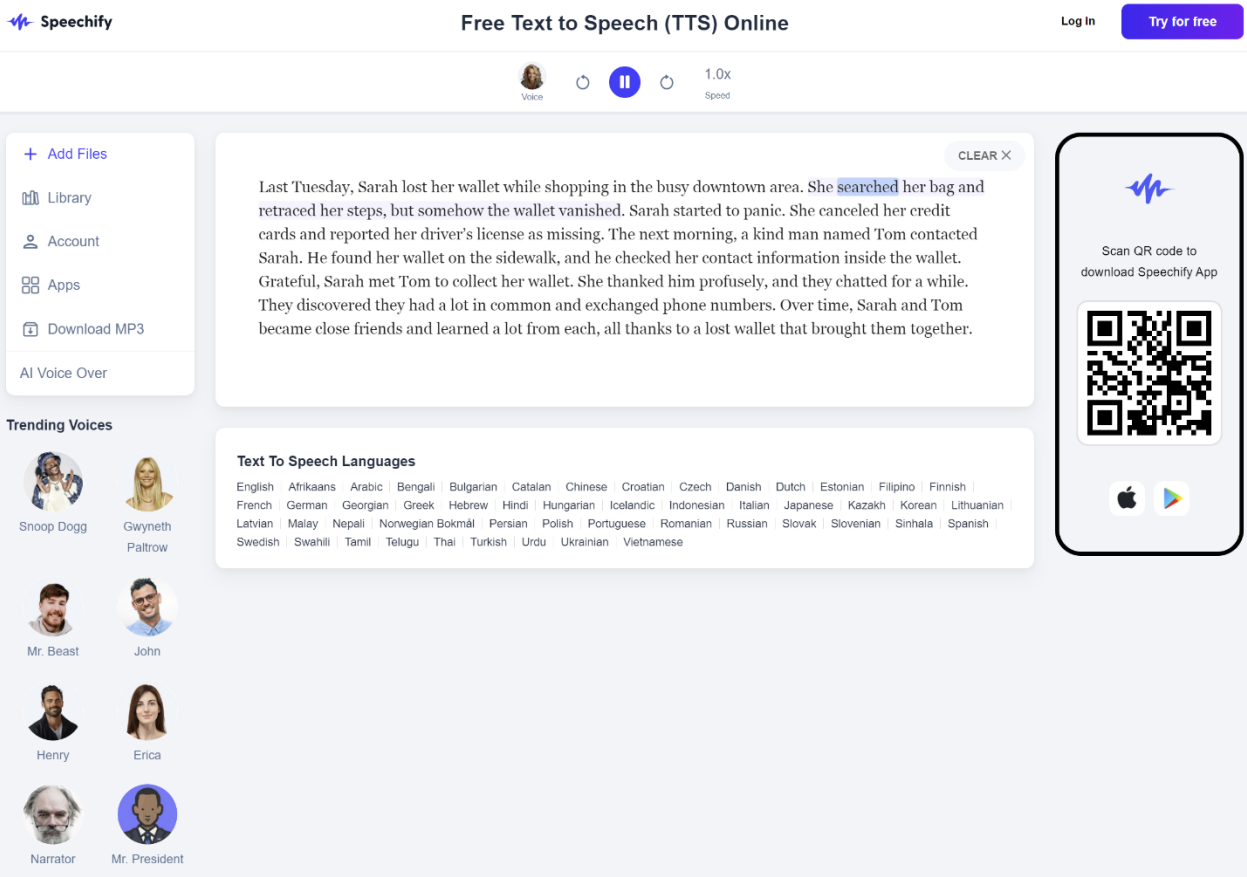


Speechify: The Target TTS Tool

Speechify is a TTS tool that meets three important criteria for this study: accessibility, diversity of voices, and voice quality. First, Speechify offers a straightforward user experience across various platforms, including Web, Chrome extension, Mac, Android, and iOS. As shown in Figure 3, it features an intuitive interface and seamless functionality, making it user-friendly for individuals with different technical backgrounds. Second, aligning with the study's goal of exposing participants to variability, Speechify provides a diverse range of English voice options, encompassing variations in gender, age, speech styles, dialects, and accents. Finally, Speechify produces high-quality, natural-sounding voices with authentic intonation and emotion, and provides a free basic plan that includes unlimited use of voices.

Figure 3

Speechify TTS Interface: Free Version



Consent Form

The researcher met with potential participants via Zoom and provided them with a digital copy of the consent form (Appendix A-1). During the meeting, the researcher explained the consent form and clarified the study's scope to the participants. Those who agreed to take part in the study were asked to sign a digital copy of the consent form and return it to the researcher.

Demographic Surveys

Participants completed a demographic survey (Appendix B). The survey included questions about age, gender, linguistic background, and knowledge of technology and familiarity with TTS.

Pre-tests

After completing the demographic surveys, participants completed the pre-tests (approximately 30 minutes) to evaluate their knowledge in two key areas: the discrete aspects of past *-ed* (i.e., phonological awareness) and their holistic pronunciation in English (i.e., comprehensibility and accentedness). To ensure comprehensive data collection and explore various aspects of the same phenomenon, we used a total of four instruments. As illustrated in Table 1 and detailed below, two instruments assessed knowledge of past *-ed* morpheme with a focus on phonological awareness, while the other two instruments were employed for holistic measures (Speech Evaluation).

Table 1

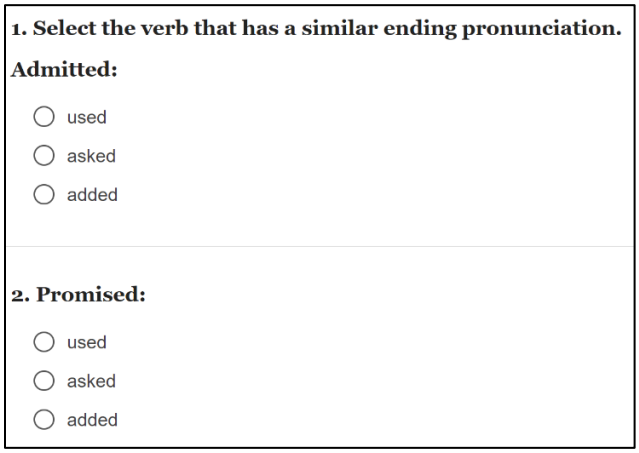
Assessment Instruments for Data Collection

Instrument Type	Focus of Assessment	Assigned to	Number of instruments
Phonological awareness	Discrete Analysis	Learners	2
Speech Evaluation	Holistic Analysis	Raters	2
		Total:	**4**

Phonological Awareness. In the first test, participants answered open-ended questions to assess their understanding of how past *-ed* is pronounced (e.g., "Do you think past tense *-ed* is pronounced differently for different verbs?"; see Appendix C). The second test, as shown in Figure 4, required participants to pair an inflected verb (e.g., *needed*) with the way they believe the inflected endings are pronounced (i.e., *added* in this case). This involved drawing connections with established and highly-frequent verb forms, like 'used /d/', 'added /id/', and 'asked /t/' (see Appendix D for the complete list of items).

Figure 4

Phonological Awareness Test 2: A Sample



Holistic Assessment. Finally, participants were evaluated based on comprehensibility and accentedness, both of which are often assessed by raters or listeners. This assessment aimed to elicit spontaneous speech, facilitating a natural and comprehensive evaluation of their pronunciation skills. To maintain a counterbalanced design and minimize the potential influence of question order, participants were divided into two groups, Group A and Group B. In the pre-test phase, Group A was asked to describe their activities from the previous summer, while Group B was prompted to describe their last birthday. For the post-test, the questions were reversed between the two groups (see Appendix E). Participants' responses were audio-recorded during Zoom meetings.

Upon collecting all recordings, each file was uniquely identified to preserve the anonymity of the participant (e.g., P01, P02). The researcher then selected a 20-30-second excerpt from each participant's recording and distributed these excerpts to the assigned raters. After signing the consent form for participation (Appendix A-2), the raters assessed each participant's pronunciation using a 9-point Likert scale in terms of comprehensibility and accentedness, following Munro and Derwing (1995, 2020) (see Appendix F):

- Comprehensibility: Scale ranging from “Extremely Easy to Understand” to “Extremely Hard to Understand”;
- Accentedness: Scale ranging from “Strongly accented” to “Not accented at all.”

All raters completed an evaluation form for each participant and sent their reports to the researcher via email.

Materials

Instructional video

All participants watched a brief (10-minute) instructional video (see Appendix G) that explained the written formation of regular past tense inflection. Notably, the video deliberately avoided addressing the pronunciation of past *-ed* allomorphy. Instead, it emphasized that there are different pronunciations for past-*ed*, assigning participants the task of identifying these variations, and understanding the contexts in which they are employed. Additionally, the video provided guidance on utilizing Speechify as a TTS tool for pronunciation practice. This step aimed to ensure that all participants had a better understanding of the target feature (e.g., how it functions in English), encouraged them to independently discover the nuances of *-ed* allomorphy, and offered them with instructions on using the target TTS tool effectively.

Training

Over the course of one month, participants engaged in a series of 10 self-paced pronunciation learning sessions, allowing them the flexibility to complete these tasks remotely (e.g., from the convenience of their personal space at a time of their choosing). Each session was designed to take approximately 30 minutes, similar to a typical homework assignment. Participants utilized Speechify in combination with the FlexQuiz platform to complete exercises such as fill-in-the-blanks and multiple-choice tasks. During these activities, the control group used a single

input voice (i.e., Jaime, North American female) for all 10 sessions. In contrast, the treatment group used five different voices including American and British English (3 females, 2 males), providing exposure to a range of English language varieties. Table 2 illustrates how the texts were distributed across the sessions, with each text being read aloud by a specific designated voice.

Table 2

The Distribution of Voices Across Sessions

Treatment Group Input			**Control Group Input**		
Voice Name	**Voice Features**	**Sessions**	**Voice Name**	**Voice Features**	**Sessions**
Jamie	North American Female	1, 6	*Jamie*	North American Female	All sessions
Micheal	British Male	2, 7			
Sydney	Australian Female	3, 8			
Nate	North American Male	4, 9			
Stephanie	British Female	5, 10			

Note. See Training Materials: Activity #1 as an example in Appendix H

Following HVPT protocols, the underlying hypothesis for these training sessions is that exposing participants to various English accents using multiple voices from a TTS tool will enhance their ability to acquire phonological knowledge regarding past *-ed* allomorphy.

This study explored how pronunciation knowledge develops in a TTS-based HVPT learning setting. As such, each training session included three TTS-assisted activities:

Activity 1. In the first part of each session, participants engaged in a listening activity involving fill-in blank questions designed to enhance their awareness and recognition of past *-ed* within context. This exercise involved working with a short story that included a balanced distribution of the three *-ed* allomorphs (four of each) and two to three distractors (refer to Appendix H). Participants were instructed to copy and paste each story text into Speechify, select

the designated voice, and then listen to the short story without reading the original text. While listening, they were expected to complete the blanks with the appropriate regular past tense verbs. In line with HVPT principles, immediate feedback was incorporated into the exercise. If a participant submitted an incorrect answer, the system identified the mistake and provided feedback before proceeding to the next question. Upon a correct response, the system immediately proceeded to the next item.

Activity 2. The second activity consisted of a list of 120 regular simple past tense verbs, all extracted from the 600 most frequently used English verbs (see Appendix I). In each session, 12 verbs were included, with four representing each allomorph. Participants were instructed to copy and paste these verbs into Speechify to listen to their synthesized pronunciations. Subsequently, they were tasked with matching the verbs to the perceived pronunciations of their inflected endings by associating them with established exemplars such as “used” for /d/, “added” for /id/, and “asked” for /t/. As was the case for the first activity, immediate feedback was provided for participants’ responses.

Activity 3. The third activity involved a listening comprehension task focusing on past tense events (see Appendix J). This activity aimed to address a larger discourse that incorporated past *-ed* inflection. Each training session included a story in the past tense (10 in total), followed by five comprehension questions in a multiple-choice format. Participants were instructed to copy and paste a text into Speechify, select the designated voices, and listen to the text without viewing the original content. Immediate feedback was also provided within this activity.

Post-tests

After completing the training phase, the post-tests were administered, lasting approximately 30 minutes. Like the pre-tests, these assessments evaluated the participants’

progress in both discrete and holistic aspects of pronunciation. These post-tests employed the same items as the pre-tests but presented in a randomized order.

Interviews

In the final phase, brief semi-structured interviews were conducted via Zoom. Participants responded to a series of questions regarding their experience with pronunciation training, their perceptions of Speechify and its speech capabilities, strategies they employed to learn about past *-ed*, and any challenges they encountered throughout the study (see Appendix K). These interview questions were adapted from Venkatesh and Davis' (2000) Technology Acceptance Model (TAM) survey, allowing us to explore how participants perceive and accept TTS as a pedagogical tool for improving English pronunciation.

Analysis

Given the ordinal nature of our data and the relatively small sample sizes in each group, we employed non-parametric statistical methods for our analysis. Specifically, we used the Wilcoxon Signed-Rank Test to evaluate changes within each group from pre-test to post-test (within-group analysis) and the Mann-Whitney U Test to compare changes between the treatment and control groups (between-group analysis). To address potential Type I errors from multiple comparisons, we applied the Holm-Bonferroni method for p-value adjustment (corrected p-values reported).

For the discrete analysis, participants' awareness of past *-ed* morphophonemics was measured on a scale from 1 to 4. A score of 1 indicated no knowledge of past *-ed* pronunciation. Scores of 2 or 3 reflected varying levels of partial knowledge: a score of 2 was assigned to participants who demonstrated knowledge of only one *-ed* form, while those who identified two *-ed* allomorphs received a score of 3. Participants who showed full awareness by articulating all

three distinct *-ed* allomorphs were assigned a score of 4. In the second awareness test, participants were evaluated out of a total of 21 points, with seven items assigned to each *-ed* allomorph.

For the holistic measures, participants' speeches were evaluated by eleven experienced ESL teachers using a 9-point Likert scale. In evaluating comprehensibility, a score of 1 indicated that the speech was highly comprehensible, while a score of 9 indicated that it was extremely difficult or impossible to understand. Regarding accentedness, a score of 1 indicated that the rater found the target utterance not accented at all, while a score of 9 indicated an accent that was highly accented. For both measures, lower scores indicate more favorable outcomes, with speech being easier to understand and not accented at all, respectively.

Lastly, the interviews were transcribed and coded to provide additional insights into participants' experiences with TTS and its efficacy as a pedagogical tool, offering a better understanding on the study's findings where applicable.

Results

This study investigated the use of TTS technology to enhance ESL learners' pronunciation through HVPT, comparing the effects of TTS technology with a single native accent versus multiple English accents. We examined both morphophonemic features (to which we also refer as segmental features) and holistic aspects of pronunciation to address two research questions: First, can the integration of HVPT through TTS technology improve ESL learners' phonological awareness of the English past tense *-ed* morpheme (discrete analysis)? Second, can the integration of HVPT through TTS technology enhance ESL learners' pronunciation in terms of comprehensibility and accentedness (holistic analysis)? The results are detailed in subsections for discrete and holistic analyses, including statistical analyses and interpretations.

Discrete Analysis of Pronunciation: Past -ed Allomorphy

Phonological Awareness 1: Measured answers to open-ended questions

The first test of phonological awareness assessed participants' knowledge of the pronunciation of the English past *-ed* allomorphy through general open-ended questions (e.g., Do you know how past tense *-ed* is pronounced? Explain). Responses were rated on a scale from 1 to 4: 1 = no knowledge, 2 = knowledge of one allomorph, 3 = knowledge of two allomorphs, 4 = knowledge of all three allomorphs.

Within-Group Analysis. The Wilcoxon Signed-Rank Test revealed a statistically significant improvement in phonological awareness scores for the treatment group, $z = 0.0$, $p > 0.0001$ (corrected p-value = 0.0002), with the median score increasing from 3 before training to 4 after training Conversely, the control group showed a significant but smaller change, $z = 0.0$, $p = 0.0005$ (corrected p-value = 0.0001), with a median score remaining at 3 before and after training.

Between-Group Analysis. The Mann-Whitney U Test yielded a statistic of $z = 144.0$, $p = 0.1104$ (with the same corrected p-value = 0.1104), indicating that the improvement in phonological awareness in the treatment group was not significantly different from the control group.

Overall, these findings suggest that while the treatment group exhibited more gains in phonological awareness knowledge, the difference between groups was not statistically significant. Table 3 summarizes the descriptive and inferential statistics for Phonological Awareness 1. Statistically significant differences are marked in bold.

Table 3

Summary of Statistical Results for Awareness 1

Measure	Group	Test statistic (z)	p-value		Median	
			Uncorrected	Corrected	Pre	Post
Awareness 1	Treatment	0.0	>0.0001	***0.0002**	3	4
	Control	0.0	0.0005	***0.0001**	3	3
	Between Groups	144.0	0.1104	0.1104	-	-

Phonological Awareness 2: Associating inflected -ed forms

The second awareness test evaluated participants' ability to associate 21 target *-ed* forms (7 for each allomorph) with their corresponding allomorphic pronunciation.

Within-Group Analysis. The Wilcoxon Signed-Rank Test showed statistically significant improvements in the treatment group for all allomorphs. For the /t/ allomorph, $z = 0.0$, $p > 0.0001$ (corrected p-value = 0.0005), with the median score rising from 2.27 before training to 5.07 after training. For the /d/ allomorph, $z = 0.0$, $p > 0.0001$ (corrected p-value = 0.0005), with the median score increasing from 1.67 to 5.80. The /id/ allomorph also showed significant improvement, with $z = 0.0$, $p \geq 0.0008$ (corrected p-value = 0.0049), and the median score improving from 5.20 to 6.93.

In contrast, the control group showed significant changes but with smaller effect sizes. For the /t/ allomorph, $z = 2.5$, $p = 0.0023$ (corrected p-value = 0.0092), with the mean score increasing from 2.4 to 4.4. For the /d/ allomorph, $z = 0.0$, $p > 0.0001$ (corrected p-value = 0.0005), with the mean score rising from 1.6 to 4.6. The /id/ allomorph had $z = 0.0$, $p = 0.0009$ (corrected p-value = 0.0049), with the mean score improving from 4.27 to 6.8. In all cases, the control group's improvements were notable but less pronounced compared to the treatment group.

Between-Group Analysis. Mann-Whitney U Tests showed no significant differences between groups for the /t/ allomorph (z = 141.0, p = 0.2306, corrected p-value = 0.2783) and /id/ allomorph (z = 77.5, p = 0.1391, corrected p-value = 0.2783). For the /d/ allomorph, there was a trend toward greater improvement in the treatment group (z = 163.5, p = 0.0262, corrected p-value = 0.0788), although this did not reach statistical significance.

In summary, the results from the second awareness assessment reveal that the treatment group made some gains across all allomorphs, particularly for /t/ and /d/, compared to the control group, which also showed improvements but to a lesser degree. Although between-group comparisons did not yield significant differences for /t/ and /id/, there was a trend toward greater improvement for /d/ in the treatment group. Table 4 provides a summary of the descriptive and inferential statistics for Phonological Awareness 2, including the three allomorphs /t/, /d/, and /id/ for both groups, with statistically significant changes highlighted in bold.

Table 4

Summary of Statistical Results for Awareness 2.

Measure (Allomorph)	Group	Test statistic (z)	p-value		Mean	
			Uncorrected	Corrected	Pre	Post
/d/	Treatment	0.0	>0.0001	***0.0005**	1.67	5.80
	Control	0.0	>0.0001	***0.0005**	1.6	4.6
	Between Groups	163.5	0.0262	0.0788	-	-
/t/	Treatment	0.0	>0.0001	***0.0005**	2.27	5.07
	Control	2.5	0.0023	***0.0092**	2.4	4.4
	Between Groups	141.0	0.2306	0.2783	-	-
/id/	Treatment	0.0	≥ 0.0008	***0.0049**	5.20	6.93
	Control	0.0	0.0009	***0.0049**	4.27	6.8
	Between Groups	77.5	0.1391	0.2783	-	-

Holistic Analysis of Pronunciation: Comprehensibility and Accentedness

For this assessment, participants' spontaneous speech tasks (e.g., describing previous summer activities) were evaluated before and after the training by eleven experienced ESL teachers using a 9-point Likert scale to measure comprehensibility and accentedness. Lower scores signify more favorable outcomes, indicating that the speech is easier to understand and not accented at all.

Prior to conducting inferential tests, Intraclass Correlation Coefficient (ICC) values were calculated to assess rater reliability, with ICC(3,k) values of 0.927 (95% CI [0.88, 0.96], $p < 0.001$) for comprehensibility and 0.924 (95% CI [0.88, 0.96], $p < 0.001$) for accentedness, suggesting a high degree of agreement among the raters.

Comprehensibility

Within-Group Analysis. The Wilcoxon Signed-Rank Test revealed a statistically significant improvement in comprehensibility ratings for the treatment group, $z = 0.0$, $p = 0.0028$ (corrected p-value = 0.0055), with a median rating improving from 3 before training to 2 after training. Conversely, no significant change was observed in the control group, $z = 20.5$, $p = 0.4584$ (with the same corrected p-value), and a median rating of 4 before training and 3 after training.

Between-Group Analysis. The Mann-Whitney U Test indicated a significant difference between the treatment and control groups, with a statistic of $z = 27.0$, $p = 0.0002$ (corrected p-value = 0.0007), showing that the treatment group achieved greater improvements in comprehensibility compared to the control group.

Accentedness

Within-Group Analysis. The Wilcoxon Signed-Rank Test revealed a significant reduction in accentedness for the treatment group in the post-test, with $z = 0.0$, $p = 0.0013$ (corrected p-value = 0.0027), and median scores improving from 7 before training to 5 after training. In contrast, the

control group showed no significant changes, z = 16.0, p = 0.4263 (corrected p-value = 0.4263), with a median score of 7 both before and after training.

Between-Group Analysis. The Mann-Whitney U Test confirmed that the treatment group had significantly less accented speech than the control group, with z = 23.5, p = 0.0002 (corrected p-value = 0.0005), indicating a significant difference between the groups.

Overall, the holistic measures revealed significant improvements for the treatment group in both comprehensibility and accentedness, validating the effectiveness of the HVPT intervention. Table 5 offers an overview of the statistical results for both measures, with statistically significant differences highlighted in bold.

Table 5

Summary of Statistical Results for Holistic Ratings

Measure	Group	Test statistic (z)	p-value		Median	
			Uncorrected	Corrected	Pre	Post
Comprehensibility	Treatment	0.0	0.0028	***0.0055**	3	2
	Control	20.5	0.4584	0.4584	4	3
	Between Groups	27.0	0.0002	***0.0007**	-	-
Accentedness	Treatment	0.0	0.0013	***0.0027**	7	5
	Control	16.0	0.4263	0.4263	7	7
	Between Groups	23.5	0.0002	***0.0005**	-	-

Discussion

The current study evaluated the impact of a TTS-assisted HVPT approach on adult ESL learners' pronunciation. The research aimed to address two primary research questions: (1) Can HVPT using TTS technology enhance ESL learners' phonological awareness of the English past tense *-ed* morpheme (discrete analysis)? (2) Can HVPT using TTS technology improve ESL learners' pronunciation in terms of comprehensibility and accentedness (holistic analysis)?

The results suggest that integrating HVPT with TTS technology has the potential to enhance participants' phonological awareness of the English past *-ed* allomorphy and lead to improvements in comprehensibility and accentedness. While the treatment group showed greater improvements in both discrete and holistic pronunciation measures, these differences were not statistically significant for the former measures. The following sections explore potential explanations for these findings and, whenever relevant and appropriate, provide excerpts from the interviews to further shed light on the discussion.

Discrete Analysis: Awareness to Past -ed Allomorphy

The analysis of phonological awareness revealed that both groups demonstrated improvements in their knowledge of English past *-ed* allomorphs, regardless of the approach used (HVPT or one-voice TTS input). This finding supports previous studies indicating that TTS, with or without HVPT, can be effective in developing phonological awareness (Bione & Cardoso, 2020; Cardoso, 2018). Participants' feedback from both groups confirmed this observation, as they reported an increased knowledge of the allomorphs following the training. For instance, one participant noted "I initially thought there was only one way for pronouncing past *-ed*, but by the end of the study, I could distinguish them as /d/, /id/," while another mentioned, "I also discovered /t/, something I never thought existed."

Despite these improvements, the results did not reveal statistically significant differences between the treatment and control groups. This suggests that both groups benefitted from the TTS technology to a similar extent. The significant within-group improvement observed in the treatment group highlights HVPT's effectiveness in improving knowledge of *-ed* allomorphy. However, the lack of significant between-group differences suggests that TTS technology, whether

applied through highly variable or one-voice input, can contribute to improvements in phonological awareness. These findings are discussed in detail below.

Among the past *-ed* allomorphs, /d/ showed the highest improvement in the treatment group, highlighting the pedagogical benefits of HVPT for this specific allomorph. Although the control group also exhibited improvement, it was less pronounced in comparison. This notable gain for the /d/ allomorph may be linked to its orthographic representation. The clear orthographic correspondence between "-ed" and the /d/ allomorph (due to the presence of "d" in *-ed*) could facilitate its acquisition. This analysis was also proposed by Delatorre (2010), who highlighted the facilitative role of orthography in learning past tense forms. Participants observed this relationship, with comments such as, "I think /id/ is the most obvious one because that is exactly how we write it 'ed'; followed by /d/, it was easy to learn because the letter 'd' appears in the endings of past tense verbs."

Although there were no significant differences between the groups, the trend towards greater improvement observed in the treatment group for the /d/ allomorph aligns with research highlighting the advantages of high variability input for phonological learning (Bradlow & Bent, 2008; Thomson, 2018). The effectiveness of HVPT could potentially be more evident in this case: while HVPT offers potential benefits through its variable input, the general effectiveness of TTS is also notable.

Initially, the /t/ allomorph was not identified at all in the first awareness test and showed low levels of awareness in the second test across both groups. However, the post-tests revealed a statistically significant improvement in awareness of the /t/ allomorph in both groups, although the increase was less pronounced compared to its voiced counterpart, /d/. This indicates that while HVPT positively impacted learners' awareness of the /t/ allomorph, its effect was somewhat

weaker than for /d/. This observation is consistent with previous research suggesting that certain allomorphs are more resistant to acquisition due to their phonetic context (e.g., see Lively et al., 1993, who found similar challenges with /r/ and /l/ in word-initial consonant clusters).

Additionally, the relatively smaller improvement observed for /t/ in awareness supports the notion that this allomorph is particularly challenging to acquire (see Cardoso, 2018 and Dwight, 2012 for similar claims). One possible explanation for this difficulty is the opaque orthographic correspondence between orthographic "-ed" and /t/, which lacks a clear sound-to-letter mapping. As Jackson and Cardoso (2022) claim, discrepancies between L1 and L2 orthographic systems can impede L2 phonological development, suggesting that the /t/ allomorph's opaque orthographic cues might contribute to the observed difficulty.

The absence of significant between-group differences for both /d/ and /t/ allomorphs may suggest that the TTS-based input was sufficiently robust to support learning in both conditions. This level of effectiveness may be attributed to the high-quality auditory input provided by TTS technology, which has been noted as an effective tool for delivering auditory input (Bione et al., 2017; Cardoso et al., 2015). For example, some participants in the control group reported that their assigned TTS voice enhanced their ability to identify /d/ more distinctly, as one participant noted: "Because Jamie's voice was very clear and realistic, I now understand that the verb 'drag' is pronounced 'drag/d/' rather than 'drag/id/." Participants from both groups expressed surprise upon discovering the /t/ allomorph, with one stating: "I never thought that a sound like /t/ could come from the spelling of -ed, but TTS's accurate voices helped me notice and discover a new sound for past tense verbs." This feedback highlights the effectiveness of TTS in addressing the challenges associated with the irregularities of past *-ed* marking in English, as suggested by Bione et al. (2017) and John and Cardoso (2017).

Lastly, the /id/ allomorph showed the smallest enhancement between pre-test and post-test scores compared to /t/ and /d/, possibly due to a ceiling effect. As discussed above, because of the transparent orthographic correspondence between *-ed* and the /id/ allomorph, participants in both groups exhibited high initial awareness of this allomorph prior to the training, which remained unaffected on the post-test. Accordingly, participants already had robust knowledge of the /id/ allomorph, leaving little room for further improvement. This ceiling effect likely explains the absence of statistically significant differences between groups regarding /id/, indicating that the auditory input provided by TTS, regardless of its integration with HVPT, was likely sufficient to consolidate learner's existing knowledge (see Delatorre, 2010 for similar claims regarding the acquisition of English /id/).

Other factors may account for the relative ease of acquisition and the ceiling effect observed for the /id/ allomorph. Research by Barros (2003) and Zimmer et al. (2009) highlights the significant impact of L1 phonological structures on L2 acquisition. For example, languages like Arabic, which do not allow complex coda clusters, can hinder the phonological development of *-ed*. This issue is compounded by articulatory difficulties: complex coda sequences, such as /gd/ in "drag/gd/," are inherently more challenging to produce than singleton codas (Easterday, 2019). This challenge is particularly evident for Arabic speakers, who often insert an epenthetic vowel (as in "drag/id/") to ease the articulation of complex /gd/ (Kharma & Hajjaj, 1997; Salim & Mohammed, 2023). As one participant remarked, "It is easier to pronounce /id/ for past tense verbs because articulating two consecutive consonants feels awkward."

Overall, while HVPT provides some pedagogical benefits via its ability to provide learners with variable input, the findings from the discrete analysis suggest that TTS, with or without high

variability training, can be a valuable tool for enhancing phonological awareness (for similar claims, see De Araújo Gomes et al., 2018).

Holistic Analysis: Comprehensibility and Accentedness

This study also examined the impact of TTS-assisted HVPT on holistic measures of pronunciation, focusing on comprehensibility and accentedness. The results revealed significant improvements in both measures for the treatment group, whereas the control group showed no notable changes. These findings align with previous research, which highlights HVPT's effectiveness in enhancing both aural perception and oral production (e.g., Bradlow & Bent, 2008; Lively et al., 1993; Logan et al., 1991; Thomson, 2018).

The significant enhancements in comprehensibility and reductions in accentedness observed in the treatment group highlight the effectiveness of TTS in improving pronunciation skills through exposure to a variety of accents and voice types. These results further emphasize the importance of diverse linguistic input in developing robust phonetic representations among L2 learners (Flege, 1999; Moyer, 2009). Several factors contribute to these observed improvements. First, the variable speech samples provided by TTS likely facilitated more effective phonetic category formation and likely reduced L1 interference (Flege, 1995), confirming Bione and Cardoso's (2020) claim that TTS can be a valuable tool for HVPT in L2 pronunciation pedagogy. Additionally, the self-paced nature of the training enabled learners to engage with the material at their own convenience, which may lead to more sustained learning outcomes. This aspect is supported by Cardoso (2018, 2022), and Kiliçkaya (2008), who emphasize TTS's potential to enhance L2 pronunciation while fostering learner autonomy.

Notably, the integration of TTS technology represents a novel application of HVPT, extending its benefits beyond traditional laboratory settings (Thomson, 2018). TTS technology

facilitates self-paced, autonomous learning, providing a practical solution to the constraints often faced in L2 education, such as limited access to fluent or native speakers and diverse speech models (Cardoso 2018; Collins & Muñoz, 2016; Bione & Cardoso, 2020). As one participant noted, "My favorite thing about this TTS tool is that I can use it any time I want and as much as I want… this made the learning experience way easier than traditional methods." Such feedback highlights the value of TTS technology in offering flexible and accessible learning opportunities.

Conclusion

The purpose of this study was to evaluate the impact of a TTS-assisted HVPT approach on adult ESL learners' pronunciation, with a focus on two key aspects: phonological awareness of the English past tense *-ed* morpheme (discrete analysis) and overall pronunciation in terms of comprehensibility and accentedness (holistic analysis).

The results show that the of integration of HVPT with TTS has the potential to enhance learners' phonological awareness of the past *-ed* morpheme. Participants in the treatment group exhibited improvement in their knowledge of *-ed* allomorphy, particularly the /d/ allomorph, compared to those exposed to a single TTS input (control). Despite the observed improvements, the differences in phonological awareness between the two groups were not statistically significant, indicating that while the HVPT approach offered some pedagogical advantages, the one-voice TTS input also had a beneficial effect. In terms of holistic pronunciation, the results revealed significant improvements in both comprehensibility and accentedness for the treatment group, whereas the control group did not exhibit notable changes in these pronunciation measures. Overall, these findings highlight that while TTS-based HVPT can improve phonological awareness, its impact on comprehensibility and accentedness is even more pronounced.

While this study provides some evidence for the efficacy of TTS-assisted HVPT, several limitations should be acknowledged. First, the study focused exclusively on phonological awareness as a measure of discrete analysis. Although phonological awareness is fundamental to the stages of pronunciation development (Celce-Murcia et al., 2010), a more comprehensive evaluation should include aural perception and oral production (including both controlled and spontaneous speech). Incorporating these additional measures would offer a fuller understanding of the impact of TTS-assisted HVPT on all facets of pronunciation development. Second, the study's duration was relatively short, with participants engaging in HVPT over a one-month period. Extended exposure to HVPT with TTS technology, incorporating spaced practice and long-term interaction, could provide more robust insights into the sustainability of pronunciation improvements. Future research should adopt a longitudinal design to explore the enduring effects of TTS-assisted HVPT on pronunciation development. Lastly, the study utilized a modest sample size and was conducted within a specific linguistic and cultural context among Kuwait ESL learners. To enhance the generalizability of the findings, future research should involve a larger and more diverse sample, encompassing various language backgrounds and educational settings.

In an era where technology is continually reshaping language learning, the integration of TTS technology into HVPT represents an exciting frontier for enhancing L2 pronunciation. This study demonstrates that TTS-assisted HVPT can not only refine learners' phonological awareness but also lead to significant gains in comprehensibility and reduced accentedness. TTS technology proves to be a powerful tool in bridging gaps in pronunciation instruction by offering diverse and authentic auditory input, particularly in a setting where learners can manage their own leaning experience.

Chapter Three

In this chapter, we will summarize the key findings and conclusions drawn in Chapter 2, discuss their significance for research and language education, and outline possible directions for future research.

Summary of Goals and Findings

This study aimed to evaluate the integration of TTS technology into HVPT to enhance pronunciation skills among ESL learners. Specifically, the project examined both morphophonemic features (i.e., consisting of individual segments or clusters such as /d/ and /dg/ respectively) and holistic aspects of pronunciation (i.e., comprehensibility, accentedness). The morphophonemic (discrete) analysis assessed the phonological awareness of regular past tense marking in English (-*ed*, as in walk/t/, play/d/ and visit/id/), following Celce-Murcia et al.'s (2010) framework for pronunciation development. This framework posits that acquisition of phonology adheres to a developmental hierarchy that begins with phonological awareness, progresses through aural perception and oral production, and culminates in oral fluency. Based on insights from Munro and Derwing (1995), the holistic analysis evaluated two broader aspects of pronunciation, namely comprehensibility and accentedness.

A mixed-methods research design was employed, involving pre-test and post-test measurements. Participants were divided into two groups: a Treatment group, which engaged with TTS-assisted HVPT training, and a Control group, which used a TTS single-input approach. Both interventions consisted of self-paced sessions conducted over four weeks.

The findings revealed that both groups improved in their ability to display knowledge of phonological awareness of past -*ed* morphophonemics, particularly regarding the /d/ allomorph. However, despite these improvements, no statistically significant differences were observed

between the two groups in the discrete analysis, suggesting that both types of TTS input can effectively support the development of phonological awareness. In contrast, the holistic analysis demonstrated significant improvements in comprehensibility and reductions in accentedness for the Treatment group, whereas the Control group showed no notable changes. These results underscore the potential effectiveness of TTS-assisted HVPT in enhancing pronunciation skills, particularly in terms of comprehensibility and accentedness.

Implications for L2 Education

The findings of this study have some implications for language education. The observed improvements in both phonological awareness (for both groups) and holistic pronunciation measures (for the HVPT Group) highlight the potential of integrating TTS technology into pronunciation training. Educators can leverage TTS technology to offer students a range of auditory experiences that can help mitigate L1 interference and establish more robust perceptual accuracy, consequently enhancing pronunciation learning. In general, our findings suggest that TTS-assisted HVPT can be an effective tool for providing opportunities for autonomous learning, thereby extending the classroom experience and enabling learners to engage with L2 material beyond traditional instructional settings.

Further Research

Future research should explore some of the limitations encountered in this research project. Longitudinal studies are necessary to gain insights into the long-term effects of TTS-assisted HVPT on pronunciation development, especially regarding the sustainability of improvements over time. Additionally, involving a larger and more diverse sample of ESL learners could enhance the generalizability of the results, providing a broader understanding of how TTS-assisted HVPT performs across various linguistic and cultural contexts. Expanding the evaluation to include

measures of aural perception as well as guided and spontaneous oral production would offer a more comprehensive view of the impact of TTS-assisted HVPT on pronunciation development. These research directions have the potential to contribute to refining TTS-assisted HVPT pedagogy and improving their effectiveness in diverse educational environments.

Conclusion

This study demonstrates that TTS-assisted HVPT can enhance ESL learners' phonological awareness and significantly improve their pronunciation in terms of comprehensibility and accentedness. While the findings indicate that both TTS-assisted and one-voice TTS inputs have beneficial effects on phonological awareness, the more pronounced improvements observed in the treatment group for holistic pronunciation measures highlight the added value of HVPT. These results support the integration of TTS technology into pronunciation training as a practical and effective approach to enhancing ESL learners' overall pronunciation skills. Ultimately, this study contributes to the body of knowledge on the use of speech technology in L2 education and highlights the importance of embracing new tools (e.g., TTS) and methods (e.g., HVPT) to support learners' pronunciation development.

References

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

Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation*. Oxford University Press.

Barcomb, M., & Cardoso, W. (2020). Rock or lock? Gamifying an online course management system for pronunciation instruction: Focus on English /r/ and /l/. *CALICO Journal*, *37*(2), 127–147. https://doi.org/10.1558/cj.36996

Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition*, *27*(3), 387–414. https://doi.org/10.1017/S0272263105050175

Barriuso, T. A., & Hayes-Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *CATESOL Journal*, *30*(1), 177-194.

Barros, A. M. V. (2003). *Pronunciation difficulties in the consonant system experienced by Arabic speakers when learning English after the age of puberty* [Unpublished Master's thesis]. West Virginia University, Morgantown. https://doi.org/10.33915/etd.766

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

Bione, T., Grimshaw, J., & Cardoso, W. (2016). An evaluation of text-to-speech synthesizers in the foreign language classroom: learners' perceptions. In S. Papadima-Sophocleous, L. Bradley & S. Thouësny (Eds), *CALL communities and culture – short papers from*

EUROCALL 2016 (pp. 50-54). Research-publishing.net. https://doi.org/10.14705/rpnet.2016.eurocall2016.537

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

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*(5), 977–985. https://doi.org/10.3758/bf03206911

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101*(4), 2299–2310. https://doi.org/10.1121/1.418276

Cardoso, W. (2011). The development of coda perception in second language phonology: A variationist perspective. *Second Language Research, 27*(4), 433-465. https://doi.org/10.1177/0267658311413540

Cardoso, W. (2018). Learning L2 pronunciation with a text-to-speech synthesizer. In Taalas, P., Jalkanen, J., Bradley, L., & Thouësny, S., (Eds.), *Proceedings of the European*

Association for Computer-Assisted Language Learning – EUROCALL 2018 (pp.16-21). https://doi.org/10.14705/rpnet.2018.26.806

Cardoso, W. (2022). Technology for Speaking Development. In T. Derwing, M. Munro, & R. Thomson (Eds), *Routledge Handbook on Second Language Acquisition and Speaking* (p. 299-313). Routledge, Taylor & Francis Group.

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

Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge University Press.

Crosby, C. (2020). Adding Production to High Variability Phonetic Training. *Honors Theses*, (1471). Retrieved from https://egrove.olemiss.edu/hon_thesis/1471

Collins, L., & Muñoz, C. (2016). The foreign language classroom: Current perspectives and future considerations. *The Modern Language Journal*, *100*(S1), 133–147. https://doi.org/10.1111/modl.12305

Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some input on the easy/difficult grammar question: An empirical study. *The Modern Language Journal, 93*(3), 336–353. https://doi.org/10.1111/j.1540-4781.2009.00894.x

Cook, V. J. (1992). Evidence for multicompetence. *Language Learning, 42*(4), 557–591. https://doi.org/10.1111/j.1467-1770.1992.tb01044.x

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect, 21*(1), 46–66.

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

Delatorre, F. (2010). The role of orthography on the production of regular verbs ending in ed by Brazilian EFL learners. *In Proceedings of the 9th Seminário do Círculo de Estudos Linguísticos do Sul* (pp. 1-13). Florianópolis: Federal University of Santa Catarina.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*(3), 379-397. https://doi.org/10.2307/3588486

Dwight, V. (2012). *Regular Past Tense Acquisition in L2 English: The Roles of Perceptual Salience and Readiness* [Unpublished Master's thesis]. Concordia University.

Easterday, S. (2019). *Highly complex syllable structure: A typological and diachronic study* (Studies in Laboratory Phonology 9). Language Science Press.

Eksi, G. Y., & Yesilcinar, S. (2016). An investigation of the effectiveness of online text-to-speech tools in improving EFL teacher trainees' pronunciation. *English Language Teaching, 9*(2), 205-214. https://doi.org/10.5539/elt.v9n2p205

Everly, P. (2019). Expanding pronunciation instructional time beyond the classroom: Microsoft Office 2016 OneNote Class Notebook as an interactive delivery platform. *TESOL Journal, 10*(2), e421. https://doi.org/10.1002/tesj.421

Farhat, P., & Dzakiria, H. (2017). Pronunciation barriers and computer assisted language learning (CALL) coping the demands of 21st century in second language learning classroom in

Pakistan. *International Journal of Research in English Education, 2*(2), 53–62. https://doi.org/10.18869/acadpub.ijree.2.2.53

Flege, J. E. (1988). The production and perception of speech sounds in a foreign language. In H. Winitz (Ed.), *Human communication and its disorders: A review* (pp. 224–401). Ablex.

Flege, J. E. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Huebner & C. A. Ferguson (Eds.), *Cross Currents in Second Language Acquisition and Linguistic Theory* (pp. 249–289). John Benjamins. https://doi.org/10.1075/lald.2.15fle

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

Flege, J. E. (1999). Age of learning and second language speech. In D. Birdsong (Ed.), *Second language learning and the critical period hypothesis*. (pp. 101–131). Erlbaum.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*(4), 527–552. https://doi.org/10.1017/s0272263101004041

Fraser, H. (2000). *Coordinating improvements in pronunciation teaching for adult learners of English as a second language*. DETYA (ANTA Innovative project).

Frese, R. A. (2006). *The relationship between perception and production of words ending in-ed by Brazilian EFL learners* [Unpublished Master's thesis]. Universidade Federal de Santa Catarina, Florianópolis.

Gass, S. M., & Varonis, E. M. (1994). Input, interaction, and second language production. *Studies in Second Language Acquisition*, *16*(3), 283–302. https://doi.org/10.1017/s0272263100013097

González, D. (2007). Text-to-speech applications used in EFL contexts to enhance pronunciation. *TESL-EJ, 11*(2), 1-11.

Hamm, O. (2020). *Where does pronunciation teaching stand in the globalization and internationalization of English?* [Unpublished Master's thesis]. University of Jyväskylä, Finland. Retrieved from http://urn.fi/URN:NBN:fi:jyu202006053992.

Herd, W., Jongman, A., & Sereno, J. (2013). Perceptual and production training of intervocalic /d, ɾ, R/ in American English learners of Spanish. *The Journal of the Acoustical Society of America, 133*(6), 4247–4255. https://doi.org/10.1121/1.4802902

Honeybone, P. (2011). Variation and linguistic theory. In W. Maguire & A. McMahon (Eds), *Analysing Variation in English* (pp. 151–177). Cambridge University Press. https://doi.org/10.1017/cbo9780511976360.008

Ingvalson, E. M., Ettlinger, M., & Wong, P. C. M. (2014). Bilingual speech perception and learning: A review of recent trends. *International Journal of Bilingualism*, *18*(1), 35–47. https://doi.org/10.1177/1367006912456586

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866–877. https://doi.org/10.1121/1.3148196

Jackson, S., & Cardoso, W. (2022). Orthographic interference in the acquisition of English /h/ by francophones. *Second Language Pronunciation,* 229–248. https://doi.org/10.1515/9783110736120-009

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.

Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics, 24*(1), 131–161. https://doi.org/10.1017/s0142716403000079

Jing, Z. (2010). A new approach to college English pronunciation teaching. *Shandong Foreign Language Teaching Journal, 31*(03), 60–63.

Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. In T. Johns & P. King (Eds.), *Classroom concordancing. English Language Research, 4,* 1-16.

John, P., & Cardoso, W. (2017). A comparative study of text-to-speech and native speaker output. In J. Demperio, E. Rosales & S. Springer (Eds.), *Proceedings of the meeting on English language teaching* (pp. 78-96). Université du Québec à Montréal Press.

Kachru, B. B. (1992). *The other tongue: English across cultures*. University of Illinois Press.

Kangatharan, J., Giannakopoulou, A., & Uther, M. (2021). The effect of HVP training in vowel perception on bilingual speech production, *Research in English Language Teaching, 2*(1), 15-32.

Kartushina, N., & Martin, C. D. (2018). Talker and acoustic variability in learning to produce nonnative sounds: Evidence from articulatory training. *Language Learning, 69*(1), 71–105. https://doi.org/10.1111/lang.12315

Kennedy, S., & Trofimovich, P. (2019). Comprehensibility: A useful tool to explore listener understanding. *Canadian Modern Language Review, 75*(4), 275–284. https://doi.org/10.3138/cmlr.2019-0280

Kharma, N., & Hajjaj, A. (1997). *Errors in English among Arabic speakers: Analysis and remedy*. York Press.

Kiliçkaya, F. (2008). Improving pronunciation via accent reduction and text-to-speech software. In T. Koyama (Ed.), *Proceedings of the WorldCALL 2008 conference* (pp. 135–137). Nagoya, Japan: The Japan Association for Language Education and Teaching.

Kim, S. (2018). Exploring media literacy: Enhancing English oral proficiency and autonomy using media technology. *Studies in English Education, 23*(2), 473–500. https://doi.org/10.22275/see.23.2.03

Kingston, J. (2003). Learning foreign vowels. *Language and Speech, 46*(2–3), 295–348. https://doi.org/10.1177/00238309030460020201

Krashen, S. (1985). *The Input Hypothesis: Issues and Implications*. Longman.

Layton, L., Deeny, K., Upton, G., & Tall, G. (1998). A Pre-school training programme for children with poor phonological awareness: Effects on reading and spelling. *Journal of Research in Reading, 21*(1), 36–52. https://doi.org/10.1111/1467-9817.00041

Levis, J. M. (2016). Research into practice: How research appears in pronunciation teaching materials. *Language Teaching, 49*(3), 423–437. https://doi.org/10.1017/s0261444816000045

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

Liakin, D., Cardoso, W., & Liakina, N. (2017a). Mobilizing instruction in a second-language context: Learners' perceptions of two speech technologies. *Languages, 2*(3), 11. https://doi.org/10.3390/languages2030011

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

Linebaugh, G., & Roche, T. (2015). Evidence that L2 production training can enhance perception. *Journal of Academic Language & Learning, 9*(1), 1-17.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*(3), 1242–1255. https://doi.org/10.1121/1.408177

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America, 96*(4), 2076–2087. https://doi.org/10.1121/1.410149

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America, 89*(2), 874–886. https://doi.org/10.1121/1.1894649

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience, 2*(2), 89–108. https://doi.org/10.3758/cabn.2.2.89

Moghaddam, M. S., Nasiri, M., Zarea, A., & Sepehrinia, S. (2012). Teaching pronunciation: The lost ring of the chain. *Journal of Language Teaching and Research, 3*(1), 215-219. https://doi.org/10.4304/jltr.3.1.215-219

Moon, D. (2012). Web-based text-to-speech technologies in foreign language learning: Opportunities and challenges. In T. Kim, J. Ma, W. Fang, Y. Zhang, A. Cuzzocrea (Eds.), *Computer Applications for Database, Education, and Ubiquitous Computing* (pp. 120-125). Springer. https://doi.org/10.1007/978-3-642-35603-2_19

Morin, R. (2007). A neglected aspect of the standards: Preparing foreign language Spanish teachers to teach pronunciation. *Foreign Language Annals, 40*(2), 342–360. https://doi.org/10.1111/j.1944-9720.2007.tb03206.x

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly, 25*(3), 481-520. https://doi.org/10.2307/3586981

Moyer, A. (2009). Input as a critical means to an end: Quantity and quality of experience in L2 phonological attainment. In T. Piske & M. Young-Scholten (Eds.), *Input Matters in SLA* (pp. 159–174). Multilingual Matters.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., & Derwing, T. M. (2020). Foreign accent, comprehensibility and intelligibility, Redux. *Journal of Second Language Pronunciation, 6*(3), 283–309. https://doi.org/10.1075/jslp.20038.mun

Nunan, D., & Richards, J. (2015). *Language learning beyond the classroom*. Routledge.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472. https://doi.org/10.1121/1.3593366

Prashant, P. D. (2018). Importance of pronunciation in English language communication. *Pronunciation and Communication, 7*(2), 16-17.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America, 119*(3), 1684–1696. https://doi.org/10.1121/1.2161427

Reinders, H., & White, C. (2010). The theory and practice of technology in materials development and task design. In N. Harwood (Ed.), *English Language Teaching Materials: Theory and Practice* (pp. 58-80). Cambridge University Press.

Richards, J. C. (2015). The changing face of language learning: Learning beyond the classroom. *RELC Journal, 46*(1), 5–22. https://doi.org/10.1177/0033688214561621

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science, 12*(2), 339–349. https://doi.org/10.1111/j.1467-7687.2008.00786.x

Royani, I., & Rahmi, A. (2019). Phonetic transcription on allomorphs of past participle form of regular verbs. *English Education: English Journal for Teaching and Learning, 7*(2), 197-210.

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.01318

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics, 39*(1), 187–224. https://doi.org/10.1017/s0142716417000418

Salim, E. A. E., & Mohammed, F. A. H. (2023). Mother tongue interference in teaching English. *Multicultural Education, 9*(6). https://doi.org/10.5281/zenodo.8025208

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237-326). Newbury House.

Shin, D. J., & Iverson, P. (2013). Training Korean second language speakers on English vowels and prosody. *Proceedings of Meetings on Acoustics, 19*(1). https://doi.org/10.1121/1.4801046

Shinohara, Y., & Iverson, P. (2013). Computer-based English /r/-/l/ perceptual training for Japanese children. *Proceedings of Meetings on Acoustics, 19*(1). https://doi.org/10.1121/1.4800136

Sifakis, N. C., & Sougari, A.-M. (2005). Pronunciation issues and EIL pedagogy in the periphery: A survey of Greek state school teachers' beliefs. *TESOL Quarterly, 39*(3), 467 -488. https://doi.org/10.2307/3588490

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

Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning*, *62*(4), 1231–1258. https://doi.org/10.1111/j.1467-9922.2012.00724.x

Thomson, R. I. (2018). High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation, 4*(2), 208–231.

Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference*. Oct. 2015. (pp. 88‑97). Iowa State University.

Venkatesh, V., & Davis, F. (2000). A theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science, 46*(2), 186–204.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America, 106*(6), 3649–3658. https://doi.org/10.1121/1.428217

Wong, J. W. (2014). The effects of high and low variability phonetic training on the perception and production of English vowels /e/-/æ/ by Cantonese ESL learners with high and low L2 proficiency levels. *Interspeech 2014*. https://doi.org/10.21437/interspeech.2014-129

Zhang, X., Cheng, B., & Zhang, Y. (2021). The role of talker variability in nonnative phonetic learning: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research, 64*(12), 4802–4825. https://doi.org/10.1044/2021_jslhr-21-00181

Zimmer, M., Alves, U. K., & Silveira, R. (2009). *Pronunciation instruction for Brazilians: Bringing theory and practice together*. Cambridge Scholars.

Appendices

Appendix A-1

Consent form (for learners)

INFORMATION AND CONSENT FORM

Study Title: Exploring High-Variability Phonetic Training through TTS Technology in ESL Pronunciation Pedagogy

Researcher: Forcan Al-Shami (Master's student in Applied Linguistics)

Researcher's Contact Information: f_lshami@live.concordia.ca

Faculty Supervisors: Dr. Walcir Cardoso Applied Linguistics/Education

Faculty Supervisor's Contact Information:

Dr. Walcir Cardoso: walcir.cardoso@concordia.ca; 514-848-2424 x2451; office S-FG 6441

Faubourg Ste-Catherine Building,

1610 St. Catherine W.

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

A. PURPOSE

The purpose of the research is to investigate the effectiveness of Text-To-Speech technology in improving English pronunciation skills among Arabic-speaking ESL learners, with a focus on its integration into language learning settings, as well as your perceptions of the technology and the overall learning experience.

B. PROCEDURES

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

- Fill out the demographic survey to provide us with some background information (approximately 5 minutes)

- Complete a set of online pre-tests to assess your English pronunciation skills (approximately 30 minutes).
- Watch a short video explaining English past *-ed* and how to use Text-To-Speech for pronunciation practice (approximately 10 minutes).
- Engage in pronunciation-learning exercises using Text-To-Speech technology over a period of 10 sessions (approximately 20-30 minutes for each session).
- Complete a set of online post-tests after the training to assess your English pronunciation skills (approximately 30 minutes).
- Participate in a short oral interview to share your insights related the learning experience (approximately 10-15 minutes)

Note:

- All aspects of the study will be conducted remotely.
- The interview will be recorded through Zoom for further data analysis. Neither your face nor your name will appear in the recording or be published.

In total, participating in this study will take around six weeks to be completed.

C. RISKS AND BENEFITS

Your participation in this research carries no known risks. Instead, it offers potential benefits such as gaining insights into English pronunciation and discovering a potentially beneficial technology to enhance your English pronunciation.

D. CONFIDENTIALITY

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

- During the pre-tests and post-tests, we will collect data on your language skills and performance as part of this research. These assessments will be recorded for analysis purposes.
- During the interview, we will gather data on your learning strategies and perceptions of using Text-to-speech as a learning tool. The interview will be recorded for transcription and analysis purposes.

We want to assure you of the following regarding the confidentiality of your information:

- We will not allow anyone to access your information, except people directly involved in conducting the research. We will only use the information for the purposes of the research described in this form.
- The information gathered will be coded and interviews will be transcribed. We will not include your name; instead, a pseudonym will be used. This means your name will not appear anywhere in the written study, and no one else will have access to your personal information.

- We will protect the information by keeping all digital files on a password protected computer.
- We intend to publish the results of the research. However, it will not be possible to identify you in the published results.
- We will destroy the information three years after the end of the study.

F. CONDITIONS OF PARTICIPATION

- You do not have to participate in this research. It is purely your decision. If you do participate, you can stop at any time. You can also ask that the information you provided not be used, and your choice will be respected. If you decide that you don't want us to use your information, you must tell the researcher before June 30, 2024.

- As a compensatory indemnity for participating in this research, you will receive $30. If you withdraw before the end of the research, you will receive $10 regardless of when you withdraw.

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

G. PARTICIPANT'S DECLARATION

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

NAME (please print)_____

SIGNATURE _____

DATE _____

If you have questions about the scientific or scholarly aspects of this research, please contact the researcher. Their contact information is on page 1. You may also contact the faculty supervisor (Dr. Walcir Cardoso: walcir.cardoso@concoria.ca)

If you have concerns about ethical issues in this research, please contact the Manager, Research Ethics, Concordia University, 514.848.2424 ex. 7481 or oor.ethics@concordia.ca.

Appendix A-2

Consent Form (for raters)

INFORMATION AND CONSENT FORM

Study Title: Exploring High-Variability Phonetic Training through TTS Technology in ESL Pronunciation Pedagogy

Researcher: Forcan Al-Shami (Master's student in Applied Linguistics)

Researcher's Contact Information: f_lshami@live.concordia.ca

Faculty Supervisors: Dr. Walcir Cardoso Applied Linguistics/Education

Faculty Supervisor's Contact Information:

Dr. Walcir Cardoso: walcir.cardoso@concordia.ca; 514-848-2424 x2451; office S-FG 6441

Faubourg Ste-Catherine Building,

1610 St. Catherine W.

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

A. PURPOSE

The purpose of the research is to investigate the effectiveness of Text-To-Speech technology in improving English pronunciation skills among Arabic-speaking ESL learners, with a focus on its integration into language learning teaching and how much learners can benefit from the experience.

B. PROCEDURES

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

- Listen to 60 audio recordings of 30 participants (estimate). Each audio file contains English speech samples produced by language students (approximately 20-30 seconds for each audio recording).

- Evaluate each participant's pronunciation in terms of comprehensibility and accentedness by completing a 9-point Likert scale (approximately 3 minutes for each participant).
- Submit completed evaluation forms via email to the researcher (approximately 5 minutes).

Note:

- All aspects of the study will be conducted remotely.
- Your name will not appear in dissemination format such as publications, presentations, and reports.

In total, participating in this study will take approximately 4 hours to be completed.

C. RISKS AND BENEFITS

Your participation in this research carries no known risks and any major benefit. However, it offers potential benefits such as gaining insights into students' English pronunciation and discovering a potentially beneficial technology to enhance your English pronunciation classes.

D. CONFIDENTIALITY

We want to assure you of the following regarding the confidentiality of your information:

- We will not allow anyone to access your information, except people directly involved in conducting the research. We will only use the information for the purposes of the research described in this form.
- The information gathered for the evaluation forms will be coded. We will not include your name; instead, a pseudonym will be used. This means your name will not appear anywhere in the written study, and no one else will have access to your personal information.
- We will protect the information by keeping all digital files on a password protected computer.
- We intend to publish the results of the research. However, it will not be possible to identify you in the published results.
- We will destroy the information three years after the end of the study.

F. CONDITIONS OF PARTICIPATION

- You do not have to participate in this research. It is purely your decision. If you do participate, you can stop at any time. You can also ask that the information you provided not be used, and your choice will be respected. If you decide that you don't want us to use your information, you must tell the researcher before July 30, 2024.
- As a compensatory indemnity for participating in this research, you will receive $30. If you withdraw before the end of the research, you will receive $10 regardless of when you withdraw.
- There are no negative consequences for not participating, stopping in the middle, or asking us not to use your information.

G. PARTICIPANT'S DECLARATION

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

NAME (please print)_____

SIGNATURE _____

DATE _____

If you have questions about the scientific or scholarly aspects of this research, please contact the researcher. Their contact information is on page 1. You may also contact the faculty supervisor (Dr. Walcir Cardoso: walcir.cardoso@concoria.ca)

If you have concerns about ethical issues in this research, please contact the Manager, Research Ethics, Concordia University, 514.848.2424 ex. 7481 or oor.ethics@concordia.ca.

Appendix B

Demographic questionnaire

Please fill out the following form. (You can use ✓ to answer some of the questions)

1. Name: _____

2. Age: ( )

3. Gender: Female ( ) Male ( ) Prefer not to answer ( )

4. Nationality: _____

5. Where do you live? _____

6. What is your native language? _____

7. What is your course level (if you are already attending an English course): _____

8. How many languages do you know? And what is your proficiency level in them?

Language	Beginner	Lower-intermediate	Upper-intermediate	Advanced

9. How good are you with using technology in general? Please rate your overall knowledge for using technology on a scale from **1** (no knowledge at all) and **9** (very advanced knowledge):

No Knowledge at all	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Very Advanced Knowledge**

10. Are you familiar with text-to-speech (TTS) technology? Yes ( ) No ( )

11. How often do you use text-to-speech feature for language learning or pronunciation practice?

Always ( ) Often ( ) Sometimes ( ) Rarely ( ) Never ( )

12. Do you have any history of hearing or speech disorders? Yes ( ) No ( )

Appendix C

Pre- and Post-test: Awareness #1

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

- Follow-up questions:
    - Please explain what sound is produced when -*ed* is added to the end of a verb?
    - Do you think past tense -*ed* is pronounced differently for different verbs?
    - Give examples if you can.

Appendix D

Pre- and Post-test: Awareness #2

- **Select the verbs that have similar ending pronunciations:**

[Different pronunciations of the past tense *-ed* ending are color-coded as follows]:

- /t/ are color coded in red
- /d/ are color coded in blue
- /id/ are color coded in green

1) admitted a) used b) asked c) added	**8) laughed** a) used b) asked c) added	**15) cried** a) used b) asked c) added
2) promised a) used b) asked c) added	**9) suggested** a) used b) asked c) added	**16) finished** a) used b) asked c) added
3) enjoyed a) used b) asked c) added	**10) watched** a) used b) asked c) added	**17) hated** a) used b) asked c) added
4) opened a) used b) asked c) added	**11) wanted** a) used b) asked c) added	**18) cleaned** a) used b) asked c) added
5) needed a) used b) asked c) added	**12) imagined** a) used b) asked c) added	**19) walked** a) used b) asked c) added
6) stopped a) used b) asked c) added	**13) parked** a) used b) asked c) added	**20) lived** a) used b) asked c) added
7) allowed a) used b) asked c) added	**14) decided** a) used b) asked c) added	**21) ended** a) used b) asked c) added

Note: Information within brackets [ ] and color codes are for reader reference only and will not be shared with participants.

Appendix E

Pre- and Post-test: Holistic Pronunciation (for participants)

We would like you to describe a personal experience in English. Please try to provide as much detail as possible in your answer.

Pre-Test:

- [Group A: Administered to 50% of Participants]

  Please describe what you did last summer.

- [Group B: Administered to other 50% of Participants]

  Please describe what you did on your last birthday.

Post-Test:

- [Group A: Administered to 50% of Participants]

  Please describe what you did on your last birthday.

- [Group B: Administered to other 50% of Participants]

  Please describe what you did last summer.

Note: Information within brackets [ ] is for reader reference only and will not be shared with participants.

Appendix F

Pre- and Post-test: Holistic Pronunciation (for raters)

You will have access to 60 audio recordings that are numbered from 1-60. Each audio file contains English speech samples produced by a certain speaker. Please listen carefully to each audio recording and complete this evaluation form accordingly:

Recording Number: _____

1) *Comprehensibility:* How clear and easy was the speech to understand? On a scale of 1 to 9, Where 1 indicates "Extremely Easy to Understand" and 9 indicates "Extremely Hard to Understand," please rate the overall comprehensibility of the speech by placing (✓) on the scale below:

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
Extremely Easy to Understand										**Extremely Hard to Understand**

2) *Accentedness*: How accented was the speech? On a scale of 1 to 9, Where 1 indicates "Not Accented" and 9 indicates "Strongly Accented," please rate the overall accentedness of the speech by placing (✓) on the scale below:

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
Not Accented at all										**Strongly Accented**

Appendix G

The transcript of the video

Title: Understanding Regular Past Tense Verbs

Researcher: Welcome to this brief lesson on regular past tense verbs.

[Scene transition: The screen displays text - "Why Learn Past Tense?"]

Researcher: So, why learn about the past tense? Well, the past tense is used to talk about actions that happened or were completed in the past. It's an important part of English grammar, and it's essential for effective communication.

[Scene transition: The screen displays text – "Two Types of Past Tense Verbs"]

Researcher: Past tense verbs in English are divided into two main groups: regular verbs and irregular verbs. Today, we'll focus on regular verbs, which follow a straightforward pattern.

[Scene transition: The screen displays text – "Forming Past Tense with Regular Verbs"]

Researcher: Now, let's dive into how you form the past tense with regular verbs. It's quite simple, actually. In writing, you create the past tense of regular verbs by adding *-ed* to the base form of the verb.

[Scene transition: Examples on the screen]

Researcher: Let's look at some examples to make it clear:

- For "talk," the past tense is "talked."
- "Need" becomes "needed."
- And "listen," becomes "listened."

[Scene transition: The screen displays text – "Consistency Across Subjects"]

Researcher: What's great about regular verbs is that the past tense form remains consistent across all subjects. No matter if you're talking about yourself, someone else, or even a group, the past tense stays the same.

[Scene transition: Examples of subject-verb agreement]

Researcher: For instance, take the verb "walk." In the past tense, it looks like this:

- "I walked"
- "You walked"
- "He walked"
- "She walked"
- "It walked"
- "We walked"
- "They walked"

[Scene transition: The screen displays text – “Pronunciation Variations”]

Researcher: Now, let’s explore something interesting. The past tense ending, -*ed*, can be pronounced in different ways. During this study, you will have to find out these pronunciation variations and understand when to use them in different contexts.

[Scene transition: The screen displays text – “Speechify – Text-To-Speech”]

To practice your English, we are going to use a cool tool called “Speechify”

Speechify is a creative online tool that can turn your text into speech. You can choose from many voices to create realistic and amazing speech.

To get started, visit the Speechify homepage. Click on “Text to Speech” at the top. In this white box, type or paste your words or sentences to generate speech. Then select a voice from this list to narrate the text. Here you have all these different voices, including both male and female options.

You can also adjust the speed rate of the voice using the speed rate icon, but for our study, please keep the speed rates at their **default values.**

Once you’ve entered the text and selected a voice, click the “play” icon at the top center to generate the speech.

Now, you’re all set to explore regular past tense verbs and use Speechify TTS for English pronunciation practice.

[Closing Scene: Thank you!]

Instructor: Thanks for watching, and happy learning!

- *End of Video*

Appendix H

Training Materials: Activity #1

[This table shows how the voices are distributed for this activity]:

Treatment Group Input			Control Group Input		
Voice Name	**Voice Features**	**Sessions**	**Voice Name**	**Voice Features**	**Sessions**
Jamie	North American Female	1, 6	*Jamie*	North American Female	All sessions
Micheal	British Male	2, 7			
Sydney	Australian Female	3, 8			
Nate	North American Male	4, 9			
Stephanie	British Female	5, 10			

Instructions:

In this activity, you will need to do the following steps:

1) Copy and past the following text into Speechify.
2) Choose the voice (X)
3) Listen to the story without looking at the original text.
4) Fill in the blanks to complete this exercise.

Session 1: A Birthday Celebration

A year ago, my best friend, Lisa, **celebrated** her birthday. She **invited** all of her friends to a special party at her house. It was a sunny day, and she **decorated** the backyard with balloons. We **arrived** at four o'clock. The party **started** with games and laughter. We **played** musical chairs and **danced** to our favorite songs. Lisa's mom **baked** a delicious cake, and we sang "Happy Birthday" as Lisa blew out the candles.

Afterward, we **watched** a movie in the cozy living room. We **ate** popcorn and drank soda while we **enjoyed** the movie.

As the day came to an end, Lisa **thanked** us for making her birthday so special. We **loved** the party and we **left** with smiles on our faces.

[2 Distractors]

Session 2: The Bottle Cap Collector

In a small town, there was a boy named Max. Max **loved** collecting bottle caps. He didn't collect just any bottle caps; he **liked** ones with interesting designs and colors.

Every day after school, Max would look for bottle caps. He **searched** in recycling bins and **checked** each one carefully.

By the time he was twelve years old, Max **collected** one hundred bottle caps. He **created** a unique system to organize them. He **arranged** his bottle caps by color, design, and where he found them. His friends and family **started** to give him bottle caps too.

One day, Max read about an art contest. He **used** his bottle caps and **painted** them with beautiful colors. He **won** the first prize. Collecting bottle caps **helped** Max learn about recycling and making art from ordinary things.

Max's journey of collecting bottle caps **filled** his heart with a colorful sense of creativity and learning.

[2 Distractors]

Session 3: The Lost Wallet

Last Tuesday, Sarah **lost** her wallet while shopping in the busy downtown area. She **searched** her bag and retraced her steps, but somehow the wallet **vanished**.

Sarah **started** to panic. She **canceled** her credit cards and **reported** her driver's license as missing. The next morning, a kind man named Tom **contacted** Sarah. He found her wallet on the sidewalk, and he **checked** her contact information inside the wallet.

Grateful, Sarah **met** Tom to collect her wallet. She **thanked** him profusely, and they **chatted** for a while. They **discovered** they had a lot in common and **exchanged** phone numbers.

Over time, Sarah and Tom **became** close friends and **learned** a lot from each, all thanks to a lost wallet that brought them together.

[3 Distractors]

Session 4: The Unexpected Bus Stop

On a rainy Thursday morning, Laura **waited** for her usual bus at the familiar bus stop. However, the bus was late. She **checked** her watch, growing impatient.

Just as she considered taking a taxi to work, a lady named Maya approached the same bus stop and **asked**, "Is this the bus to downtown?"

"Yes" Laura **answered,** and they both **started** talking.

As they **talked**, they discovered that they both **worked** in the same office building, but on different floors. Maya said she had recently **moved** to the city and was still learning her way around.

When the bus finally **arrived**, Laura **invited** Maya to sit next to her. She **showed** her the best places to **eat** and where to find the nearest grocery store. Maya **suggested** meeting after work for dinner. They exchanged phone numbers, and they **became** good friends ever since that morning.

[2 Distractors]

Session 5: A Day at the Park

Last Sunday, my family and I **visited** a beautiful park near our home. We **reached** the park and **spotted** a perfect spot for a picnic.

We **unloaded** our picnic basket, spread out a cozy blanket under a shady tree, and **listened** to the birds **singing**. Everyone was excited, and my younger brother **started** playing with his favorite toys.

I **looked** around and saw a group of children **playing** near the swings. They **enjoyed** the sunny day.

After our picnic, we **packed** up and **cleaned** our area. It was essential to **leave** the park clean. As we **walked** back to the car, I **spotted** a beautiful bird in the distance.

We **arrived** home in the evening, tired but happy, with memories of a wonderful day at the park.

[3 Distractors]

Session 6: Sarah's Cookies

Sarah, a loving mom, **decided** to bake cookies for her family. She **gathered** flour, sugar, eggs, and chocolate chips. Sarah **mixed** all the ingredients and **placed** the doughs onto a baking sheet.

She **put** them in the oven and set a timer, but a phone call **distracted** her. When she **came** back, the cookies were burnt to a crisp.

Sarah didn't give up. She **repeated** the whole thing and made a fresh batch. This time, she **watched** the oven and **waited** carefully, so the cookies didn't overbake.

When the second batch came out, they were perfect. Her family **loved** them and had a good laugh about the ones she **burned** earlier. They **thanked** her for making the most delicious cookies.

Sarah **realized** that even when things go wrong, you can **turn** them into something wonderful.

[3 Distractors]

Session 7: Lily's Garden

Lily, a young girl, **loved** planting flowers in her garden. She spent lots of time caring for her plants. She **watered** them, sang to them, and read books about them.

One spring, Lily **planted** many colorful flowers. She **liked** each flower for its beauty and smell. She **learned** about gardening by taking classes and talking to other plant lovers.

One day, her neighbor, Mrs. Johnson, **visited** Lily's garden. She **admired** the pretty flowers and got a potted plant as a gift. Mrs. Johnson **thanked** Lily and **placed** the plant on the shelf of her window, making her home feel brighter.

As years **passed**, Lily became famous in her neighborhood for her lovely garden. She **hosted** garden parties and taught her neighbors about planting and taking care of flowers.

Lily's love for gardening taught her about patience and how to **take** care of living things. Her garden **made** her happy and brought joy to her neighborhood. It **reminded** everyone of how wonderful nature could be.

[2 Distractors]

Session 8: A Hardworking Dad

A dedicated dad named Michael **worked** really hard to support his family. Every morning, he **prepared** breakfast and **kissed** his kids goodbye as they left for school.

Michael had a job at a factory, where he **operated** heavy machines. He wore safety gear and **took** his job seriously because he knew his family **depended** on him.

During his lunch break, he read letters from his children. These letters **reminded** him of why he put in so much effort and **showed** love and appreciation.

In the evening, after work, Michael came home to his family. He **played** with his kids, **helped** them with homework, and listened to their stories.

On weekends, he **fixed** things around the house and taught his kids how to do it too. He **smiled** every day and was grateful for his family.

Michael's hard work and love **created** a warm and happy home where they **became** a close and strong family.

[2 Distractors]

Session 9: Lost in Paris

Sophie and Alex were best friends who **went** on a trip to Paris. They **explored** many famous places like the Eiffel Tower and the Louvre Museum.

One evening, they were walking by the river, but they took a wrong turn and got lost. The beautiful streets they had seen earlier now **looked** like a confusing maze. Both Sophie and Alex **started** to feel scared, but they **needed** to stay strong.

Instead of giving up, they **asked** people for help, using gestures and simple French words. The kind Parisians **guided** them into the right direction, leading them to **discover** new beautiful places. They **discovered** new shops and cozy cafes. Their adventure made them appreciate the city even more.

At last, they **reached** a busy area they **recognized**. Their adventure **ended** but they **realized** that getting lost had made their trip more interesting.

The experience taught them that even when things go wrong, they could **find** new and exciting things. Sophie and Alex became even closer friends as they **faced** the challenges together.

(3 Distractors)

Session 10: A Weekend Adventure

Last summer, my friends and I **decided** to go on a thrilling adventure. We **planned** to hike to the top of a nearby mountain. The weather **predicted** a sunny day, and we **looked** forward to the beautiful views from the summit.

As we **started** our hike, the path **went** through a beautiful forest. The trees **whispered** in the wind, and the birds **sang** above us.

About halfway up the mountain, we **spotted** a small stream. We **stopped** to rest and **dipped** our feet in the cool, refreshing water. It was a perfect break.

After a few hours of hiking, we **reached** the top. The view was breathtaking, and we **captured** it with our cameras.

As the sun set, we **began** our descent. It was a challenging but rewarding day, and we **returned** home with a sense of accomplishment and a collection of beautiful photos.

[3 Distractors]

Note: Information within brackets [ ] and color codes are for reader reference only and will not be shared with participants.

Appendix I

Training Materials: Activity #2

[This table shows how the voices are distributed for this activity]:

Treatment Group Input			Control Group Input		
Voice Name	**Voice Features**	**Sessions**	**Voice Name**	**Voice Features**	**Sessions**
Jamie	North American Female	1, 6	*Jamie*	North American Female	All sessions
Micheal	British Male	2, 7			
Sydney	Australian Female	3, 8			
Nate	North American Male	4, 9			
Stephanie	British Female	5, 10			

Instructions:

In this activity, you will need to do the following steps:

1) Copy and paste the verbs below one by one (12 verbs, 4 for each variation) in Speechify
2) Choose the voice (X)
3) Listen to the verbs' pronunciation.
4) Select the right verb based on their ending pronunciation of past *-ed*.

Session 1:

1) guided a) used b) asked c) added	**2) called** a) used b) asked c) added	**3) jumped** a) used b) asked c) added
4) attracted a) used b) asked c) added	**5) filmed** a) used b) asked c) added	**6) divided** a) used b) asked c) added
7) brushed a) used b) asked c) added	**8) played** a) used b) asked c) added	**9) talked** a) used b) asked c) added
10) waited a) used b) asked c) added	**11) rearranged** a) used b) asked c) added	**12) baked** a) used b) asked c) added

Session 2:

1) hunted a) used b) asked c) added	**2) picked** a) used b) asked c) added	**3) belonged** a) used b) asked c) added
4) completed a) used b) asked c) added	**5) walked** a) used b) asked c) added	**6) needed** a) used b) asked c) added
7) liked a) used b) asked c) added	**8) rained** a) used b) asked c) added	**9) passed** a) used b) asked c) added
10) ended a) used b) asked c) added	**11) enjoyed** a) used b) asked c) added	**12) moved** a) used b) asked c) added

Session 3:

1) wished a) used b) asked c) added	**2) planted** a) used b) asked c) added	**3) wasted** a) used b) asked c) added
4) answered a) used b) asked c) added	**5) washed** a) used b) asked c) added	**6) afforded** a) used b) asked c) added
7) burned a) used b) asked c) added	**8) stuffed** a) used b) asked c) added	**9) decided** a) used b) asked c) added
10) filled a) used b) asked c) added	**11) named** a) used b) asked c) added	**12) relaxed** a) used b) asked c) added

Session 4:

1) wanted a) used b) asked c) added	**2) knocked** a) used b) asked c) added	**3) noted** a) used b) asked c) added
4) offered a) used b) asked c) added	**5) painted** a) used b) asked c) added	**6) sniffed** a) used b) asked c) added
7) stayed a) used b) asked c) added	**8) created** a) used b) asked c) added	**9) looked** a) used b) asked c) added
10) followed a) used b) asked c) added	**11) guessed** a) used b) asked c) added	**12) cried** a) used b) asked c) added

Session 5:

1) finished a) used b) asked c) added	**2) welcomed** a) used b) asked c) added	**3) dragged** a) used b) asked c) added
4) accepted a) used b) asked c) added	**5) stopped** a) used b) asked c) added	**6) described** a) used b) asked c) added
7) kissed a) used b) asked c) added	**8) tested** a) used b) asked c) added	**9) expanded** a) used b) asked c) added
10) cared a) used b) asked c) added	**11) searched** a) used b) asked c) added	**12) hated** a) used b) asked c) added

Session 6:

1) allowed a) used b) asked c) added	**2) attached** a) used b) asked c) added	**3) shopped** a) used b) asked c) added
4) appreciated a) used b) asked c) added	**5) amused** a) used b) asked c) added	**6) attacked** a) used b) asked c) added
7) admired a) used b) asked c) added	**8) attempted** a) used b) asked c) added	**9) bleached** a) used b) asked c) added
10) admitted a) used b) asked c) added	**11) announced** a) used b) asked c) added	**12) approved** a) used b) asked c) added

Session 7:

1) attended a) used b) asked c) added	**2) covered** a) used b) asked c) added	**3) calculated** a) used b) asked c) added
4) cracked a) used b) asked c) added	**5) counted** a) used b) asked c) added	**6) camped** a) used b) asked c) added
7) compared a) used b) asked c) added	**8) clapped** a) used b) asked c) added	**9) complained** a) used b) asked c) added
10) chopped a) used b) asked c) added	**11) claimed** a) used b) asked c) added	**12) alerted** a) used b) asked c) added

Session 8:

1) coughed a) used b) asked c) added	**2) participated** a) used b) asked c) added	**3) boiled** a) used b) asked c) added
4) banned a) used b) asked c) added	**5) provided** a) used b) asked c) added	**6) bounced** a) used b) asked c) added
7) begged a) used b) asked c) added	**8) blushed** a) used b) asked c) added	**9) rated** a) used b) asked c) added
10) booked a) used b) asked c) added	**11) prevented** a) used b) asked c) added	**12) behaved** a) used b) asked c) added

Session 9:

1) realized a) used b) asked c) added	**2) quoted** a) used b) asked c) added	**3) blinked** a) used b) asked c) added
4) negotiated a) used b) asked c) added	**5) battled** a) used b) asked c) added	**6) limited** a) used b) asked c) added
7) choked a) used b) asked c) added	**8) blessed** a) used b) asked c) added	**9) reacted** a) used b) asked c) added
10) borrowed a) used b) asked c) added	**11) balanced** a) used b) asked c) added	**12) recovered** a) used b) asked c) added

Session 10:

1) tasted a) used b) asked c) added	**2) bumped** a) used b) asked c) added	**3) listed** a) used b) asked c) added
4) learned a) used b) asked c) added	**5) neglected** a) used b) asked c) added	**6) poured** a) used b) asked c) added
7) packed a) used b) asked c) added	**8) pulled** a) used b) asked c) added	**9) loaded** a) used b) asked c) added
10) mentioned a) used b) asked c) added	**11) marked** a) used b) asked c) added	**12) laughed** a) used b) asked c) added

Note: Information within brackets [ ] and color codes are for reader reference only and will not be shared with participants.

Appendix J

Training Materials: Activity #3

[This table shows how the voices are distributed for this activity]:

Treatment Group Input			Control Group Input		
Voice Name	**Features**	**Sessions**	**Voice Name**	**Voice Features**	**Sessions**
Jamie	North American Female	1, 6	*Jamie*	North American Female	All sessions
Micheal	British Male	2, 7			
Sydney	Australian Female	3, 8			
Nate	North American Male	4, 9			
Stephanie	British Female	5, 10			

Note: Information within brackets [ ] is for reader reference only and will not be shared with participants.

Instructions:

In this activity, you will need to do the following steps:

1) Copy and past the following text into Speechify.
2) Choose the voice (X)
3) Listen to the story without looking at the original text.
4) Answer the multiple-choice questions.

Session 1: The Lost Cat

Last month, Jessica was very worried because she had lost her beloved cat, “Fluffy”. Fluffy was a beautiful white cat with bright green eyes.

Jessica **looked** everywhere inside her house, but Fluffy was nowhere to be found. So, she **decided** to go outside and call out for Fluffy.

Jessica **walked** around the neighborhood. She **shouted** for Fluffy and **hoped** for a response. She met some neighbors who **offered** to help in the search. She also **posted** posters with Fluffy's photo around the area, wishing that someone might recognize her cat and provide some news.

As the days **passed**, Jessica’s worry only **deepened**. She couldn’t sleep at night, thinking about her lost cat. She **persisted** in her search every day, checking the nearby parks and alleys. Jessica **remained** heartbroken.

Then, one evening, as she was putting up more posters, a kind neighbor **called** her. He told her that Fluffy was in his backyard. Jessica was very happy. She **rushed** to her neighbor's place, and there, she found Fluffy safe. At last, they were together again.

Comprehension Questions:

1) What did Jessica lose?

a) Her bicycle b) Her favorite book

c) Her cat, Fluffy d) Her phone

2) How did Jessica try to find her lost cat?

a) She posted pictures of her cat. b) She asked her friends for help.

c) She searched her room more thoroughly. d) She went on a vacation.

3) Why did Jessica have trouble sleeping at night?

a) She had too much homework to do. b) She was worried about her lost cat.

c) She was watching TV late. d) She had a lot of friends over.

4) Where was Fluffy found?

a) In a police station b) In her neighbour's backyard

c) In a supermarket d) In a park

5) How did Jessica feel when she was reunited with Fluffy?

a) Angry b) Sad

c) Happy d) Tired

Session 2: A Day at the Beach

Last summer, John and his family went to the beach. They **wanted** to make it a special day by having a picnic on the sandy shore. The sun was bright, and the breeze **drifted** through the palm leaves. They spread a colorful beach blanket and **opened** their picnic basket filled with sandwiches, fruit, and lemonade. As they **enjoyed** their delicious meal, seagulls **circled** overhead.

After lunch, John's younger sister, Emily, built a sandcastle with the help of their dad. John and his mom **collected** seashells and small rocks along the waterline. The waves rolled in, and John couldn't resist diving into the cool, clear water. He **splashed** around and gave bodyboarding a try for the first time. His laughter **filled** the air as he **balanced** on the waves.

As the day gradually **faded** into evening, they all felt tired and ready to leave. They **packed** their things to go back home. The sunset was a breathtaking sight, with the sky painted in shades of pink, orange, and purple. They took pictures to remember this beautiful day.

Comprehension Questions:

1) What did John and his family do last summer?

a) They went shopping. b) They went to the beach.

c) They went to a theme park. d) They went to a museum.

2) What did they have for their picnic?

a) Pizza b) Sandwiches, fruit, and lemonade

c) Burgers and fries d) Sushi

3) What did Emily do at the beach?

a) She collected seashells. b) She played with a beach ball.

c) She built a sandcastle. d) She surfed on the waves.

4) How did John spend his time in the water?

a) He collected seashells. b) He played with his sister.

c) He built a sandcastle. d) He rode the waves and tried bodyboarding.

5) What was the sky like during the sunset?

a) Gray and cloudy b) Painted in shades of pink, orange, and purple

c) Bright and sunny d) Covered in stars

Session 3: Tom's Big Exam

Tom was an ordinary boy who **loved** playing soccer and exploring the woods near his house. One day, he had a big science exam and he felt both excited and a little nervous about it.

Tom's mom **noticed** her son's anxiety and said, "I see you're worried about your exam tomorrow, Tom. Let's go over your notes and make sure you're well-prepared." They spent the evening going over his textbooks and notes. Tom's mom patiently **explained** some of the difficult topics, making sure he understood them.

After a while, she **suggested**, "It's important to take a break and relax your mind too. Shall we play soccer to clear your mind?" Tom **agreed**, and they **headed** to the backyard. They **kicked** the soccer ball around and **laughed** a lot. It was a wonderful break from the studying. After the break, Tom's mom made a delicious cup of hot cocoa for him to help him study better. By the time they **finished** studying, Tom felt ready for his exam.

The next morning, Tom woke up early and **revised** his notes one more time. He felt a bit nervous, but he **remembered** his mom's encouraging words, "You've got this, Tom." At school, Tom

completed the exam with success. Happy with his achievement, Tom went home and **thanked** his mom for her support.

Comprehension Questions:

1) What was Tom feeling about his science exam?

a) Excited but a little nervous b) Happy and prepared

c) Bored and disinterested d) Hungry and tired

2) How did Tom's mom help him prepare for the science exam?

a) She baked cookies for him. b) She took him to the park.

c) She reviewed the notes with him. d) She watched TV with him.

3) What did Tom and his mom do during their break from studying?

a) They went for a long walk. b) They played soccer in the backyard.

c) They read a novel. d) They ate a big meal

4) What did Tom's mom make for Tom after the break?

a) Chocolate cake b) Cheese sandwich

c) Orange juice d) Hot cocoa cup

5) How did Tom feel after taking the science exam?

a) happy and thankful b) Nervous and unprepared

c) Sad and disappointed d) Tired and sleepy

Session 4: The School Play

Three years ago, Adam **participated** in a school play. He was excited because he got the role of the main character, a brave pirate. He **practiced** his lines with his friends and **rehearsed** for weeks.

When the day of the play **arrived**, Adam put on his pirate costume and stood backstage with his fellow actors. His heart was pounding, and he was a little nervous.

As the curtains **opened**, Adam **walked** onto the stage. He said his lines and **acted** out his part. The audience **clapped** and **cheered**, which made him feel more confident. He **enjoyed** every moment of the play, and when it was over, he felt a great sense of accomplishment.

After the play, Adam and his friends **celebrated** with pizza and ice cream. He was happy for doing a good job in the play, and he would always remember that special day at his old school.

Comprehension Questions:

1) What role did Adam have in the school play?

a) A teacher b) The main character, a pirate

c) The audience d) A musician

2) How did Adam feel before going on stage?

a) Confident b) Nervous

c) Bored d) Hungry

3) How did the audience react to the play?

a) They booed the actors. b) They clapped and cheered.

c) They fell asleep. d) They left the theater.

4) What did Adam do after the play?

a) He went home and took a nap. b) He practiced more for the next play.

c) He went to another play. d) He celebrated with friends and had pizza and ice cream.

5) How did Adam feel after the play?

a) Confused b) Happy and proud

c) Hungry d) Bored

Session 5: The Lost Necklace

Last year, during summer, Lily went to the park with her family. She wore a beautiful necklace that her grandmother had given her for her birthday. The necklace had a shiny silver heart and Lily **loved** it very much.

Lily and her family had a picnic at the park. They **played** games and **enjoyed** the day. After a fun day of activities, they **packed** their things and **headed** home.

At home, Lily **realized** that her lovely necklace was missing. She was very upset and told her family. They **suggested** to retrace their steps and look for the necklace. So, they **returned** to the park and **searched** everywhere. They **looked** around the picnic area, the playground, and the walking path. Lily even **checked** the area around the swing she had used. However, they couldn't find the necklace. Lily **started** to cry.

Just as they were about to leave, Lily's brother, Danny, **spotted** something shiny near the picnic area. It was her lost necklace! It had fallen on the ground when Lily was playing. Lily was so happy and **hugged** her brother.

Lily was very thankful for her family's help and **decided** to take better care of her special necklace.

Comprehension Questions:

1) What did Lily lose at the park?

a) Her family b) Her necklace

c) Her bicycle d) Her favorite book

2) What was special about Lily's necklace?

a) It was made of gold. b) It was a gift from her brother.

c) It had a shiny silver heart. d) It was her mother's necklace.

3) How did Lily and her family spend their day at the park?

a) They watched a movie b) They had a picnic and played games

c) They went shopping d) They had a barbecue

4) Where did Lily find her lost necklace?

a) In the playground b) On the walking path

c) Near the swing d) Near the picnic area

5) What did Lily learn from this experience?

a) She learned to be more careful with her special necklace.

b) She learned to never wear necklaces.

c) She learned to stop going to the park.

d) She learned to give her necklace to her brother.

Session 6: The Birthday Surprise

Three years ago, for Sarah's 25th birthday, her best friends **planned** a special surprise to make her day unforgettable. They secretly **asked** Sarah's family to help them organize a surprise party in the family's backyard. Sarah's favorite color was purple, so they **decorated** the garden with purple balloons, flowers, and a big banner that read, "Happy Birthday, Sarah!"

On her birthday, Sarah's family **invited** her to a movie night at home. Meanwhile, her friends **sneaked** into the house to prepare for the surprise. Sarah **arrived**, and her family **pretended** to watch a movie, but their hearts were pounding with excitement.

When the time was right, they **handed** her a clue that led her to the backyard. As she **stepped** through the door, everyone **jumped** out and **shouted**, "Surprise!".

Sarah was both shocked and delighted. She had a huge smile on her face.

They **played** games, sang karaoke, and **enjoyed** delicious homemade cake and ice cream. Sarah had a fantastic time, and she said it was the best birthday she had ever had.

Comprehension Questions:

1) Who planned a birthday surprise for Sarah?

a) Her teacher b) Her best friends

c) Her neighbor d) Her sibling

2) What was Sarah's favorite color?

a) Green b) Blue

c) Purple d) Red

3) Where did they decide to throw the surprise party?

a) In a restaurant b) In the backyard of Sarah's family

c) In the park d) Inside Sarah's house

4) How did they trick Sarah into going to the backyard?

a) They told her the truth about the surprise party.

b) They pretended to watch a movie at the backyard.

c) They handed her a clue that led her to the backyard.

d) They called her on the phone and asked her to come outside.

5) How did Sarah feel about the surprise party?

a) Angry b) Sad

c) Bored d) Shocked and delighted

Session 7: The New Student

Two years ago, Anna **moved** to a new town from a faraway city and **joined** a new school. She was nervous on her first day at school because her English was not very good.

A girl named Tina became friends with Anna and **helped** her adjust to the new school. Tina sat next to Anna in class, and during lunch breaks, Tina **introduced** Anna to other friends. Anna was a bit shy at first, but she soon began to smile and feel more comfortable.

During the breaks, Tina and the other students **assisted** Anna with her assignments. They also **practiced** English with her, which **boosted** her confidence in speaking.

As the weeks **passed**, Anna's English **improved**, and she made more friends. She even **founded** a new book club, where she became known as a great storyteller.

Everyone **welcomed** and **motivated** Anna at school. Anna's story tells us how being kind and helpful can really change someone's life.

Comprehension Questions:

1) Who was the new student in the school?

a) Tina b) Anna

c) The teacher d) The school principal

2) Why was Anna nervous on her first day of school?

a) She was excited to meet new friends. b) Her English was not very good.

c) She had moved from another town. d) She was a great soccer player.

3) How did the Tina help Anna with her course?

a) By ignoring her and not talking to her

b) By practicing English with her and introducing her to other friends

c) By teasing her and making fun of her English

d) By sitting alone and not talking to anyone

4) What did Anna do as time passed and her English improved?

a) She founded a book club. b) She made more friends.

c) She became less shy. d) She stopped speaking English.

5) What did you learn from Anna's story?

a) Being kind and helpful can change lives. b) English is a very easy language to learn.

c) Making friends is not important. d) Helping new students is a waste of time.

Session 8: The Snow Day

Two winters ago, my family and I woke up to a surprise. Our neighborhood **looked** like a winter wonderland, covered in deep white snow. Our school **canceled** the classes for the day! My brother and I were very excited.

We quickly **zipped** up our warmest winter clothes, **grabbed** our sleds, and **headed** outside. We built a snowman in front of our home, threw snowballs on each other, and **glided** down the nearby hill on our sleds.

In the afternoon, our parents **joined** us for a snowball fight, and we built a snow fort in the backyard. As the day **ended**, we went inside, tired but filled with the joy of a fantastic snow day.

We **warmed** up by the fireplace, **sipped** hot cocoa, and **talked** about our fun day. We will always remember that day.

Comprehension Questions:

1) What was the surprise that the family woke up to?

a) A visit from their grandparents b) A snow day

c) A power outage d) A birthday party

2) What did the family do as soon as they saw the snow?

a) They stayed inside and played video games. b) They went to a beach.

c) They cleaned their house. d) They went outside to enjoy the snow.

3) What activities did they do outside on the snow day?

a) They built sandcastles. b) They went sledding and had snowball fights

c) They played with water balloons. d) They rode their bikes.

4) What did they do in the afternoon?

a) They had a picnic in the snow. b) They went swimming in a frozen lake.

c) They built a snow fort in the backyard d) They went for a long walk.

5) What did they do as the day ended?

a) They went to a theme park. b) They played video games all night.

c) They went to sleep early. d) They warmed up by the fireplace and talked about their day.

Session 9: The Science Fair

Last year, Emily's school **organized** a science fair, and she **decided** to participate. She was excited but also a little nervous about presenting her project to judges and other students.

Emily's project **investigated** how plants grow under different light conditions. In her room, she grew three plants: one in natural sunlight, one with artificial light, and one in complete darkness.

For weeks, Emily **worked** hard caring for her plants, watering them, and measuring their growth. She carefully **recorded** the changes. She also **researched** and **prepared** a presentation to explain her project.

When the science fair day **arrived**, Emily set up her project display and **waited** for the judges to come around. She **explained** her experiment and **discussed** her findings confidently. As the fair

went on, Emily **visited** other projects and **learned** so much from her classmates. It was a fantastic opportunity to see the exciting and creative experiments they had come up with.

At the end of the day, the judges **announced** the winners, and Emily was thrilled to receive an award for her project. It was a great moment for her.

Comprehension Questions:

1) What event did Emily participate in?

a) A basketball game b) A science fair

c) A talent show d) A cooking competition

2) What was the topic of Emily's science project?

a) The history of music b) Animal behavior in the wild

c) Outer space and the planets d) Plants and their growth under different conditions

3) How many different light conditions did Emily use for her plant experiment?

a) One b) Two

c) Three d) Four

4) What did Emily do on the day of the science fair?

a) Watered the plants b) Watched TV

c) Played video games d) presented her project and explained it to judges

5) How did Emily feel at the end of the science fair?

a) Nervous and disappointed b) Excited and proud

c) Hungry and tired d) Bored and upset

Session 10: The Art Show

Last spring, Alex's school **hosted** an art show for students to showcase their artistic talents. Alex **decided** to participate and **started** painting a picture she had thought about for a long time.

Alex **picked** a beautiful scene with a colorful sunset over the ocean for her painting. She **selected** the colors carefully and **worked** on the painting for many weeks to make the picture look perfect.

As the day of the art show **approached**, Alex was excited but also a bit nervous. She carefully **framed** her painting and **placed** it on display. Many students and their families came to see the artwork, and Alex was proud to have her painting among the others. The art show was a great success, and many people **admired** the different paintings. Positive feedback on Alex’s artwork **boosted** her confidence.

In the end, the painting won an award for its beautiful colors and details. Alex was overjoyed and **realized** that her hard work and passion for art paid off. The art show was a wonderful experience, and it **inspired** Alex to keep painting and pursuing the love for art.

Comprehension Questions:

1) What event did Alex participate in?

a) A soccer game b) An art show

c) A cooking contest d) A science fair

2) What was the inspiration for Alex's painting?

a) A busy city street b) A beautiful colorful sunset over the ocean

c) A dark and gloomy forest d) A mountain covered in snow

3) How did Alex feel as the day of the art show approached?

a) Excited and nervous b) Sad and disappointed

c) Hungry and tired d) Bored and upset

4) Where did Alex place their painting for the art show?

a) In their bedroom b) On the kitchen table

c) In the backyard d) On display at the school's art show

5) What did Alex win for their painting?

a) A book b) A ticket

c) An award d) A trip to the beach

Note: Information within brackets [ ] and color codes are for reader reference only and will not be shared with participants.

Appendix K

Semi-Structured Interview Questions (adapted from Venkatesh & Davis, 2000)

Aim 1: [Perceived Usefulness – PU] - To identify the features and affordances that make TTS a **useful** technology for learning English pronunciation.

Q1: When using TTS to improve your English pronunciation, did you notice fewer or more errors compared to your pronunciation practice without it?

> Follow up: What aspects of pronunciation (e.g., word endings, intonation) did you consider when making this judgment?

Q2: Did TTS help you enhance your English pronunciation more quickly than other methods you have used in the past? Why or why not?

Q3: Do you think TTS serves as a useful tool for learning English pronunciation? Please explain your answer.

Aim 2: [Academic Relevance – AR] - To identify the features and affordances that make Speechify TTS a **relevant** technology for when learning English pronunciation.

Q1: Would you consider TTS to be an important technology for learning pronunciation? why or why not?

> Follow up: What about other skills? (grammar, vocabulary, writing)

Q2: If a friend or classmate wanted to enhance their English pronunciation, would you recommend using TTS? Please clarify your reasons. Do you think it could be relevant for learning in other courses or contexts?

Aim 3: [Output Quality – OQ] - To identify the features and affordances of TTS that affect the **quality** of the speech.

Q1: Did your pronunciation quality get better or worse using TTS for English pronunciation practice? Can you explain the effect in terms of these aspects: (clarity, fluency, accuracy, and accent)?

Q2: beneficial to English pronunciation quality? What aspects of TTS do you think impact the quality of your pronunciation?

Aim 4: [Result Demonstrability – RD] - To identify the features and affordances of TTS that affect the **results** of their pronunciation.

Q1: When using TTS, do you think that your improved pronunciation will give you better academic grades or better real-world outcomes? Why, why not??

Q2: Do you think using TTS for pronunciation practice, over an extended period, could lead to better results in your spoken English? Why or why not?

Aim 5: [Perceived Ease of Use – PEU] - To identify the features and affordances that make TTS **easy** to use when learning English pronunciation.

Q1: When you engaged with TTS, which features did you find easy to use? Were there any features you found less easy? Did you continue using them despite any difficulties? Please explain.

Q2: Can you describe how you handled any issues or challenges that arose while using TTS? Were they easy or difficult to resolve? Can you give examples of specific instances?

Aim 6: [Intention to Use – IU] - To identify if participants **intend to use** TTS to learn English pronunciation after participating in the study.

Q1: After participating in this study, do you plan to continue using TTS for improving your English pronunciation? What factors influence your decision to continue or discontinue using it?

Q2: Can you give examples of other situations you would use TTS to English?